

volume VII ♦ issue 1

2016

Logos & Episteme

an international journal
of epistemology

**Romanian Academy
Iasi Branch**



**“Gheorghe Zane” Institute
for Economic and Social
Research**

Editorial board

Director

Teodor Dima

Editor-in-Chief

Eugen Huzum

Executive Editors

Vasile Pleșca

Cătălina-Daniela Răducu

Assistant Editors

Irina Frasin

Bogdan Ștefanachi

Ioan Alexandru Tofan

Web&Graphics

Codrin Dinu Vasiliu

Virgil-Constantin Fătu

Simona-Roxana Ulman

Contact address:

Institutul de Cercetări
Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

www.logos-and-episteme.projectsbc.ro

Advisory board

Frederick R. Adams

University of Delaware, USA

Scott F. Aikin

Vanderbilt University, USA

Daniel Andler

Université Paris-Sorbonne, Paris IV, France

Alexandru Boboc

Academia Română, Romania

Panayot Butchvarov

University of Iowa, USA

Mircea Dumitru

Universitatea din București, Romania

Sanford Goldberg

Northwestern University, Evanston, USA

Alvin I. Goldman

Rutgers, The State University of New Jersey, USA

Susan Haack

University of Miami, USA

Stephen Hetherington

The University of New South Wales, Sydney, Australia

Paul Humphreys

University of Virginia, USA

Jonathan L. Kvanvig

Baylor University, USA

Thierry Martin

Université de Franche-Comté, Besançon, France

Jürgen Mittelstraß

Universität Konstanz, Germany

Christian Möckel

Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot

Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins

Universidade de Lisboa, Portugal

Duncan Pritchard

University of Edinburgh, United Kingdom

Nicholas Rescher

University of Pittsburgh, USA

Rahman Shahid

Université Lille 3, France

Ernest Sosa

Rutgers University, USA

Alexandru Surdu

Academia Română, Romania

John F. Symons

University of Texas at El Paso, USA

Logos &
Episteme

TABLE OF CONTENTS

Research Articles

Mihai HÎNCU, Games of Partial Information and Predicates of Personal Taste.....	7
Moti MIZRAHI, Why Gettier Cases Are Misleading.....	31
Alexander R. PRUSS, Being Sure and Being Confident That You Won't Lose Confidence.....	45
Michael J. SHAFFER, What If Bizet and Verdi had Been Compatriots?.....	55

Discussion Notes/Debate

Fred ADAMS, Murray CLARKE, Two Non-Counterexamples to Truth-Tracking Theories of Knowledge.....	67
Rodrigo BORGES, A Failed Twist to an Old Problem: A Reply to John N. Williams.....	75
John N. WILLIAMS, Still a New Problem for Defeasibility: A Rejoinder to Borges.....	83
Luis ROSA, Justification and the Uniqueness Thesis Again – A Response to Anantharaman.....	95
Mark SCHROEDER, Knowledge Based on Seeing.....	101
Notes on the Contributors.....	109
<i>Logos and Episteme</i> . Aims and Scope.....	113
Notes to Contributors.....	115

RESEARCH ARTICLES

GAMES OF PARTIAL INFORMATION AND PREDICATES OF PERSONAL TASTE

Mihai HÎNCU

ABSTRACT: A predicate of personal taste occurring in a sentence in which the perspectival information is not linguistically articulated by an experiencer phrase may have two different readings. In case the speaker of a bare sentence formed with a predicate of personal taste uses the subjective predicate encoding perspectival information in one way and the hearer interprets it in another way, the agents' acts are not coordinated. In this paper I offer an answer to the question of how a hearer can strategically interact with a speaker on the intended perspectival information so that both agents can optimally solve their coordination problem. In this sense, I offer a game-theoretical account of the strategic communication with expressions referring to agents' perspectives, communication which involves the interaction between a speaker who intends to convey some perspectival information and who chooses to utter a bare sentence formed with a predicate of personal taste, instead of a sentence in which the perspectival information is linguistically articulated by an experiencer phrase, and a hearer who has to choose between interpreting the uttered sentence in conformity with the speaker's autocentric use of the predicate of personal taste or in conformity with the speaker's exocentric use.

KEYWORDS: predicates of personal taste, autocentric, exocentric, communication, games of partial information, Nash equilibrium, Pareto dominance

1. Introduction

The ability of using language to communicate is an important part of human agency which involves, without any doubts, elements of cooperation. In order to successfully communicate by means of language, agents must coordinate on the intended meaning of the uttered sentences. In everyday life, it is not uncommon for people to talk to each other about their likes and dislikes, about their tastes and preferences, or about their perspectives from which they conceive the reality. In order to do that, they exploit in communication the fragment of natural language which consists of subjective predicates. One subclass of these predicates which the agents often use to express perspectival information is the class of predicates of personal taste.

Depending on which particular perspective the speaker refers to when he intends to communicate to the hearer perspectival information, a predicate of

personal taste occurring in a sentence in which the perspectival information is not linguistically articulated by an experiencer phrase may have two different readings. In the case in which the speaker of a bare sentence formed with a predicate of personal taste uses the subjective predicate encoding perspectival information in one way and the hearer interprets it in another way, things go wrong. Cases of this kind, in which the agents do not coordinate on the intended perspectival information, constitute instances of a more general case in which the agents' acts are not coordinated.

In this paper I will offer an answer to the question of how a hearer can coordinate with a speaker on the intended perspectival information conveyed by an utterance of a bare sentence formed with a predicate of personal taste, and I will show that there is a systematic way in which both agents can optimally solve their coordination problem and rationally avoid problems due to miscommunication. In this sense, in order to isolate some of the semantic properties exemplified by the utterances of bare sentences formed with predicates of personal taste, I will compare them, in the next section of the paper, with utterances of sentences in which an indexical like "I" occurs. Both types of sentences have a context-sensitive profile which explains the variation, in different contexts, of their semantic contents and of their truth-values. In contrast with the speaker-oriented semantics of sentences in which the first person singular pronoun occurs, the semantics of bare sentences with predicates of personal taste is not necessarily speaker-oriented. In this sense, I will introduce the distinction between the autocentric and exocentric interpretations of predicates of personal taste, interpretations which correspond to the different values the context of utterance provides to the variable for perspective occurring in the logical forms of bare sentences formed with predicates of this kind. In the third section of the present paper, I will offer a game-theoretical account of the strategic interaction between a speaker who intends to convey some perspectival information and who chooses, in this sense, to utter a bare sentence formed with a predicate of personal taste, instead of a sentence in which the perspectival information is linguistically articulated by an experiencer phrase, and a hearer who has to choose between interpreting the uttered sentence in conformity with the speaker's autocentric use of the predicate of personal taste or in conformity with the speaker's exocentric use of the predicate. The present game-theoretical account of strategic communication with expressions referring to agents' perspectives predicts that if the situation in which the speaker intends to convey to the hearer perspectival information about himself is factual, then the unique Pareto-Nash equilibrium of the game modeling the situation will correspond to

the balance between the speaker's choice to utter a bare sentence formed with a predicate of personal taste and the hearer's choice to interpret the utterance as expressing the perspective-specific proposition which corresponds to the speaker's autocentric use of the predicate of personal taste.

2. Predicates of Personal Taste

Predicates of personal taste are linguistic devices used to convey perspectival information. In order to explain how these linguistic tools can be used in real-world communication to express perspectival information, the truth-conditional semantics had to make room for the subjective meanings encoded by the predicates of personal taste. One such semantic theory, in which the truth-conditions of sentences formed with predicates of personal taste are accommodated with the subjective meanings lexicalised by these predicates, is *meaning perspectivalism*.¹ There is, however, a second variety of perspectivist semantics, more precisely, *truth perspectivalism*, according to which the predicates of personal taste are monadic predicates whose extensions vary depending on the values of a parameter which represents the perspective and which is placed in the circumstances with respect to which the truth-values of the utterances of bare sentences formed with predicates like these are evaluated.² The game-theoretical account which I will offer in the following section frames the problem of coordination of the speaker and the hearer on the intended perspectival information conveyed by utterances of bare sentences formed with predicates of personal taste in the terms and spirit of meaning perspectivalism.

According to meaning perspectivalism, at the level of the logical forms of sentences formed with predicates of personal taste, there are variables whose role is to represent the agents' perspectives.³ In this perspectivist semantics, utterances of bare sentences formed with predicates of personal taste express perspectival information by means of the values which the contexts of utterances assign to the variables representing the perspectives of the contextually salient agents. Therefore, even though the perspectival information is not represented by

¹ Jonathan Schaffer, "Perspective in Taste Predicates and Epistemic Modals," in *Epistemic Modality*, eds. Andy Egan and Brian Weatherson (Oxford: Oxford University Press, 2011), 179-226.

² Schaffer, "Perspective in Taste Predicates," 188; Herman Cappelen and John Hawthorne, *Relativism and Monadic Truth* (Oxford: Oxford University Press, 2009); Claudia Bianchi, "Contextualism," in *Philosophical Perspectives for Pragmatics*, eds. Marina Sbisa, Jan-Ola Östman, and Jef Verschueren (Amsterdam: John Benjamins, 2011), 64-66; Emma Borg, *Pursuing Meaning* (Oxford: Oxford University Press, 2012), 23-27.

³ Schaffer, "Perspective in Taste Predicates," 191.

a lexical unit from the surface syntax of a sentence formed with a predicate of personal taste, this information enters into the semantic content expressed by uttering the sentence in a context. In order to facilitate comprehension, consider, as an example, the following sentence:

[1] Philosophy is fun.

Consider also that the function μ is a semantics, that u is a situation in which an arbitrary expression s is uttered, and that the pair $\langle w, t \rangle$, which consists of a possible world w and of a time t , represents the index with respect to which the extension of s is determined. According to meaning perspectivalism, the semantic value of the predicate of personal taste which occurs in the above displayed sentence is functionally represented in the following way:⁴

$$\mu(\text{fun})^{u, \langle w, t \rangle} = \lambda x_e. \lambda y_e. x \text{ is fun to } y \text{ in } w \text{ at } t.$$

Depending on the values contextually assigned to the variable y , a sentence like [1] above will express, in different contexts, different propositions. Insofar as the perspectival information is a syntactically unprofiled constituent of the proposition expressed by uttering the sentence [1] in a particular context, this proposition is considered to be, in meaning perspectivalism, a *perspective-specific proposition*.⁵

Meaning perspectivalism conceives the semantics of the natural language fragment containing predicates of personal taste in the same way in which the semantics of the natural language fragment containing indexicals is conceived. In this regard, the conceptual framework by means of which the semantic values of predicates of personal taste are construed is that of Kaplanian semantics. What justifies, in meaning perspectivalism, the methodological import which amounts to semantically treating predicates of personal taste as indexicals, is the context-sensitivity exhibited by these predicates. In order to capture the context-dependence of the semantic values of indexicals and demonstratives, Kaplan has distinguished between the *character* of an expression and its *content*.⁶ While the former is represented as a function from the set of contexts of utterance to the set of semantic contents, the latter is understood as a function whose domain is the set of circumstances in which simple or complex expressions are evaluated and whose range is the set of their extensions. In order to better grasp the difference between these levels of meaning, consider that one and the same sentence

⁴ Eric Snyder, "Binding, Genericity, and Predicates of Personal Taste," *Inquiry* 56 (2013): 282.

⁵ Schaffer, "Perspective in Taste Predicates," 184.

⁶ David Kaplan, "Demonstratives," in *Themes from Kaplan*, eds. Joseph Almog, John Perry, and Howard Wettstein (Oxford: Oxford University Press, 1989), 481–563.

containing an indexical, like [2] below, is uttered by Mihai in a context c_1 and by Irina in a context c_2 :

[2] I love philosophy.

With regard to this particular example, the semantic explanation offered to the intuition according to which Mihai and Irina said the same thing appeals to the fact that the character of the sentence uttered by Mihai in the context c_1 does not differ from the character of the sentence uttered by Irina in the context c_2 . Likewise, what explains the intuition according to which the agents of c_1 and c_2 said different things when they utter the sentence [2], is the fact that [2] expresses different semantic contents, one corresponding to the proposition [Mihai loves philosophy], the other to the proposition [Irina loves philosophy].

At this point, the similarities between the semantic behavior of indexicals and that of predicates of personal taste become more transparent. One feature that both classes of expressions have in common is their *alethic variability*. In this sense, one and the same sentence in whose surface syntax occurs an indexical item or a predicate of personal taste can have, in different contexts, different truth-values. Consider that only Mihai finds philosophy fun and loves reading philosophy papers. In the case in which Mihai utters the sentences [1] and [2] in a context c_3 and Irina utters them in a context c_4 , Mihai's utterances are both true, while Irina's utterances of the same sentences are false. What explains the variation in truth-values of the different utterances of the sentence [2] is the occurrence, at the level of the logical form of [2], of a variable, representing the speaker, whose values are provided by the contexts in which [2] is uttered. In the same vein, the truth-conditional effects of the contexts c_3 and c_4 in which the sentence [1] is uttered are traced to the presence, at the level of the logical form of [1], of a variable representing the perspective of the contextually salient agent. Insofar as the values that the context c_3 provides to the variables present in the logical forms of [1] and [2] are different from the values assigned by the context c_4 to the same variables, c_3 and c_4 have different contextual contributions to the semantic contents of the utterances of the sentences [1] and [2]. In this case, what the agent of the context c_3 says, when he utters the sentence [1], can be equated with the perspective-specific proposition [Philosophy is fun for Mihai], while the semantic content of the utterance of the same sentence by the agent of the context c_4 will correspond to the perspective-specific proposition [Philosophy is fun for Irina]. Likewise, the semantic contents of the utterances of the sentence [2] in c_3 and c_4 are, as I already said, the perspective-neutral proposition [Mihai loves philosophy] and, respectively, the perspective-neutral proposition [Irina loves philosophy]. Hence, the variation of the values contextually assigned to the

variables which occur at the level of the logical forms of sentences formed with predicates of personal taste or with indexicals, like [1] and [2] above, explains the variation of the semantic contents of their different utterances which, in its turn, explains the variation of the truth-values of the propositions which these utterances contextually express.⁷ The alethic variability of utterances of sentences like [1] and [2] above is warranted by the fact that predicates of personal taste and indexicals are context-sensitive expressions whose characters, according to the Kaplanian semantics, are not constant functions.⁸ Since the character of a context-sensitive expression is not a constant function, the sentences syntactically constructed with expressions of this kind, will express, depending on the contexts in which they are uttered, variable semantic contents.

Another feature that predicates of personal taste and indexicals have in common is the way in which pragmatic factors intervene in the process by means of which the semantic values of these expressions are determined. Even though an interpreter of a sentence in which a predicate of personal taste or an indexical occurs knows the semantic roles associated with these expressions, he is forced to consult the context in which the sentence was uttered and to extract from there the needed information in order to determine the semantic content of the utterance and the semantic values of its parts. Insofar as the interpreter who exploits contextual information to resolve the references of indexicals and of predicates of personal taste is guided in his task by their semantics, both types of expressions involve a *semantic use* of context.⁹ In order to understand how a context can be used in a semantic way, it is better to bear in mind the distinction that Recanati drew between saturation and modulation. Even though both saturation and modulation belong to the class of primary pragmatic processes, there is a sharp contrast between them. While modulation is, from a linguistic viewpoint, an *optional* pragmatic process whose function is to informationally enrich the semantic contents of utterances, saturation is a linguistically *mandatory* pragmatic process by means of which the references of indexicals and demonstratives are contextually resolved and of which the values of the variables from the logical forms of sentences are assigned.¹⁰ Even though in both cases the interpreter exploits contextual information, only in the case of saturation the appeal to context is induced by the presence, in the surface or deep syntax, of

⁷ Bianchi, "Contextualism," 65.

⁸ Kaplan, "Demonstratives," 481-563.

⁹ John Perry, *Reference and Reflexivity* (Stanford: CSLI Publications, 2001), 39-42.

¹⁰ François Recanati, *Literal Meaning* (Cambridge: Cambridge University Press, 2004); François Recanati, *Truth-Conditional Pragmatics* (Oxford: Oxford University Press, 2010).

elements which demand to be contextually filled. The contextual provision of values to the variables which represent, at the level of the logical form of sentences like [1] and [2] above, the perspective of the salient agent, and, respectively, the speaker, is linguistically controlled, which means that the same pragmatic mechanism (i.e., saturation) operates both in the case of indexicals and in the case of predicates of personal taste.

In spite of all the features which sentences containing an indexical like “I” and bare sentences formed with predicates of personal taste have in common, there is one aspect which highlights a pragmatic contrast between them. Consider again the sentences [1] and [2] displayed above. If Mihai utters the sentence [2] in the context c_3 and Irina utters it in the context c_4 , the referent of the indexical occurring in [2] is, in both cases, a constituent of the semantic content expressed by uttering [2]. But insofar as the contexts c_3 and c_4 are different, because the agents of c_3 and c_4 are not the same, the referents of the indexical occurring in [2] will be different, and therefore the propositions which contain these referents will be different. The constituent of the proposition expressed by uttering the sentence [2] in the context c_3 is the agent of c_3 , that is, the speaker of [2] in c_3 , while the constituent of the proposition expressed by uttering [2] in the context c_4 is the agent who utters [2] in c_4 . This has to do with the semantics of the indexical occurring in [2] which requires that the value contextually assigned, by means of the pragmatic process of saturation, to the variable present in the logical form of [2], has to be the agent of the context in which [2] is uttered, that is, the speaker of [2]. In this sense, it can be said that the semantics of sentences in which the first person singular pronoun occurs is a *speaker-oriented semantics*. But the generalization licensed by the semantics of the indexical occurring in [2], according to which the utterances of sentences like [2] refer to their speakers, is not supported by the semantics of predicates of personal taste. Consider that Mihai utters the sentence [1] in the context c_3 and Irina utters it in the context c_4 . Among the constituents of the proposition expressed by uttering [1] in the context c_3 we find the perspective of the agent of c_3 . Similarly, one of the constituents of the proposition expressed by uttering [1] in the context c_4 is the perspective of the agent who utters [1] in c_4 . In both these cases, the perspective of the agent who utters the sentence [1] is a constituent of the semantic content expressed by uttering [1]. But it cannot be inferred from these data that the utterances of bare sentences formed with predicates of personal taste always refer to the perspectives of their speakers. The semantics of bare sentences in which predicates of personal taste occur does not impose the restriction that the value contextually assigned to the variable present in the logical form of [1], has to be the perspective of the

agent who utters [1]. The latter semantic requirement is relaxed in the case of the predicates of personal taste and this can easily be seen if we take into consideration a scenario in which, even though Irina does not find philosophy fun, she intends to utter the sentence [1] in order to communicate to her interlocutor information about Mihai's perspective, not about her. Even though the variable which occurs at the level of the logical form of [1] contextually receives a value which corresponds to a specific perspective of an agent, this does not imply that the value assigned to the variable must correspond to the perspective of the agent who utters [1], that is, to Irina's perspective. Instead, in the latter considered scenario, it corresponds to Mihai's perspective. Therefore, in contrast with the speaker-oriented semantics of a sentence in which the first person singular pronoun occurs, the semantics of a bare sentence formed with a predicate of personal taste is not necessarily speaker-oriented: the value contextually assigned to the variable present at the level of the logical form of a sentence of this kind can correspond to the perspective of the agent who utters the sentence, but, as well, to the perspective of another agent.

Consider that e is a bare sentence formed with a predicate of personal taste, like [1] above. In the case in which the value contextually provided to the variable for perspective present in the logical form of e corresponds to the perspective of an agent different from the speaker who utters e , or even to the perspective of an entire group, the predicate of personal taste occurring in e is used by the speaker of e in an *exocentric* way.¹¹ In this case, the utterance of e will not convey perspectival information about the agent of the utterance and the speaker's perspective will not be a constituent of the perspective-specific proposition contextually expressed by uttering e . In contrast, when the value contextually provided to the variable for perspective present in the logical form of e corresponds to the perspective of the speaker who utters e , the predicate of personal taste occurring in e is used by the speaker of e in an *autocentric* way.¹² In this case, the utterance of e will express perspectival information about the agent of the utterance and the speaker's perspective will be a constituent of the perspective-specific proposition contextually expressed by uttering e .

¹¹ Peter Lasersohn, "Context Dependence, Disagreement and Predicates of Personal Taste," *Linguistics and Philosophy* 28 (2005): 643-686; Cappelen and Hawthorne, *Relativism and Monadic Truth*, 104; John MacFarlane, *Assessment Sensitivity. Relative Truth and Its Applications* (Oxford: Oxford University Press, 2014).

¹² Lasersohn, "Context Dependence, Disagreement and Predicates of Personal Taste," 643-686; Cappelen and Hawthorne, *Relativism and Monadic Truth*, 104; MacFarlane, *Assessment Sensitivity*.

The fact that speakers who utter bare sentences formed with predicates of personal taste use these predicates in more than one way can put the utterances' interpreters in difficult situations. Situations like these constitute instances of the more general and classical problem of coordination. If a speaker who utters in a context a sentence *e*, in whose surface syntax the perspectival information is not explicitly mentioned by an experiencer phrase, autocentrically uses the predicate of personal taste occurring in *e*, while the hearer assigns to the utterance of *e* an interpretation which would correspond to the speaker's exocentric use of the predicate, the two agents' acts are not coordinated. Similarly, if the speaker of *e* has the intention to communicate to the hearer perspectival information about an agent different from the speaker, while the hearer understands that the perspective of the speaker is a constituent of the perspective-specific proposition expressed by the utterance of *e*, the possibility of successful communication is compromised. In order to restore it, the agents' acts of utterance and of interpretation must be aligned. This would ensure that the speaker's choice of an utterance of a sentence like *e*, in which the perspectival information is not linguistically articulated, and the hearer's choice of its interpretation are balanced. But how can the speaker and the hearer arrive at this optimal solution of their coordination problem? Is there a systematic way which specifies how the speakers of bare sentences formed with predicates of personal taste and the interpreters have to act in order to be rational and to solve and avoid problems due to miscommunication?

In what follows, I will show how a hearer can coordinate with a speaker on the intended perspectival information conveyed by uttering a sentence like *e*. In this regard, I will offer, in the next section of the present paper, a game-theoretical account of the strategic interaction between a speaker who intends to convey some perspectival information and who chooses, in this sense, to utter a bare sentence formed with a predicate of personal taste, instead of a sentence in which the perspectival information is linguistically articulated by an experiencer phrase, and a hearer who has to choose between interpreting the uttered sentence in conformity with the speaker's autocentric use of the predicate of personal taste or in conformity with the speaker's exocentric use of the predicate.

3. Games of Partial Information

In order to model the strategic communication between a hearer and a speaker of a sentence formed with a predicate of personal taste in which the perspectival information is not linguistically articulated by an experiencer phrase, and to show how the hearer can coordinate with the speaker on the intended perspectival

information conveyed by an utterance of a bare sentence formed with a predicate of personal taste, I will use the conceptual framework of games of partial information elaborated by Parikh and the format in which van Rooy has framed the games designed by Parikh.¹³

Given that the autocentric uses of the predicates of personal taste which occur in sentences in which the perspectival information is not linguistically articulated are more frequent¹⁴ than their exocentric uses, I choose to model, in this paper, only the situations in which the hearers have to interpret the bare sentences formed with predicates of personal taste which the speakers choose to utter in order to convey information about their own perspectives. The present model can be extended and accordingly adapted also for the cases in which a hearer has to strategically interact with a speaker who intends to convey perspectival information about another agent and who utters, in this sense, a sentence formed with a predicate of personal taste in which this information is not linguistically articulated by an experiencer phrase.

Consider a scenario in which *A* has recently met *B*, they moved together and they want to invite an old friend *C* of *A*'s to dinner at their home. *A* and *B* begin to talk about the food they will serve to *C* and about their likes and dislikes. In this context, *A*, who has the intention to talk about his tastes and to convey to *B* information about his own perspective, utters the following sentence:

[3] Lasagna is delicious.

Let e_1 abbreviate the above bare sentence formed with a predicate of personal taste which the speaker *A* uses, in the utterance situation *u*, to convey perspectival information to the hearer *B*. Insofar as the agent's perspective is not profiled in the surface syntax of sentences like e_1 , and as sentences of this kind are used in contexts to express perspectival information, it is common knowledge to both *A* and *B* that *A* can use the predicate of the sentence e_1 in two ways. The speaker can use the predicate of personal taste occurring in e_1 in an autocentric way, intending to convey information about his own perspective, or in an exocentric way, having in mind the intention to convey to the hearer *B* information about the perspective of another agent, namely *C*. If *A* uses in *u* the predicate of personal taste occurring

¹³ Prashant Parikh, *The Use of Language* (Stanford: CSLI Publications, 2001); Prashant Parikh, *Language and Equilibrium* (Cambridge: MIT Press, 2010); Robert van Rooy, "Signalling Games select Horn Strategies," *Linguistics and Philosophy* 27 (2004): 493-527.

¹⁴ Lasersohn, "Context Dependence, Disagreement and Predicates of Personal Taste," 673-674; Tamina Stephenson, "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste," *Linguistics and Philosophy* 30 (2007): 520-521; Hazel Pearson, "A Judge-Free Semantics for Predicates of Personal Taste," *Journal of Semantics* 30 (2013): 115.

in e_1 in an autocentric way, as it happens in the present scenario, the sentence e_1 expresses the perspective-specific proposition [Lasagna is delicious to A]. Likewise, if the predicate of e_1 is used by A in an exocentric way, the sentence e_1 will express in u the perspective-specific proposition [Lasagna is delicious to C]. Let p_1 abbreviate the perspective-specific proposition expressed in u by the autocentric reading of the predicate of e_1 , and p_2 abbreviate the perspective-specific proposition expressed in u by the exocentric reading of the very same predicate. Regarding what B considers possible in this scenario and the information he has, the meaning of the above mentioned sentence [3] can be functionally represented as follows:

$$\mu(e_1)^{u, \langle w, t \rangle} = \{p_1, p_2\}.$$

This means that, in this scenario, the hearer B is confronted with the following two choices: either he interprets the sentence e_1 as meaning p_1 , or he interprets it as meaning p_2 . Even though B does not know which of p_1 and p_2 is the particular perspective-specific proposition A intends to communicate by uttering e_1 , it is plausible to reckon that, in this scenario, based on the information provided by the utterance situation u and on the fact that bare sentences with predicates of personal taste are usually used in autocentric ways, B will correctly choose to interpret e_1 as meaning p_1 .

The scenario described above constitutes an example of situation in which the agents strategically interplay with one another. In order to game-theoretically model the strategic interaction between A and B , I will adopt the general assumptions¹⁵ on which relies the conceptual framework of games with partial information elaborated by Parikh and I will adapt them to the present case. Hence, I will assume that:

- (1) Both the speaker A and the hearer B are rational agents.
- (2) E is the fragment of language containing sentences with predicates of personal taste.
- (3) A and B competently use E .
- (4) The function μ is the semantics of E .
- (5) The target set of μ is the power set of the set of perspective-specific propositions.
- (6) A intends to linguistically express the perspective-specific proposition p_1 .
- (7) A uses in the utterance situation u an element e_1 of E .
- (8) The perspectival information is not profiled in the surface syntax of e_1 .
- (9) B intends to interpret the bare sentence e_1 uttered by A in u .

¹⁵ Parikh, *The Use of Language*, 21-23; Parikh, "Communication, Meaning, and Interpretation," *Linguistics and Philosophy* 23 (2000): 193-194.

- (10) B interprets e_1 .
- (11) According to B , $\mu(e_1)^{u, < w, t >} = \{p_1, p_2\}$.
- (12) $\mu(e_1)^{u, < w, t >} = \{p_1\}$, if A autocentrically uses the predicate of e_1 .
- (13) $\mu(e_1)^{u, < w, t >} = \{p_2\}$, if A exocentrically uses the predicate of e_1 .
- (14) p_1 is more likely than p_2 .
- (15) The effort of producing linguistic forms which explicitly express p_1 and p_2 is greater than the effort of producing e_1 .
- (16) The effort of processing linguistic forms which explicitly express p_1 and p_2 is greater than the effort of processing e_1 .
- (17) All of the above, except (6) and (9), are common knowledge to A and B .

The above assumptions ensure that, in the scenario previously described, the speaker A will successfully communicate, by using the sentence e_1 in the utterance situation u , the perspectival information p_1 to the hearer B .

The strategic interaction between a speaker A who utters in a context a bare sentence formed with a predicate of personal taste and a hearer B who tries to figure out whether A is using autocentrically or exocentrically the predicate of personal taste occurring in the received sentence, can be modeled as a two-agent game of partial information which A and B play, more precisely, a game whose unique solution is a Pareto-efficient Nash Equilibrium.¹⁶ In what follows, I will show that in the game which models the above described scenario, the optimal choice of A is to utter, in u , the sentence e_1 and the optimal choice of B is to assign to e_1 the interpretation p_1 , that is, that perspective-specific proposition which corresponds to the speaker's autocentric use of the predicate of personal taste occurring in e_1 and to his intention to convey information about his own perspective.

According to the contextual assumption (6), the agent A intends to linguistically communicate the perspective-specific proposition p_1 to the agent B . In order to accomplish this task, A has the following two possibilities: either he chooses to utter the sentence e_1 , that is, a linguistic expression belonging to E in whose surface syntax the perspectival information is not profiled, or he chooses to utter another sentence e_2 in which the perspectival information is syntactically represented by an experiencer phrase which refers to his own perspective. In this sense, one such sentence in which A 's perspective is explicitly mentioned by some linguistic material can be the following:

[4] Lasagna is delicious to me.

Let e_2 abbreviate the above sentence which A could utter in order to explicitly convey perspectival information to the hearer B . A sentence like e_2

¹⁶ Parikh, "Communication, Meaning," 207.

would make transparent to B the fact that the speaker uses the predicate of personal taste which occurs in e_2 in an autocentric way and that the perspective-specific proposition which A intends to express by uttering e_2 in the situation u is p_1 . According to the contextual assumption (15), the effort of A to produce the sentence e_2 in which the perspectival information is linguistically articulated is greater than the effort of producing the sentence e_1 which does not linguistically articulate the perspectival information, even though it expresses, in the above described scenario, the same perspective-specific proposition as e_2 , that is p_1 . Insofar as the speaker A is, according to the assumption (1), a rational agent, A has to take into consideration¹⁷ the consequences of his choice to utter the sentence e_1 along with the consequences of his choice to utter the sentence e_2 , and to compare the former consequences with the latter ones in order to decide which of e_1 and e_2 is the optimal linguistic form to express the semantic content p_1 .

In the above scenario, the speaker A has decided that in order to communicate to B information about his own perspective, his optimal action is to utter the sentence e_1 and to autocentrically use the predicate of personal taste occurring in e_1 . According to the contextual assumption (9), B intends to interpret A 's utterance, in the situation u , of the bare sentence formed with a predicate of personal taste. Insofar as the utterance of the sentence e_1 in u can express, according to the contextual assumption (11), either the perspective-specific proposition p_1 or the perspective-specific proposition p_2 , the hearer B cannot decide whether A uses the predicate of personal taste occurring in e_1 in an autocentric or in an exocentric way. B knows only that if, in the above scenario, the speaker A autocentrically uses the predicate of personal taste which occurs in the sentence e_1 , then, according to the contextual assumption (12), the utterance of e_1 in the situation u expresses the perspective-specific proposition p_1 . Likewise, he knows that in the case in which A exocentrically uses the predicate of personal taste occurring in e_1 , then, according to the contextual assumption (13), the utterance of the sentence e_1 in the situation u will express the perspective-specific proposition p_2 . Let s_1 denote the situation in which the speaker A intends to communicate to the hearer B , by uttering e_1 , the perspectival information corresponding to his autocentric use of the predicate of personal taste occurring in e_1 , and let s_2 denote the situation in which A intends to convey to B , by uttering e_1 , the perspectival information corresponding to the exocentric use of the very same predicate. If A is in the situation s_1 , he intends to use the sentence e_1 to linguistically communicate to the hearer the perspective-specific proposition p_1 ,

¹⁷ Parikh, "Communication, Meaning," 196.

and to transmit information about his location¹⁸ in s_1 . Similarly, if the speaker is in the situation s_2 , he intends to use the bare sentence containing a predicate of personal taste to convey to B the perspective-specific proposition p_2 , and to transmit information about his location in s_2 . According to the above described scenario, s_1 is a factual situation, s_2 is a counterfactual one and only A can discriminate between them. While A knows which of s_1 and s_2 is the factual situation, B does not know and this is common knowledge to both agents.¹⁹ Insofar as B does not have enough information to decide which of s_1 and s_2 is the factual situation, he does not know A 's intention yet and both epistemic possibilities form B 's information set. What the hearer B knows instead, according to the contextual assumptions (14) and (17), is that the perspective-specific proposition p_1 is more likely than the perspective-specific proposition p_2 , which means that B knows that the situation in which A autocentrically uses the predicate of personal taste which occurs in the sentence e_1 is more probable than the situation in which he uses the very same predicate in an exocentric way. In fact, this is known to both A and B and this fact is common knowledge to both agents involved in the strategic interaction described by the above scenario.²⁰ Let $\rho(s_1)$ represent the probability that the speaker A is located in the situation s_1 , that is, the probability that A intends to linguistically communicate to B , by uttering in u the sentence e_1 , the perspective-specific proposition p_1 , and consider also that $\rho(s_1) = 0.9$. Hence, the probability that A is located in the situation s_2 and that he intends to express, by uttering e_1 , the perspective-specific proposition p_2 , will be $\rho(s_2) = 1 - \rho(s_1)$, that is 0.1.²¹

At this point, the hearer's choice problem becomes more transparent. In our scenario, B has two choices: either he chooses to assign to the sentence e_1 uttered by the speaker A in u the semantic content which corresponds to the perspective-specific proposition p_1 , or he chooses to assign to the utterance of e_1 the semantic content corresponding to the perspective-specific proposition p_2 . But the optimal choice of B depends on which of s_1 and s_2 is the factual situation. B knows that if s_1 is the factual situation, then his optimal choice would be to interpret the utterance of the sentence e_1 as meaning p_1 rather than p_2 . Similarly, B knows that if s_2 is the factual situation, then his optimal choice would not be to assign to the utterance of e_1 the semantic content p_1 , but instead, it would be to interpret the utterance of the bare sentence formed with the predicate of personal taste which A uses in an

¹⁸ Parikh, "Communication, Meaning," 196; van Rooy, "Signalling Games," 499.

¹⁹ Parikh, *The Use of Language*, 27-29; Parikh, "Communication, Meaning," 196-198.

²⁰ Parikh, *The Use of Language*, 28; Parikh, "Communication, Meaning," 197.

²¹ In what follows, I will use p_1 instead of $\rho(s_1)$ and p_2 instead of $\rho(s_2)$.

exocentric way as meaning p_2 rather than p_1 . Insofar as the hearer B does not know which of s_1 and s_2 is factual, he does not know which interpretation of the utterance of e_1 is correct and, in consequence, he does not know what to choose between p_1 and p_2 , even though he knows that p_1 is, in the above described scenario, the most likely interpretation of the utterance of e_1 .

To solve this problem, the hearer B has to take into consideration the speaker's possible choices²² and to relate them to his actual choice of uttering e_1 . In this sense, B knows that A might have chosen from the fragment of natural language E containing sentences formed with predicates of personal taste alternative sentences in which the perspectival information is syntactically represented by experiencer phrases whose semantic role is to make salient the relevant perspective. Hence B knows that if s_1 is the factual situation, then A might have chosen to utter a sentence like e_2 mentioned above in order to make transparent to B the fact that he uses the predicate of personal taste in an autocentric way and that he intends to communicate the perspective-specific proposition p_1 . Similarly, B knows that if s_2 would be the factual situation, then A would make transparent to B the fact that he uses the predicate of personal taste in an exocentric way and that he intends to convey the perspective-specific proposition p_2 , only if A would choose to utter a sentence like the following:

[5] Lasagna is delicious to C .

Let e_3 abbreviate the above sentence which A might have uttered in order to explicitly signal to B that s_2 is the factual situation. In consequence, both agents A and B have to take into consideration the alternative sentence e_2 , which explicitly expresses only the perspective-specific proposition p_1 , and the alternative sentence e_3 , which explicitly expresses only the perspective-specific proposition p_2 , and to contrast these two linguistic variants with the sentence e_1 which can be used to express both p_1 and p_2 .

In order to show how to solve the two-agent cooperation game which models the strategic interaction between A and B in the above described scenario, I will use the format in which van Rooy has framed the games of partial information designed by Parikh. I will also adopt from van Rooy the assumption that the players of the game simultaneously choose strategies.²³ A strategy specifies what an agent chooses in different situations when he is involved in a strategic interaction with other agents. In van Rooy's framework, the speaker's strategy is

²² Parikh, *The Use of Language*, 30; Parikh, "Communication, Meaning," 199; van Rooy, "Signalling Games," 499.

²³ van Rooy, "Signalling Games," 500.

modeled as a function from the set of situations to the set of sentences, while the hearer's strategy is modeled as a function from the set of sentences to the set of situations.²⁴ More precisely, a speaker's strategy A_i ,²⁵ where $i = 1, 2, 3, 4$, is an element of

$$[\{s_1, s_2\} \rightarrow \{e_1, e_2, e_3\}]$$

and a hearer's strategy B_j , where $j = 1, 2$, is an element of

$$[\{e_1, e_2, e_3\} \rightarrow \{s_1, s_2\}].$$

The following two tables depict the strategies of both agents involved in the present strategic communication. The first table displays the strategies of the speaker A , while the second those of the hearer B .

	s_1	s_2
A_1	e_1	e_1
A_2	e_2	e_1
A_3	e_1	e_3
A_4	e_2	e_3

	e_1	e_2	e_3
B_1	s_1	s_1	s_2
B_2	s_2	s_1	s_2

What both agents A and B choose when they play the game which corresponds to the above described scenario, depends on what they prefer in this scenario. Insofar as successful communication is preferred to miscommunication,²⁶ the speaker and the hearer have, in this regard, the same preference. In general, one decision maker's preferences are modeled by a utility function U which assigns numerical values to his choices in conformity with the order of his

²⁴ van Rooy, "Signalling Games," 500.

²⁵ I will use, in what follows, the same letters A and B because I believe that the context makes clear when the letters have, in the economy of the text, the function to refer to agents or to their strategies.

²⁶ Parikh, *Language and Equilibrium*, 94.

preferences.²⁷ In the present two-agent game, the function U is defined over the profiles of strategies, where a profile of strategies is a pair whose first member is one of the speaker's strategies and the second member is one of the hearer's strategies. Hence, the domain of the utility function contains the elements from the Cartesian product of the set of the speaker's strategies and the set of the hearer's strategies. Assuming that the speaker and the hearer involved in the strategic interaction described by the above mentioned scenario have the same utility function, and that s is a situation from the set $\{s_1, s_2\}$, the successful communication between agents can be represented by letting U to output 1, while the miscommunication which occurs between them can be represented by letting U to output 0, as below:²⁸

$$\begin{aligned} U(s, A(s), B(A(s))) &= 1, \text{ if } B(A(s)) = s \\ &= 0 \text{ otherwise.} \end{aligned}$$

In order to solve the game which models the situation in which the speaker utters a bare sentence formed with a predicate of personal taste and the hearer has to infer the perspectival information which the speaker intends to communicate, we have to determine the expected utilities for each profile of strategies. Taking into consideration the probability distribution ρ over the situations, the expected utility of each joint strategy can be computed according to the following formula:²⁹

$$EU(A, B) = \sum_s \rho(s) \times U(s, A(s), B(A(s))).$$

The following two tables display the utilities assigned by the function U to the profiles of strategies in the situations s_1 and s_2 :

s_1	B_1	B_2
A_1	1	0
A_2	1	1
A_3	1	0
A_4	1	1

s_2	B_1	B_2
A_1	0	1
A_2	0	1
A_3	1	1
A_4	1	1

Using the above mentioned probabilities (i.e., $\rho_1 = 0.9$, $\rho_2 = 1 - \rho_1 = 0.1$) and the utilities assigned by U to all the profiles in the situations s_1 and s_2 , the expected

²⁷ Robin Clark, *Meaningful Games. Exploring Language with Game Theory* (Cambridge, Massachusetts: MIT Press, 2012), 71.

²⁸ van Rooy, "Signalling Games," 502.

²⁹ van Rooy, "Signalling Games," 501.

utilities, calculated for each profile of strategies, are displayed in the following table.

<i>EU</i>	<i>B₁</i>	<i>B₂</i>
<i>A₁</i>	0.9	0.1
<i>A₂</i>	0.9	1
<i>A₃</i>	1	0.1
<i>A₄</i>	1	1

The data from the table in which are depicted the expected utilities of all the profiles of the players' strategies can now be used to see whether there is an optimal combination of a strategy from the speaker's set of strategies and a strategy belonging to the hearer's set of strategies. If there is an optimal profile of strategies, this would constitute the solution to the game of partial information which models the strategic interaction between *A* and *B* in the above described scenario. Such a combination of a speaker strategy and a hearer strategy would be a Nash equilibrium of the present two-agent game, and this would basically mean that neither *A*, nor *B*, will benefit by changing his strategy while the other agent keeps his strategy fixed.³⁰ But a quick look at the table of expected utilities shows that the game depicted there in strategic form has multiple equilibria, or, more precisely, that the set of Nash equilibria contains the following four elements (*A₃*, *B₁*), (*A₄*, *B₁*), (*A₂*, *B₂*), (*A₄*, *B₂*). Since the utility function measures how successful the communication between *A* and *B* is, the multiple equilibria of the game show that the strategic interaction between *A*, who utters a bare sentence formed with a predicate of personal taste, and *B*, who has to decide whether *A* uses the predicate of personal taste in an autocentric or in an exocentric way, involves four cases of successful communication between them. Insofar as the structure of the game is common knowledge to both *A* and *B*, *B* knows which profiles of strategies form Nash equilibria. Since *B* knows that the game involves four optimal combinations of strategies which warrant that the agents successfully communicate one with the other, *B* does not know what is the best strategy for him to play, and, in consequence, *B* cannot decide which of the perspective-specific propositions *p₁* and *p₂* is the intended meaning of *A*'s utterance of the sentence *a₁*.

In order to solve this problem, Parikh's proposal is to fine-grain the agents' preferences. This means that the agents, beside preferring successful communication to miscommunication, will prefer, this time, not just to

³⁰ Parikh, *The Use of Language*, 38.

successfully communicate with a simple expression rather than to successfully communicate with a more complex expression, but also to successfully communicate with a complex expression rather than to miscommunicate.³¹ The previously mentioned contextual assumptions (15) and (16) encapsulate these preferences. According to the contextual assumption (15), the effort of the speaker A to produce the sentence e_2 , which makes transparent to the hearer B the fact that A autocentrically uses the predicate of personal taste, or his effort to produce the sentence e_3 , which makes transparent to B the fact that A uses in an exocentric way the predicate of personal taste, is greater than the effort of producing the sentence e_1 in which the perspectival information is not linguistically articulated. Similarly, according to the contextual assumption (16), the effort of B to process the sentence e_2 , which expresses only the perspective-specific proposition p_1 , or his effort to process the sentence e_3 , which expresses only the perspective-specific proposition p_2 , is greater than his effort to process the simpler sentence e_1 . What both contextual assumptions (15) and (16) signal is that the utility function U has to be sensitive to the fact that the costs involved in producing and processing a bare sentence formed with a predicate of personal taste, like the sentence e_1 which A uses in u to convey perspectival information to B , are inferior to the costs involved in producing and processing alternative sentences from E , like e_2 and e_3 , in which the presence of an experiencer phrase in the surface syntax makes transparent to the hearer what the speaker intends to communicate. In order to capture the agents' preferences for shorter and more economical expressions, I will follow van Rooy's proposal to define a complexity measure and to let the value of the utility function to incorporate this measure. Considering that the complexity of sentences containing predicates of personal taste can be measured by a function $\delta: E \rightarrow N$ from the set of sentences forming the fragment of language E to the set of natural numbers, the utility function will have now the following format:³²

$$U(s, A(s), B(A(s))) = 1/\delta(A(s)), \text{ if } B(A(s)) = s \\ = 0 \text{ otherwise.}$$

Assuming the following values of the function which measures the complexity of e_1 , that is, of the bare sentence with a predicate of personal taste used by A , in the above mentioned scenario, to communicate to B the perspective-specific proposition p_1 , and of its more complex alternatives e_2 and e_3 ,

$$\delta(e_1) = 1$$

³¹ Parikh, *Language and Equilibrium*, 94; Clark, *Meaningful Games*, 252; van Rooy, "Signalling Games," 502.

³² van Rooy, "Signalling Games," 502.

Mihai Hîncu

$$\delta(e_2) = 2$$

$$\delta(e_3) = 2,$$

the function U will assign to the profiles of strategies in the situations s_1 and s_2 the utilities displayed in the following tables:

s_1	B_1	B_2
A_1	1	0
A_2	0.5	0.5
A_3	1	0
A_4	0.5	0.5

s_2	B_1	B_2
A_1	0	1
A_2	0	1
A_3	0.5	0.5
A_4	0.5	0.5

The table below displays, for each profile of strategies, the expected utilities which I have calculated with the help of the above established probabilities, that is, $\rho_1 = 0.9$ and $\rho_2 = 0.1$, and of the utilities assigned by U to all the profiles of strategies in the situations s_1 and s_2 :

EU	B_1	B_2
A_1	0.9	0.1
A_2	0.45	0.55
A_3	0.95	0.05
A_4	0.5	0.5

The data from the table in which the expected utilities are depicted, show that there are now two solutions to the game of partial information that models the strategic interaction between A , who utters a bare sentence formed with a predicate of personal taste, and B , who has to decide whether A uses the predicate of personal taste in an autocentric or in an exocentric way. Now, the optimal combinations of strategies which warrant that A and B will successfully communicate one with the other are the two elements (A_3, B_1) and (A_2, B_2) of the set of Nash equilibria. What the first Nash equilibrium (A_3, B_1) basically means is that the speaker A reserves the more complex and costlier sentence e_3 for the situation s_2 , and that, given that s_1 is the factual situation, A chooses to utter the bare sentence formed with a predicate of personal taste and the hearer B chooses to interpret A 's utterance as expressing the perspective-specific proposition p_1 . According to the second Nash equilibrium (A_2, B_2) of the game, A chooses to utter, in the situation s_2 , the simplest and economical sentence e_1 , he reserves the more

complex sentence e_2 for the more probable situation s_1 , and the hearer B chooses to interpret A 's utterance as expressing the perspective-specific proposition p_2 .

In order to find a unique solution to the game of partial information which models the strategic interaction between A and B in the above described scenario, I will use Parikh's proposal to appeal to the idea of Pareto dominance as a second-order criterion.³³ The idea of Pareto dominance allows us to reduce the cardinality of the already determined set of Nash equilibria and to transform this set into a singleton. A Nash equilibrium of a two-agent game satisfies the condition of being Pareto dominant only if the expected utility it yields is at least as high as the expected utility yielded by any other Nash equilibrium of the game.³⁴ Applying the idea of Pareto dominance to the set determined above of Nash equilibria, it can be seen that the profile (A_3, B_1) has a higher expected utility than the profile (A_2, B_2) . Insofar as the first contextual assumption guarantees that both A and B are rational agents, and the last assumption (17) ensures that the structure of the game of partial information which models the strategic interaction between A and B is common knowledge, both agents will choose the profile of strategies which maximizes their expected utilities. This implies that, in the present game, both agents choose to play the strategies which form the profile (A_3, B_1) because the Nash equilibrium which corresponds to this profile Pareto dominates the Nash equilibrium corresponding to the profile of strategies (A_2, B_2) . Therefore, the unique solution of the present game of partial information which models the process of interpretation of an utterance of a bare sentence formed with a predicate of personal taste is the Pareto-Nash equilibrium (A_3, B_1) , according to which the speaker A reserves the complex sentence e_3 for the counterfactual situation s_2 , and, given that s_1 is the factual situation, he chooses to utter the simple sentence e_1 in which the perspectival information is not linguistically articulated by an experiencer phrase, while the hearer B chooses to interpret A 's utterance as expressing the perspective-specific proposition p_1 which corresponds to A 's autocentric use of the predicate of personal taste occurring in e_1 .

In more general terms, the present game-theoretical account of strategic communication with expressions referring to agents' perspectives predicts that if the situation in which the speaker intends to convey to the hearer perspectival information about himself is factual, then the unique Pareto-Nash equilibrium of the game modeling the situation will correspond to the balance between the speaker's choice to utter a bare sentence formed with a predicate of personal taste and the hearer's choice to interpret the utterance as expressing the perspective-

³³ Parikh, *The Use of Language*, 39; Parikh, *Language and Equilibrium*, 114.

³⁴ Parikh, "Communication, Meaning," 205; Clark, *Meaningful Games*, 92.

specific proposition which corresponds to the speaker's autocentric use of the predicate of personal taste. Similarly, the model predicts that if the situation in which the speaker intends to convey to the hearer perspectival information about another agent is factual, then the unique Pareto-Nash equilibrium of the game modeling this situation will correspond to the balance between the speaker's choice to utter a bare sentence formed with a predicate of personal taste and the hearer's choice to interpret the utterance as expressing the perspective-specific proposition which corresponds to the speaker's exocentric use of the predicate of personal taste.

4. Conclusion

I have focused, in this paper, on bare sentences formed with predicates of personal taste. These sentences are used in communication to express perspectival information even though this information is not linguistically articulated by the occurrences of experiencer phrases at the level of sentences' surface syntax. In order to list some of the semantic properties exemplified by the utterances of bare sentences formed with predicates of personal taste, I have compared them with utterances of sentences in which an indexical like "I" occurs. In this sense, I have shown that both a sentence formed with an indexical corresponding to the first person singular pronoun and a bare sentence formed with a predicate of personal taste, express, in different contexts, different propositions, and have, in consequence, different truth-values. Their context-sensitive profile, which explains the variation of the semantic content expressed, and also the variation in truth-value of both types of sentences, is due to the occurrence, at the level of their logical forms, of a variable which represents the agent who utters the first person singular pronoun and, in the case of sentences with predicates of personal taste, the perspective of an agent. Insofar as it is semantically required that the value contextually assigned to the variable present in the logical form of a sentence in which the indexical "I" occurs, has to be the agent who utters the indexical, the semantics of sentences of this kind is speaker-oriented. In contrast, the semantics of a bare sentence with a predicate of personal taste is not necessarily speaker-oriented: the value contextually assigned to the variable present in its logical form can correspond to the perspective of the agent who utters the sentence, but also to the perspective of another agent. In this sense, I introduced the distinction between the autocentric and exocentric interpretations of predicates of personal taste, interpretations which correspond to the different values the context of utterance provides to the variable for perspective occurring in the logical forms of bare sentences formed with predicates of this kind. The

situations in which an agent, who intends to communicate perspectival information to another agent by uttering a sentence in which this information is not linguistically articulated by an experiencer phrase, uses in one way the predicate of personal taste occurring in the sentence, while the other agent interprets it in another way, constitute instances of a more general case in which the agents' acts are not coordinated. In this sense, I have tried to answer the question of how a hearer can coordinate with a speaker on the intended perspectival information conveyed by the utterance of a bare sentence formed with a predicate of personal taste, and I have showed that there is a systematic way in which both agents can optimally solve their coordination problem and rationally avoid problems due to miscommunication. In this regard, I have proposed, in the present paper, a game-theoretical account of the strategic interaction between a speaker who intends to convey some perspectival information and who chooses, in this sense, to utter a bare sentence formed with a predicate of personal taste, instead of a sentence in which the perspectival information is linguistically articulated by an experiencer phrase, and a hearer who has to choose between interpreting the uttered sentence in conformity with the speaker's autocentric use of the predicate of personal taste or in conformity with the speaker's exocentric use of the predicate. The present game-theoretical account predicts that, in the situations in which the speaker intends to talk about his perspective and utters, in this sense, a bare sentence formed with a predicate of personal taste, the solution to the game which models situations of this kind is a unique Pareto-Nash equilibrium according to which the speaker does not linguistically articulate the perspectival information by an experiencer phrase and the hearer interprets the utterance as expressing the perspective-specific proposition which corresponds to the speaker's autocentric use of the predicate of personal taste.³⁵

³⁵ This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU 159/1.5/S/133675.

WHY GETTIER CASES ARE MISLEADING

Moti MIZRAHI

ABSTRACT: In this paper, I argue that, as far as Gettier cases are concerned, appearances are deceiving. That is, Gettier cases merely appear to be cases of epistemic failure (i.e., failing to know that p) but are in fact cases of semantic failure (i.e., failing to refer to x). Gettier cases are cases of reference failure because the candidates for knowledge in these cases contain ambiguous designators. If this is correct, then we may simply be mistaking semantic facts for epistemic facts when we consider Gettier cases. This, in turn, is a good reason not to assign much, if any, evidential weight to Gettier intuitions (i.e., that S doesn't know that p in a Gettier case).

KEYWORDS: ambiguous designator, analysis of knowledge, Gettier cases, justified true belief, semantic reference, speaker's reference

1. Introduction

It is often said that “the JTB analysis was refuted by Edmund Gettier.”¹ In his seminal paper “Is Justified True Belief Knowledge?” Gettier presents counterexamples to the Justified True Belief (JTB) analysis of knowledge, according to which S knows that p if and only if p is true, S believes that p , and S is justified in believing that p .² Gettier's argument against JTB can be summed up as follows:

G1. If knowledge is JTB, then S knows that p in a Gettier case.

G2. S doesn't know that p in a Gettier case.

Therefore,

G3. It is not the case that knowledge is JTB.

Premise G2 is based on what appears to be the case in Gettier cases. That is, in a Gettier case, it seems (at least to some) that S doesn't know that p . Then, the content of this intellectual appearance, namely, $\langle S$ doesn't know that p in a Gettier case \rangle , is used as a premise in an argument, like the aforementioned *modus*

¹ Timothy Williamson, “Knowledge First Epistemology,” in *The Routledge Companion to Epistemology*, eds. Sven Bernecker and Duncan Pritchard (New York: Routledge, 2011), 209.

² Edmund L. Gettier, “Is Justified True Belief Knowledge?” *Analysis* 23 (1963): 121-123.

tollens argument from premises G1 and G2 to conclusion G3, which is taken by many epistemologists to amount to a refutation of JTB.

According to Weatherson, however, “maybe respecting the Gettier intuitions was the wrong reaction, we should instead have been explaining why we are all so easily misled by these kinds of cases.”³ This is what I aim to do in this paper. That is, in what follows, I will try to explain why Gettier cases are misleading. I will argue that, contrary to appearances, Gettier cases are actually cases of semantic, not epistemic, failure. This is because Gettier cases involve what Kripke calls “ambiguous designators.” If this is correct, then we should not respect Gettier intuitions. That is, there is a good reason not to assign much, if any, evidential weight to Gettier intuitions.⁴ To be clear, I will not be defending the JTB analysis of knowledge against Gettier and Gettier-style cases. Nor will I be offering a “solution” to the so-called Gettier problem. Rather, I will show why Gettier and Gettier-style cases are misleading. Since I think that Gettier cases are misleading, I don’t think we are in a position to say whether the subjects in Gettier cases have knowledge or not, whether the beliefs are true or not, whether the beliefs are justified or not, and so on.

In what follows, then, I will explain “why we are all so easily misled by these kinds of cases [namely, Gettier and Gettier-style cases].”⁵ I will proceed by considering five Gettier and Gettier-style cases. The first two are Gettier’s original cases, namely, Case I and Case II. The third is the sheep-in-the-meadow case (due to Chisholm). The fourth is the Fake Barn case (due to Ginet). The fifth is the Stopped Clock case (due to Russell). Unlike Gettier’s original cases, the other Gettier-style cases are supposed to be cases in which the candidate for knowledge is not inferred from any falsehoods.⁶ I will argue that all these cases are misleading

³ Brian Weatherson, “What Good Are Counterexamples?” *Philosophical Studies* 115 (2003): 1-31.

⁴ Another challenge to the alleged evidential status of Gettier intuitions comes from experimental philosophy. On this debate, see the following: Jennifer Nagel, “Intuitions and Experiments: A Defense of the Case Method in Epistemology,” *Philosophy and Phenomenological Research* 85 (2012): 495-527. Stephen Stich, “Do Different Groups Have Different Epistemic Intuitions? A Reply to Jennifer Nagel,” *Philosophy and Phenomenological Research* 87 (2013): 151-178. Jennifer Nagel, “Defending the evidential value of epistemic intuitions: a reply to Stich,” *Philosophy and Phenomenological Research* 87 (2013): 179-199. The main argument of this paper does not depend on experimental results concerning Gettier intuitions.

⁵ Weatherson, “What Good,” 1.

⁶ For the “no false lemmas” response to Gettier cases, see David M. Armstrong, *Belief, Truth, and Knowledge* (New York: Cambridge University Press, 1973), 152 and Michael Clark, “Knowledge and Grounds: A Comment on Mr. Gettier’s paper,” *Analysis* 24 (1963): 46-48. See also Robert K.

insofar as they merely appear to be cases of epistemic failure (i.e., failing to know that p), when in fact they are cases of semantic failure (i.e., failing to refer to x). Gettier cases are cases of reference failure because the candidates for knowledge in these cases contain ambiguous designators. If this is correct, then we may simply be mistaking semantic facts for epistemic facts when we consider Gettier cases. This, in turn, is a good reason not to assign much, if any, evidential weight to Gettier intuitions (i.e., that S doesn't know that p in a Gettier case).

2. Gettier's Original Cases

In Gettier's first case, the candidate for knowledge, which Smith infers from "Jones is the man who will get the job, and Jones has ten coins in his pocket,"⁷ is the following proposition:

(I) The man who will get the job has ten coins in his pocket.

Since Smith gets the job and has ten coins in his pocket, (I) is true. Smith believes (I) and is justified in believing (I), insofar as he has some evidence for (I), so all the conditions of the JTB analysis of knowledge are supposedly met. But Smith does not know that (I) is true, or so it seems to many philosophers and non-philosophers.⁸

This case is misleading, however, because it merely appears to be a case of epistemic failure (i.e., failing to know that p) but is in fact a case of semantic failure (i.e., failing to refer to x). To see why, notice that, when Smith believes that the man who will get the job has ten coins in his pocket,⁹ the "coins" that Smith wishes to talk about are *not* the coins that are actually in his pocket. In other words, 'coins' in Gettier's Case I is what Kripke calls an "ambiguous designator."¹⁰

Shope, *The Analysis of Knowing: A Decade of Research* (Princeton, NJ: Princeton University Press, 1983), 24 and the Appendix in John L. Pollock, *Contemporary Theories of Knowledge* (Totowa: Rowman and Littlefield, 1986). Cf. Michael Levin, "Gettier Cases Without False Lemmas," *Erkenntnis* 64 (2006): 381-392.

⁷ Gettier, "Is Justified True Belief," 122.

⁸ For a recent study on Gettier intuitions among non-philosophers, see Edouard Machery, et al, "Gettier across cultures," *Noûs*. DOI: 10.1111/nous.12110.

⁹ Gettier, "Is Justified True Belief," 122.

¹⁰ Saul Kripke, "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy* 2 (1977): 255-276.

Along the lines of Grice's distinction between what a speaker's words mean and what the speaker means in saying these words,¹¹ Kripke draws a distinction between speaker's reference and semantic reference. To illustrate the distinction, Kripke gives the following example:

Two people see Smith in the distance and mistake him for Jones. They have a brief colloquy: "What is Jones doing?" "Raking the leaves." "Jones," in the common language of both, is a name of Jones; it *never* names Smith. Yet, in some sense, on this occasion, clearly both participants in the dialogue have referred to Smith, and the second participant has said something true about the man he referred to if and only if Smith was raking the leaves (whether or not Jones was).¹²

According to Kripke, then, "the speaker's referent of a designator [is] that object which the speaker wishes to talk about, on a given occasion, and believes fulfills the conditions for being the semantic referent of the designator."¹³ In Kripke's example, the speaker's referent of 'Jones' is Smith, whereas the semantic referent of 'Jones' is Jones. In Gettier's Case I, the speaker's referent of 'coins' is the ten coins that are in Jones' pocket, whereas the semantic referent of 'coins' is the ten coins that are in Smith's pocket. For this reason, 'coins' is an ambiguous designator in Gettier's Case I.¹⁴

If this is correct, then it is not clear that, by using 'coins', Smith manages to successfully refer to the ten coins that fulfill the conditions for being the semantic referent of 'coins', i.e., the ten coins that make (I) true. After all, Smith wishes to talk about the ten coins that are in Jones' pocket. But the ten coins that are in Jones' pocket do not fulfill the conditions for being the semantic referent of 'coins' in Case I, i.e., the referent of 'coins' that makes (I) true, since the "coins" that Smith wishes to talk about are *not* the coins that are in the pocket of the man who will get the job. If this is correct, then Gettier's Case I is a case of reference failure, which is a semantic failure, not an epistemic failure, and hence not knowledge failure.¹⁵

¹¹ H. P. Grice, "Utterer's Meaning and Intentions," *The Philosophical Review* 78 (1969): 147-177.

¹² Kripke, "Speaker's Reference," 263.

¹³ Kripke, "Speaker's Reference," 264.

¹⁴ In Gettier's Case I, 'the man' is an ambiguous designator as well. Cf. Christoph Schmidt-Petri, "Is Gettier's First Example Flawed?" *Knowledge and Belief*, eds. W. Löffler and P. Weingartner (ALWS, 2003), 317-319.

¹⁵ Cf. Adrian Heathcote, "Truthmaking and the Gettier Problem," in *Aspects of Knowing: Epistemological Essays*, ed. Stephen Hetherington (Amsterdam: Elsevier, 2006), 151-168 on a different sort of semantic ambiguity in Case I. Unlike Heathcote, I am not trying to offer a

What makes Gettier's Case I misleading, then, is the presence of the ambiguous designator 'coins.' Given that 'coins' is an ambiguous designator, it is not clear that, by using 'coins,' Smith manages to successfully refer to what fulfills the conditions for being the semantic referent of 'coins,' which is different from what Smith wishes to talk about. This means that, upon considering Gettier's Case I, we may be confusing the fact that Smith fails to refer to what actually fulfills the conditions for being the semantic referent of 'coins,' which is a semantic fact about the case, with an epistemic fact, namely, that Smith doesn't know that (I) is the case.

In Gettier's second case, the candidate for knowledge, which Smith infers from "(g) Either Jones owns a Ford, or Brown is in Boston,"¹⁶ is the following proposition:

(h) Either Jones owns a Ford, or Brown is in Barcelona.

Since the second disjunct of (h) happens to be true, (h) is true. Smith believes (h) and is justified in believing (h), insofar as he has some evidence for "(f) Jones owns a Ford," from which Smith gets (g) and then infers (h), so all the conditions of the JTB analysis of knowledge are supposedly met. But Smith does not know that (h) is true, or so it seems to many philosophers and non-philosophers.

It is important to note that, unlike Case I, Case II involves two inferences. The first inference is from

(f) Jones owns a Ford.

to

(g) Either Jones owns a Ford, or Brown is in Boston.

The second inference is from (g) to

(h) Either Jones owns a Ford, or Brown is in Barcelona.

Smith's evidence for (f) is that "Jones has at all times in the past within Smith's memory owned a car, and always a Ford, and that Jones has just offered Smith a ride while driving a Ford."¹⁷ In using 'Jones,' then, Smith wishes to talk about the person who offered Smith a ride, has always owned a Ford, etc. But the person who offered Smith a ride, has always owned a Ford, etc. does not fulfill the

"solution" to the so-called Gettier problem, since I think the cases are misleading, as I try to show in this paper.

¹⁶ Gettier, "Is justified true belief," 122.

¹⁷ Gettier, "Is justified true belief," 122.

conditions for being the semantic referent of 'Jones' in Case II, i.e., the referent of 'Jones' that makes (g) true, since, by stipulation the person that Smith wishes to talk about by using 'Jones' does not own a Ford. In Gettier's Case II, then, the speaker's referent of 'Jones' is the person who offered Smith a ride, has always owned a Ford, etc., whereas the semantic referent of 'Jones,' i.e., the referent of 'Jones' that makes (g) true, cannot be that person, since Jones does not own a Ford, by stipulation. For this reason, 'Jones' is an ambiguous designator in Gettier's Case II. If this is correct, then, like Case I, Case II is a case of reference failure, which is a semantic failure, not an epistemic failure, and hence not knowledge failure.

What makes Gettier's second case misleading, then, is the presence of the ambiguous designator 'Jones.' Given that 'Jones' is an ambiguous designator, it is not clear that, by using 'Jones,' Smith manages to successfully refer to what fulfills the conditions for being the semantic referent of 'Jones,' which is different from what Smith wishes to talk about. This means that, upon considering Gettier's second case, we may be confusing the fact that Smith fails to refer to what actually fulfills the conditions for being the semantic referent of 'Jones,' which is a semantic fact about the case, with an epistemic fact, namely, that Smith doesn't know that (h) is the case.

3. The Sheep-in-the-Meadow Case

The sheep-in-the-meadow case¹⁸ is supposed to be a Gettier-style case without false lemmas (i.e., without inferences from falsehoods):¹⁹

It's a bright sunny day; I'm out in the country; and it looks to me like there's a sheep in a certain meadow. [...] So I believe that there's a sheep in the meadow. And let's suppose that there *is* a sheep in the meadow, so that my belief is true. [...] So I have a justified true belief that there's a sheep in the meadow. Do I *know* that there's a sheep in the meadow?²⁰

¹⁸ Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice Hall, 1966).

¹⁹ An earlier so-called Gettier-style case without false lemmas can be found in Gilbert Harman, *Thought* (Princeton, NJ: Princeton University Press Harman, 1973), 75. Cf. William G. Lycan, "On the Gettier Problem Problem," in *Epistemology Futures*, ed. S. Hetherington (New York: Oxford University Press, 2006), 148-168. Lycan defends JTB with the addition of the "no false lemmas" condition. Another epistemologist who defends the tripartite analysis of knowledge is Stephen Hetherington, *Good Knowledge, Bad Knowledge: On Two Dogmas of Epistemology* (New York: Oxford University Press, 2001). Cf. Anthony R. Booth, "The Gettier Illusion, the Tripartite Analysis, and the Divorce Thesis," *Erkenntnis* 79 (2014): 625-638.

²⁰ Jay F. Rosenberg, *Three Conversations about Knowing* (Indianapolis: Hackett Publishing Co., 2000), 30.

Since there is a rock that looks like a sheep in the meadow, and there is an actual sheep behind the rock that looks like a sheep, it seems that the subject *S* in the sheep-in-the-meadow case does not know that there's a sheep in the meadow because *S*'s belief that there's a sheep in the meadow is accidentally true or true as a matter of epistemic luck.²¹

However, I think that the sheep-in-the-meadow case is misleading in much the same way that Gettier's original cases are misleading. Like 'coins' in Case I, and 'Jones' in Case II, 'sheep' is an ambiguous designator in the sheep-in-the-meadow case. In terms of semantic reference, 'sheep' designates the actual sheep that makes <there's a sheep in the meadow> true. In terms of speaker's reference, 'sheep' designates what *S* wishes to talk about, which is a rock that to *S* looks like a sheep, not the actual sheep that makes <there's a sheep in the meadow> true. If this is correct, then it is not clear that, by using 'sheep', *S* manages to successfully refer to the sheep that makes <there's a sheep in the meadow> true. After all, *S* wishes to talk about the "sheep" that *S* sees. But the "sheep" that *S* sees does not fulfill the conditions for being the semantic referent of 'sheep,' i.e., the referent of 'sheep' that makes <there's a sheep in the meadow> true, given that it is a rock that to *S* looks like a sheep. If this is correct, then the sheep-in-the-meadow case is a case of reference failure, which is a semantic failure, not an epistemic failure, and hence not knowledge failure.

What makes the sheep-in-the-meadow case misleading, then, is the presence of the ambiguous designator 'sheep'. Given that 'sheep' is an ambiguous designator, it is not clear that, by using 'sheep', *S* manages to successfully refer to what fulfills the conditions for being the semantic referent of 'sheep', which is different from what *S* wishes to talk about. This means that, upon considering the sheep-in-the-meadow case, we may be confusing the fact that *S* fails to refer to what actually fulfills the conditions for being the semantic referent of 'sheep', which is a semantic fact about the case, with an epistemic fact, namely, that *S* doesn't know that there's a sheep in the meadow.

4. The Fake Barn Case

The same diagnosis, I submit, applies to other so-called Gettier-style cases without false lemmas. Consider Goldman's²² Fake Barn case (due to Carl Ginet):²³

²¹ Duncan Pritchard, *Epistemic Luck* (New York: Oxford University Press, 2005), 145-177.

²² Alvin I. Goldman, "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* 73 (1976): 771-791.

²³ Cf. Jason Stanley, *Knowledge and Practical Interests* (New York: Oxford University Press, 2005), 114.

Henry is driving in the countryside and sees a barn ahead in clear view. On this basis he believes that the object he sees is a barn. Unknown to Henry, however, the area is dotted with barn facades that are indistinguishable from real barn from the road. However, Henry happens to be looking at the one real barn in the area.²⁴

In the Fake Barn case, it seems that the JTB conditions are all met. Moreover, Henry's belief that there is a barn over there is not inferred from any falsehoods, and yet Henry does not know that there is a barn over there, or so it seems.

But the Fake Barn case is misleading as well. Like 'coins' in Case I, 'Jones' in Case II, and 'sheep' in the sheep-in-the-meadow case, 'barn' is an ambiguous designator in the Fake Barn case. In terms of semantic reference, 'barn' designates the one real barn in Barn County that makes <there's a barn over there> true. In terms of speaker's reference, 'barn' designates what *S* wishes to talk about. In that case, however, it is not clear that, by using 'barn', *S* manages to successfully refer to the one real barn that makes <there's a barn over there> true. After all, *S* could have easily referred to a mere barn-façade by using 'barn', since Barn County is peppered with barn-façades. In terms of speaker's reference, then, 'barn' refers to what to *S* looks like a barn, which could have easily been a barn façade, not the one real barn that makes <there's a barn over there> true.²⁵ If this is correct, then, like Gettier's Case I and the sheep-in-the-meadow case, the Fake Barn case is a case of reference failure, which is a semantic failure, not an epistemic failure, and hence not knowledge failure.

What makes the Fake Barn case misleading, then, is the presence of the ambiguous designator 'barn'. Given that 'barn' is an ambiguous designator, it is not clear that, by using 'barn', *S* manages to successfully refer to what fulfills the conditions for being the semantic referent of 'barn', which is different from what *S* wishes to talk about. This means that, upon considering the Fake Barn case, we may be confusing the fact that *S* fails to refer to what actually fulfills the conditions for being the semantic referent of 'barn', which is a semantic fact about the case, with an epistemic fact, namely, that *S* doesn't know that there's a barn over there.

²⁴ John Greco, *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity* (New York: Cambridge University Press, 2010), 76.

²⁵ For more on accidentality in Gettier-style cases, see Masahiro Yamada, "Getting it Right by Accident," *Philosophy and Phenomenological Research* 83 (2011): 72-105 and Karl Schafer, "Knowledge and Two Forms of Non-Accidental Truth," *Philosophy and Phenomenological Research* 89 (2014): 373-393.

5. The Stopped Clock Case

Some might think that, in Russell's Stopped Clock case,²⁶ there are no ambiguous designators. In the Stopped Clock case the candidate for knowledge is, say, that the time is five o'clock. Since the clock stopped exactly twelve hours ago, and the time is in fact five o'clock, *S* doesn't know that the time is five o'clock, or so it might seem to many.²⁷

Like the Gettier and Gettier-style cases discussed so far, however, I think that the Stopped Clock case also involves an ambiguous designator. To see why, note that, like 'sheep' in the sheep-in-the-meadow case, 'barn' in the Fake Barn case, 'coins' in Case I, and 'Jones' in Case II, 'the time' is an ambiguous designator in the Stopped Clock case. In terms of semantic reference, 'the time' designates the standard time in the time zone where it is currently five o'clock. In terms of speaker's reference, 'the time' designates what *S* wishes to talk about, which is a reading from the stopped clock. In that case, however, it is not clear that, by using 'the time', *S* manages to successfully refer to the standard time in the time zone where it is currently five o'clock, since *S* uses 'the time' to talk about something (namely, a reading from a stopped clock) that does not in fact fulfill the conditions for being the semantic referent of 'the time', given that the clock is not working properly.²⁸ After all, a clock is an instrument that indicates local time in a given time zone. Since the time indicator in the Stopped Clock case is a faulty one by stipulation, any given reading from this faulty time indicator does not fulfill the conditions for being the semantic referent of 'the time'.²⁹ If this is correct, then, like Gettier's Cases I and II, the sheep-in-the-meadow case, and the Fake Barn

²⁶ Bertrand Russell, *Human Knowledge: Its Scope and Its Limits* (London: George Allen & Unwin, 1948), 170-171.

²⁷ The Stopped Clock case rests on the assumption that a stopped clock shows the right time twice a day. For present purposes, I will not challenge this assumption although some do. See, e.g., Adrian Heathcote, "Truthmaking, Evidence of, and Impossibility Proofs," *Acta Analytica* 29 (2014): 363-375.

²⁸ Cf. Shope, *The Analysis of Knowing*, 20.

²⁹ Adrian Heathcote, "Gettier and the Stopped Clock," *Analysis* 72 (2012): 309-314 offers an explanation in terms of truthmakers for why *S* does not know that *p* (e.g., that the time is five o'clock) in Russell's Stopped Clock case. I remain noncommittal about whether *S* knows that *p* or not in the Stopped Clock case precisely because I think it is misleading. If I am right, then the Stopped Clock case is misleading in the same way that other Gettier and Gettier-style cases are misleading. That is, it merely appears to be a case of knowledge failure when in fact it is a case of reference failure. Since reference failure is a semantic, not an epistemic, failure, the Gettier intuition that *S* doesn't know that *p* in a Gettier case should not be respected.

case, the Stopped Clock case is a case of reference failure, which is a semantic failure, not an epistemic failure, and hence not knowledge failure.

What makes the Stopped Clock case misleading, then, is the presence of the ambiguous designator ‘the time’. Given that ‘the time’ is an ambiguous designator, it is not clear that, by using ‘the time’, *S* manages to successfully refer to what fulfills the conditions for being the semantic referent of ‘the time’, which is different from what *S* wishes to talk about. This means that, upon considering the Stopped Clock case, we may be confusing the fact that *S* fails to refer to what actually fulfills the conditions for being the semantic referent of ‘the time’, which is a semantic fact about the case, with an epistemic fact, namely, that *S* doesn’t know that the time is five o’clock.

6. Semantic Failure vs. Epistemic Failure

I have argued that Gettier and Gettier-style (without false lemmas) cases merely appear to be cases of epistemic failure (i.e., failing to know that *p*) but are in fact cases of semantic failure (i.e., failing to refer to *x*). This is because Gettier cases involve what Kripke calls “ambiguous designators.” But failing to refer is a semantic failure, not an epistemic failure, like failing to know that *p*.

To illustrate the difference between semantic failure (i.e., failing to refer to *x*) and epistemic failure (i.e., failing to know that *p*), suppose I believe that this table is made of matter. By ‘matter’, however, I do not mean atoms that are made of subatomic particles. Rather, I use ‘matter’ to talk about green cheese. And I believe that everything in the universe, including this table, is made of green cheese. In that case, when I believe that this table is made of matter, I actually believe that this table is made of green cheese, since I use ‘matter’ to refer to green cheese. If I were to use ‘matter’ to refer to what fulfills the conditions for being the semantic referent of ‘matter’, i.e., if I were to use ‘matter’ to refer to atoms, then perhaps I would know that this table is made of matter. But I use ‘matter’ to refer to green cheese, not atoms, and so my failure is semantic (i.e., failing to refer to what fulfills the conditions for being the semantic referent of ‘matter’), not epistemic (i.e., failing to know that this table is made of matter).

Similarly, if Smith were to use ‘coins’ to refer to the ten coins that make (I) true, i.e., the ten coins in his pocket, then perhaps Smith would know that the man who will get the job has ten coins in his pocket. But Smith uses ‘coins’ to refer to the coins in Jones’ pocket, not the coins in his pocket, and so his failure is semantic (i.e., a failure to refer to what fulfills the conditions for being the semantic referent of ‘coins’), not epistemic (i.e., a failure to know that the man who will get the job has ten coins in his pocket).

If Smith were to use 'Jones' to refer to the person that make (g) true, i.e., the person who owns a Ford, then perhaps Smith would know that Jones owns a Ford or Brown is in Barcelona. But Smith uses 'Jones' to refer to the person who offered Smith a ride, has always owned a Ford, etc., who doesn't own a Ford (by stipulation), and so Smith's failure is semantic (i.e., a failure to refer to what fulfills the conditions for being the semantic referent of 'Jones'), not epistemic (i.e., a failure to know that Jones owns a Ford or Brown is in Barcelona).

If *S* were to use 'sheep' to refer to the actual sheep that makes <there's a sheep in the meadow> true, then perhaps *S* would know that there's a sheep in the meadow. But *S* uses 'sheep' to refer to something that does not fulfill the semantic conditions for being the semantic referent of 'sheep' (namely, a rock that looks like a sheep to *S*), and so *S*'s failure is semantic (i.e., a failure to refer to what fulfills the conditions for being the semantic referent of 'sheep'), not epistemic (i.e., a failure to know that there's a sheep in the meadow).

If *S* were to use 'barn' to refer to the one real barn that makes <there's a barn over there> true, then perhaps *S* would know that there's a barn over there. But *S* uses 'barn' to refer to something that could have easily failed to fulfill the semantic conditions for being the semantic referent of 'barn' (namely, a mere barn-façade, given that Barn County is peppered with barn-façades), and so, if *S* fails, *S*'s failure is semantic (i.e., a failure to refer to what fulfills the conditions for being the semantic referent of 'barn'), not epistemic (i.e., a failure to know that there's a barn over there).

Finally, if *S* were to use 'the time' to refer the standard time in the time zone where it is currently five o'clock, then perhaps *S* would know that the time is five o'clock. But *S* uses 'the time' to refer to something that does not fulfill the semantic conditions for being the semantic referent of 'the time' (namely, a reading from a broken time indicator), and so *S*'s failure is semantic (i.e., a failure to refer to what fulfills the conditions for being the semantic referent of 'the time'), not epistemic (i.e., a failure to know that the time is five o'clock).

If this is correct, then Gettier cases are misleading because they merely appear to be cases of epistemic failure (i.e., failing to know that *p*), when in fact they are cases of semantic failure (i.e., failing to refer). Gettier cases are cases of reference failure because the candidates for knowledge in these cases contain ambiguous designators. If this is correct, then we may simply be mistaking semantic facts for epistemic facts when we consider Gettier cases. This, in turn, is a good reason not to assign much, if any, evidential weight to Gettier intuitions (i.e., that *S* doesn't know that *p* in a Gettier case). That is, there is a good reason to think that Gettier cases are misleading in a way that makes us mistake semantic

facts for epistemic facts. If this is correct, then we should not assign much, if any, evidential weight to Gettier intuitions.

7. Objections and Replies

I have argued that Gettier cases are misleading because the candidates for knowledge in these cases contain ambiguous designators. In this section, I will consider a couple of objections to my overall argument. The first objection goes like this. In evaluating Gettier cases, we are concerned with beliefs, not their verbal expressions, and so the notion of reference (either speaker's reference or semantic reference) does not apply to evaluating such cases.

In reply, I grant that it is not necessary that subjects make any claims or express their thoughts as far as the evaluation of Gettier cases is concerned. However, I think that the notion of reference is still relevant. After all, knowledge itself is supposed by many to either entail a mental state (such as belief) or be a mental state. According to Williamson, for instance, "knowing is the most general factive stative attitude" and the "characteristic expression of a factive stative attitude in language is a factive mental state operator (FMSO)."³⁰ For this reason, whether we think of subjects in Gettier cases as making claims or as having thoughts, which are supposed to express propositions, those propositions are supposed to be *about* something (e.g., coins, sheep, barns, etc.), which is why the notion of reference is relevant here.³¹

The second objection goes like this. Even if the candidates for knowledge in Gettier cases contain ambiguous designators, the epistemic facts about those cases are still clear. For example, in the sheep-in-the-meadow case, it is clear that *S* believes that there's a sheep in the meadow but doesn't know that there's a sheep in the meadow.

In reply, I would say that this is a little too quick and simple. For, if the candidates for knowledge in Gettier cases contain ambiguous designators, as I have argued, then that means that the relevant beliefs are ambiguous between two

³⁰ Timothy Williamson, *Knowledge and Its Limits* (New York: Oxford University Press, 2000),

34. See also Jessica Brown and Mikkel Gerken, "Knowledge Ascriptions: Their Semantics, Cognitive Bases, and Social Functions," in *Knowledge Ascriptions*, eds. Jessica Brown and Mikkel Gerken (New York: Oxford University Press, 2012), 1-30.

³¹ Some readers may wish to invoke the Language of Thought Hypothesis (LOTH) here. According to LOTH, "thought and thinking are done in a mental language, i.e., in a symbolic system physically realized in the brain of the relevant organisms" (Murat Aydede, "The Language of Thought Hypothesis," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Fall 2015 Edition) URL = <http://plato.stanford.edu/archives/fall2015/entries/language-thought>). Of course, a symbol is supposed to stand for (or refer to) something.

interpretations: an “objective” interpretation in terms of the conditions that make the belief true (i.e., in terms of semantic reference or what a speaker’s words mean) and a “subjective” interpretation in terms of what *S* means (i.e., in terms of speaker’s reference or what a speaker means in uttering certain words). In the sheep-in-the-meadow case, for example, the belief that there’s a sheep in the meadow is ambiguous between these two interpretations:

Objective interpretation (semantic reference): the semantic referent of ‘sheep’ in <there’s a sheep in the meadow> is the actual sheep that makes <there’s a sheep in the meadow> true; otherwise, <there’s a sheep in the meadow> would not be true.

Subjective interpretation (speaker’s reference): the speaker’s referent of ‘sheep’ in <there’s a sheep in the meadow> is what *S* sees, which is the rock that looks like a sheep, not what *S* doesn’t see, which is the actual sheep that makes <there’s a sheep in the meadow> true.

As I have argued above, interpreted “objectively,” or in terms of what the words mean, <there’s a sheep in the meadow> is not what *S* actually believes, since *S* uses ‘sheep’ to talk about what *S* sees, not what *S* doesn’t see. Interpreted “subjectively,” or in terms of what *S* means by uttering these words, <there’s a sheep in the meadow> is strictly false, since *S* uses ‘sheep’ to talk about something that does not in fact fulfill the conditions for being the semantic referent of ‘sheep.’

We can see this ambiguity in Kripke’s case as well. It might seem as if the epistemic facts of Kripke’s case are clear: the two people believe that Jones is raking the leaves but they don’t know that Jones is raking the leaves. However, I submit that the epistemic facts of the case are not as clear as they might seem precisely because ‘Jones’ is an ambiguous designator in this case. The people who mistake Smith for Jones wish to talk about Jones, and so they use ‘Jones’. Their belief that Jones is raking the leaves is thus ambiguous between two interpretations:

1. *Semantic reference*: Jones (= Smith) is raking the leaves.
2. *Speaker’s reference*: Jones (= Jones) is raking the leaves.

By stipulation, (2) is false, since the people in the case mistake Smith for Jones and Jones is not in fact raking the leaves. On (2), then, the two people in Kripke’s case simply have a false belief. On the other hand, (1) is not actually what the people in the case believe, since they wish to talk about Jones and they use ‘Jones’ to talk about what they see, which is Smith raking the leaves. To put it

crudely, on (1), what goes on in their heads does not match the facts of the case. Given this ambiguity, then, the case, like Gettier cases in general, is misleading.

8. Conclusion

My aim in this paper has been to remove “some of the rubbish that lies in the way to knowledge.”³² The “rubbish” I seek to remove is so-called Gettier intuitions elicited from Gettier cases. I have argued that Gettier cases are misleading insofar as they merely appear to be cases of epistemic failure (i.e., failing to know that p) but are in fact cases of semantic failure (i.e., failing to refer to x). Gettier cases are cases of reference failure because the candidates for knowledge in these cases contain ambiguous designators. If this is correct, then, because of this ambiguity, we may simply be mistaking semantic facts for epistemic facts when we consider Gettier cases. This, in turn, is a good reason not to assign much, if any, evidential weight to Gettier intuitions (i.e., that S doesn’t know that p in a Gettier case).

³² John Locke, “Epistle to the Reader,” in *An Essay Concerning Human Understanding* (London, 1689).

BEING SURE AND BEING CONFIDENT THAT YOU WON'T LOSE CONFIDENCE

Alexander R. PRUSS

ABSTRACT: There is an important sense in which one can be sure without being certain, i.e., without assigning unit probability. I will offer an explication of this sense of sureness, connecting it with the level of credence that a rational agent would need to have to be confident that she won't ever lose her confidence. A simple formal result then gives us an explicit formula connecting the threshold α for credence needed for confidence with the threshold needed for being sure: one needs $1-(1-\alpha)^2$ to be sure. I then suggest that stepping between α and $1-(1-\alpha)^2$ gives a procedure that generates an interesting hierarchy of credential thresholds.

KEYWORDS: credence, belief, moral certainty, certainty, sureness, martingale, closure of inquiry

1. Introduction

There are some things I am sure of. I am sure I have two hands and that the world is billions of years old. Yet I assign a probability less than one to these propositions. There is some small chance that I am currently in the hospital after the amputation of one of my hands and am dreaming, and likewise there is some small chance that our best science is wrong about the age of the world. Being sure is not the same as *being certain*, in the technical sense of assigning probability 1.¹ (Throughout, I will use "certain" in this technical sense, though I suspect that the ordinary usage of "sure" and "certain" is quite close.)

Perhaps the concept I am getting at is moral certainty? For instance, Leibniz writes:

[N]o firm demonstration can be made from the success of hypotheses. Yet I shall not deny that the number of phenomena which are happily explained by a given hypothesis may be so great that it must be taken as morally certain.²

¹ Or "super-1", if we're worried about cases like continuous processes where there are possibilities that have zero probability.

² G. W. Leibniz, *Philosophical Papers and Letters*, trans. and ed. L.E. Loemker (Dordrecht: Kluwer, 1989), 283.

Indeed, I suppose, my reasons for believing that the world is billions of years old have to do with “the success of hypotheses.” However “moral certainty” is something like certainty for all practical purposes. Leibniz expresses this by going on to say:

Indeed, hypotheses of [this] kind are sufficient for everyday use.³

And my sureness that the world is billions of years old is not just a credence sufficient for practical purposes. I might, after all, know that my belief that the world is billions of years old will never actually matter for any practical purposes. If sureness were a sufficiency for practical purposes, then in such a case it would be trivially true that I am sure, no matter what my credence was, which is absurd.

Of course, for just about any proposition one can imagine a scenario where I end up betting for or against it. But sureness is not the same as a credence rationally sufficient for betting on the proposition in all imaginable circumstances, because no credence less than one would rationally suffice for such willingness, and it seems that we should allow for being sure with a credence less than one.

The concept of being sure that I want to look for will be less pragmatic. I will offer an explication (in Carnap’s sense) of being sure in the next section, and then show how this explication offers a precise formula for how high one’s credence needs to be for sureness, in terms of how high one’s credence needs to be for *confidence*.

2. Closure of Inquiry

Being sure is stronger than just being confident. We could, of course, arbitrarily say that you’re confident if you assign a credence of at least 0.99 but sure if you assign a credence of at least 0.999, or one could do empirical research on the level of credence needed for people to claim confidence and sureness in any particular context. But it would be good to offer something more interestingly philosophical, to find something of philosophical significance close to what people mean when they talk of being sure.

One difference between confidence and sureness that is not merely an arbitrary numerical distinction is that being sure will, in some sense, suffice for closure of inquiry.⁴ It won’t necessarily suffice *practically* for rational closure of inquiry. After all, no matter my being sure, as long as the probability is less than one, the payoffs in a betting scenario and the costs of inquiry might make it

³ Leibniz, *Philosophical Papers and Letters*, 283.

⁴ Compare the knowledge account in Kraig W. Martin, *Justified Closure of Inquiry: A Non-Reductive Account* (PhD diss., Baylor University, 2014).

rational to continue inquiry – or I might just be paid to continue inquiry, no matter what I think. In such a case, although I continue the inquiry because it is practically rational to do so, I am not only confident that the proposition I am inquiring about is true, but I am also confident that the inquiry will not change my mind. Or at least will *rationally* not change my mind about the proposition, since the inquiry may be so dangerous that I have a high chance of a head-injury that causes a change of mind or I might be so irrationally stubborn that nothing would change my mind.⁵

We shouldn't understand sureness directly in terms of rational closure of inquiry. One might be irrationally sure in such a way that closure of inquiry would be quite irrational and, more controversially, one might even be rationally sure while realizing that closure of inquiry would be irrational. Rather, I want to suggest, to be sure is to have the level of credence that would be required for a certain kind of rational closure of inquiry.

To get at what that level is, suppose I am a rational agent, I am certain of my future rationality, and I am confident of p . I also am certain that I will engage in a certain line of inquiry. Let L be the event that at the end of that line of inquiry I will not be confident in p . If the probability of L isn't low enough that I be confident that future inquiry will make no difference to my confidence in p , then I am not in a position for rational closure of inquiry, epistemically speaking. And this is not a case where I am sure. I just do not have the kind of security in the face of future rational inquiry that being sure should offer.

A necessary condition for being sure of p , then, is that if one is a rational agent certain of her future rationality, one is confident that future inquiries will not make one lose confidence in p . But whether one is sure should not depend on what future inquiries will actually take place or even what future inquiries are possible. It would be a sign of irrationality to say "I was sure of my hypothesis until I found a way to get funding to test it" on the grounds that once one found a way to get funding to test the hypothesis, then one was no longer confident that one wouldn't lose confidence in the hypothesis. Of course one may well feel less sure when the possibility of being proved wrong looms larger, but (a) this is a sign of irrationality akin to being scared to fly even in cases where one knows it's safer than driving to the airport was, and (b) to *feel* sure or unsure is not the same as to *be* sure or unsure.

⁵ I'm grateful to an interlocutor whose identity has slipped from my memory for the latter suggestion.

Thus, a necessary condition for a perfectly rational agent being sure of p is that she have a level of credence that would suffice for being confident that one rationally won't lose confidence given hypothetical future inquiries.

I now make a crucial posit. Whether one is sure and whether one is confident depend only on the probability one assigns to a proposition. If I have a higher credence for p than you have for q , and you are sure or confident of q , then I am respectively sure or confident of p . (There may, however, be contextual variability as to what the standards for the thresholds are, and so the previous sentence will only be true when these are held constant. See Section 4.) Insofar as our ordinary usage of "is sure" is to some degree infected with how sure one *feels*, this won't match ordinary usage, and so what I am providing is an explication, in Carnap's sense, rather than an analysis.

The posit lets us leverage data about when a perfectly rational agent is confident to get insight on when an imperfect agent is sure: the imperfect agent is sure when her credence is sufficiently high that a perfectly rational agent with that credence would be sure. Putting together the above considerations, we can now give a necessary condition for any agent to be sure. An agent is sure of p only if she assigns a credence r to p such that r satisfies the Rational Confidence in Continued Confidence condition:

(RCCC) Necessarily any perfectly rational agent who knows she will remain perfectly rational and who assigns a credence r to some proposition q is confident that she will remain confident in q .

This condition constrains the credence needed for being sure in terms of the credence needed for being confident. And of course in a standard Bayesian setting, $r=1$ will satisfy RCCC, no matter what the threshold for confidence is.

My main proposal now is to suppose the necessary condition to be sufficient in order to arrive at an explication of being sure:

(SURE) An agent is sure of p if and only if the credence r she assigns to p satisfies RCCC.

On this proposal, being sure is related to a kind of security from rational refutation. One is sure provided that one has sufficient credence that any rational being who is certain of her future rationality is confident in her continued confidence, and hence is in a position to epistemically close inquiry. Of course, one might be sure and yet expect that future inquiry would shake one's own confidence, but that would be a sign that one isn't a rational being who is certain of her future rationality.

One might think that something stronger should be required, namely that to be sure, a rational agent who is certain of future rationality must be sure that she won't stop being sure. But that's too strong a condition if we are to leave open the possibility of being sure while assigning credence less than one, since credences do indeed vacillate, and so an agent slightly above the threshold of being sure – as long as that threshold falls short of one – cannot be confident, much less sure, that her credence won't dip slightly below that threshold. SURE allows the rational agent not to be confident that she will remain *sure*, but only requires that she be confident that her credence won't dip below the lower threshold, that of confidence.

It is a not entirely trivial question, however, whether any credence level $r < 1$ suffices for satisfying RCCC. If it turns out that the answer to this question is negative, a consequence of SURE – and even of the claim that RCCC provides a necessary condition for being sure – will be that one can't be sure without being certain, i.e., without assigning credence 1. In the next section I explore the question of what constraint RCCC places on r .

3. Being Sure

In the Appendix, I will show that the following is a consequence of a Bayesian agent's credential dynamics being a martingale.

Proposition 1 Suppose α and r are strictly between 0 and 1. A rational Bayesian agent who assigns $P(p)=r$ and is certain that she will always update in a Bayesian way assigns a probability of at least $1-(1-r)/(1-\alpha)$ that her credence assignment in p will always remain at or above the level α .

Simple algebraic manipulation then shows:

Corollary 1 Suppose $0 < \alpha < 1$. If a rational Bayesian agent assigns a probability $P(p) \geq 1-(1-\alpha)^2$ to p , then she assigns a probability of at least α that her credence assignment in p will always remain at or above α .

Taking α to be the level of credence needed for confidence, we see that a credence r satisfies RCCC if $r \geq 1-(1-\alpha)^2$, where α is the credence threshold needed for confidence.⁶ Consequently, we now know that SURE is not too rigorous to be satisfied at a credence level less than unity, and we have a sufficient condition for being sure.

⁶ I assume the threshold for confidence is a number such that confidence requires a credence greater than or equal than the threshold. I leave the modifications of my story for the case where the credence must be strictly greater than the threshold to the reader.

Next observe that the inequality in the Corollary is *sharp*. For let $\beta = P(p)$ and suppose that $\beta < 1 - (1 - \alpha)^2$. Assume $\alpha \leq \beta$ (otherwise, we don't even start at the confidence level α .) Next suppose that the line of inquiry that the agent expects to undertake is this. Another agent who knows for certain whether p is true flips a loaded coin with probability γ of heads out of sight of the agent and independently of p . (We will specify γ soon.) The agent then discloses the truth value of the disjunction $p \& h$, where h is the proposition that the coin came out heads. Note that $P(p|p \& h) = 1$ and, due to the independence of p and h ,

$$(1) \quad P(p|\sim(p \& h)) = P(p \& \sim h)/P(\sim(p \& h)) = \beta(1 - \gamma)/(1 - \beta\gamma).$$

Now we specify that $\gamma = \alpha/(1 - (1 - \alpha)^2)$. (It's easy to check that γ is strictly between 0 and 1 if α is.) The probability that $p \& h$ will be disclosed as false is

$$1 - \beta\gamma = 1 - \beta\alpha/(1 - (1 - \alpha)^2) > 1 - \alpha,$$

since we assumed that $\beta < 1 - (1 - \alpha)^2$. Now $\gamma < \alpha/\beta$ by the same assumption. Thus if $p \& h$ disclosed as false, the credence in p will fall below α , because of (1) and since

$$\begin{aligned} \beta(1 - \gamma)/(1 - \beta\gamma) &= 1 - (1 - \beta)/(1 - \beta\gamma) \\ &< 1 - (1 - \beta)/(1 - \beta(\alpha/\beta)) \\ &= 1 - (1 - \beta)/(1 - \alpha) \\ &< 1 - [1 - (1 - (1 - \alpha)^2)]/(1 - \alpha) = \alpha. \end{aligned}$$

Thus, for any credence below $1 - (1 - \alpha)^2$, we can find a case where starting with that credence we have a probability less than α that the credence will remain at or above α . And that case can be one of perfect rationality.

This sharpness shows that the inequality $r \geq 1 - (1 - \alpha)^2$ is not only sufficient for RCCC, but is necessary for it. Thus it follows from SURE that:

(FORMULA) An agent is sure of p if and only if she assigns a credence r to p such that $r \geq 1 - (1 - \alpha)^2$, where α is the credential threshold for confidence.

Thus, if confidence requires a credence of 0.99, then being sure requires $1 - (1 - 0.99)^2 = 0.9999$. If confidence requires 0.9999, then being sure calls for 0.99999999.

4 Closing Remarks

On this account, to be sure is to have a degree of credence sufficient to ensure that one can be confident that one won't lose confidence given further rational inquiry. If one's credence is rational, then in such a case, it is not merely

pragmatically but epistemically appropriate to close further rational inquiry, as one is confident that it would be rationally pointless. Our results give us a formula for what this threshold of credence must be, namely $1-(1-\alpha)^2$ where α is what is needed for confidence.

It is surprising that there is such an exact formula for when one counts as sure. Two things should alleviate this surprise. The first is that we are explicating rather than analysing. There is a natural concept, that of rational confidence in one's continued confidence, that is in the vicinity of our ordinary concept of being sure, and it is this concept that gives rise to the formula. The second thing to remember is that what counts as confidence is vague. So we have a precise formula that relates being sure to being confident, but being confident is something that is far from precise. The vagueness in being confident then transfers precisely to the vagueness in being sure.

Furthermore, it is very likely that what counts as being confident depends on contextual standards. On the above story, the standard for being sure follows in lockstep the contextual standard for being confident. In contexts where confidence is 0.9, being sure is 0.99, while in contexts where confidence is 0.999, being sure is 0.999999. Nonetheless, there is one interesting difficulty. Our formula above assumed that when we talk of *confidence* that one won't lose *confidence*, the same standard of confidence applies at both points. But it might be that the contextual standards of confidence for a first order claim p are different from those for the second order claim that one won't rationally lose that confidence in p . If so, then our formula becomes more complicated, and we leave it as an exercise to the reader to derive that formula from Proposition 1 and an analogue to the reasoning in the sharpness argument.

One might attempt to extend our hierarchy. If α is the level for confidence, and $1-(1-\alpha)^2$ is the threshold for being sure, one might think that a value γ such that $1-(1-\gamma)^2=\alpha$ (i.e., $\gamma=1-\sqrt{1-\alpha}$) is the threshold for *belief*. Thus, one is confident provided that one has a credence that would suffice for rationally believing in continued belief. And so we have a three-fold hierarchy: belief, confidence and sureness. If confidence is at 0.99, then sureness will be at 0.9999 while belief will be at 0.9.

There is also some plausibility in rejecting the above as an account of the relationship between being sure and being confident, while accepting it as an account of the relationship between being confident and simply believing.

The above hierarchy might be extended in both an upward and a downward direction, producing a natural hierarchy of level of credence α_n such

that $\alpha_{n+1} = 1 - (1 - \alpha_n)^2$. For instance, we might have something like “seeing as likely” below belief, and above sureness we might have “super-sureness” at 0.99999999. At each level, the rational agent who is certain of continued rationality will have the next lower level of credence in not falling below that level.

The hierarchy might give us a way of directly identifying a particular natural sequence of thresholds, since it is natural to start at 1/2. Then the sequence of thresholds will be approximately: 0.5, 0.75, 0.9375, 0.9961, 0.99998, 0.9999999998, ... If we wanted to, we could then think of 0.75 as the threshold for seeing as likely (Windschitl and Wells find “likely” to fit with 0.75 in their experiments⁷), 0.9375 for belief, 0.9961 for confidence, 0.99998 for being sure in the ordinary way, and 0.9999999998 for being super-sure. But rather than trying to exactly fit the numbers to ordinary language, it may be more helpful to simply recognize a natural hierarchy of levels of confidence determined by principled considerations.

In any case, intuitively, a credential difference of the sort we find between α and $1 - (1 - \alpha)^2$ marks an important difference. How exactly one matches up the hierarchy with ordinary predicates like “believes,” “is sure” and “is confident” may be less important than recognizing the kind of steps that are found in the hierarchy. Note that if evidential strength or degree of confirmation offered by evidence E to a hypothesis H is measured by the log-likelihood ratio $\log P(E|H)/P(E|\sim H)$, as has been contended by Good⁸ (1984; see also the defense in Pruss 2014), then it is easy to check that if a hypothesis starts at probability 1/2, each successive level of the hierarchy would require approximately double the degree of confirmation relative to the start that the previous did, and that does seem to be an intuitively important step.

Appendix: Argument for Proposition 1

Think of a Bayesian agent’s credences at the start of the data-gathering process as a countably-additive probability P on a probability space $\langle \Omega, \mathcal{F} \rangle$, so that events are members of the σ -field \mathcal{F} of subsets of Ω . We now want to model the evolution of the agent’s credences for a non-empty event $H \in \mathcal{F}$.

⁷ Paul D. Windschitl and Gary Wells, “Measuring Psychological Uncertainty: Verbal Versus Numeric Methods,” *Journal of Experimental Psychology: Applied* 2 (1996): 343–364.

⁸ I. J. Good, “The Best Explicatum for Weight of Evidence,” *Journal of Statistical Computation and Simulation* 19 (1984): 294–299. For a recent defense, see Alexander R. Pruss, “Independent Tests and the Log-Likelihood-Ratio Measure of Confirmation,” *Thought* 3 (2014): 124–135.

The agent's gathering of more and more data can be modeled as a sequence of finer and finer σ -fields $F_0 \subseteq F_1 \subseteq F_2 \subseteq \dots$, where F_0 is the trivial σ -field $\{\emptyset, \Omega\}$. For instance, suppose that at step n , the agent learns whether some relevant piece of evidence E_n obtains. Then F_n is the smallest σ -field generated by the set of events $\{E_1, \dots, E_n\}$. Of course, the data-gathering process can be much more complex. For instance, at step n , the agent might learn the value of some real- or vector-valued random variable Y_n , rather than just the answer to a yes-or-no question as in the case where the agent learns whether E_n obtains. In that case, F_n is the σ -field generated by the variables $\{Y_1, \dots, Y_n\}$. Additional complexity can be modeled. For instance, what experiment the agent does at step n might depend on the information obtained in steps $1, \dots, n$.

The important thing here is that the agent gets more and more information as the process continues, which is modeled by the fact that the σ -fields get finer and finer.

If ω is the agent's actual (but unknown to the agent) position in the state space Ω , then the function taking ω to an agent's credence in H at step n in the data-gathering process where the agent is at ω is equal to (a version of⁹) $P(H|F_n)$, where as usual a conditional probability $P(A|G)$ with respect to a σ -field G is a G -measurable function on Ω such that the conditional expectation of $P(A|G)$ with respect to any non-null event $B \in G$ equals $P(A|B)$.¹⁰ If the field F_n is finite (i.e., only a finite amount of information is received by step n) and its non-empty members have non-zero probability, then $P(H|F_n)(\omega)$ equals $P(H|B)$ where B is the smallest member of F_n containing ω . This models the fact that what an agent at ω by step n has found out is that her position in the state space is a member of B , and being a good Bayesian, her credence in H is of course $P(H|B)$.

Let X_n be the agent's credence at step n . This will be a random variable equal to (a version of) $P(H|F_n)$, and the sequence X_1, X_2, \dots will be a martingale.¹¹ Then $X_n(\omega)$ is the agent's credence at step n . In particular $X_0(\omega)$ is constant and equal to r (we are given that the agent's initial credence is r). Fix any natural

⁹ Conditional probabilities with respect to a σ -field are normally defined only up to sets of measure zero.

¹⁰ For background, see Kai Lai Chung, *A Course in Probability Theory* (San Diego: Harcourt, 2001).

¹¹ Cf. Simon M. Hutteger, "Learning Experiences and the Value of Knowledge," *Philosophical Studies* 171 (2014): 279-288.

number N and let $\tau_N(\omega)$ be equal to N if for all $n \leq N$ we have $X_n(\omega) \geq \alpha$; otherwise, let $\tau_N(\omega)$ be the smallest value of n such that $X_n(\omega) < \alpha$. This is a *stopping time*: a natural-number-valued function such that the event $\{\omega : \tau_N(\omega) = n\}$ is measurable with respect to F_n for each n . Define the random variable Z_N by setting $Z_N(\omega) = X_{\tau_N(\omega)}(\omega)$. This random variable represents the first credence up to time N to drop below α , if there is a credence that drops below α in that time period, and if there isn't, it's just the credence at time N .

By Doob's Optional Sampling Theorem,¹² $E(Z_N) = E(X_0)$. But $E(X_0) = r$. Let $A_N = \{\omega : \exists n (n \leq N \ \& \ X_n < \alpha)\}$ be the event of the credence dropping below α by time N . Then $E(Z_N) \leq P(A_N)\alpha + (1 - P(A_N))$, since on A_N the value of Z_N is less than α while outside A_N (indeed, everywhere) the value of Z_N is at most 1. Thus

$$r = E(Z_N) \leq P(A_N)\alpha + (1 - P(A_N)) = 1 - (1 - \alpha)P(A_N),$$

and so $(1 - r)/(1 - \alpha) \geq P(A_N)$.

Observe that $A_1 \subseteq A_2 \subseteq \dots$ (i.e., if we dip below α by time N , we certainly do so by time $N+1$). Let A be the union of all the events A_N . Now, the agent's credence dips below α at some time or other precisely on the event A , and by countable additivity $P(A) = \lim_{N \rightarrow \infty} P(A_N)$ since the sets A_N are increasing with N . Since $P(A_N) \leq (1 - r)/(1 - \alpha)$, it follows that the probability that the agent's credence ever dips below α is at most $(1 - r)/(1 - \alpha)$, from which the conclusion of Proposition 1 immediately follows.

¹² Chung, *A Course in Probability Theory*, Section 9.3.

WHAT IF BIZET AND VERDI HAD BEEN COMPATRIOTS?

Michael J. SHAFFER

ABSTRACT: Stalnaker argued that conditional excluded middle should be included in the principles that govern counterfactuals on the basis that intuitions support that principle. This is because there are pairs of competing counterfactuals that appear to be equally acceptable. In doing so, he was forced to introduce semantic vagueness into his system of counterfactuals. In this paper it is argued that there is a simpler and purely epistemic explanation of these cases that avoids the need for introducing semantic vagueness into the semantics for counterfactuals.

KEYWORDS: counterfactuals, Ramsey test, coherentism

1. Introduction

At least since Quine introduced the Bizet/Verdi case in 1950 there has been considerable controversy not only about the possibility of there being any adequate analysis of the logic of counterfactual conditionals, but also more specifically about the acceptability of the principle known as conditional excluded middle (CEM).¹ Conditional excluded middle is typically stated as follows:

(CEM) $(A > C) \vee (A > \neg C)$.

CEM is a consequence of what Bonevac calls Stalnaker's rule:²

(SR) $\neg (A > C)$

$A > \neg C$

This issue about CEM was a particular bone of contention between Stalnaker and Lewis as they developed their respective accounts of the logic and semantics of counterfactuals in the late 60s and 70s. Stalnaker ultimately argued that the principle was one that we should incorporate into the logic of counterfactuals (he favored the conditional logic C2) and that, as a result of doing

¹ W. V. O. Quine, *Methods of Logic* (New York: Holt, Reinhart and Winston, 1950).

² Daniel Bonevac, *Deduction*, 2nd ed. (Oxford: Blackwell, 2003).

so we must introduce vagueness into the semantics for such conditionals.³ In point of fact, he advocated doing this specifically by the use of the theory of supervaluations developed by Van Fraassen.⁴ The result then is a semantic theory whereby conditionals in Stalnaker's logic can be true, false or indeterminate.

The main reasons why he advocated this fairly radical approach to the semantics of counterfactuals are twofold. First, it is supposed to explain our inability to choose among competing conditionals like those in the Bizet/Verdi case a unique one that is most acceptable. Second, it supports Stalnaker's intuition that CEM is a plausible principle of conditional logic. Here it will be argued that we can explain our inability to choose a unique most epistemically acceptable conditional from among competing conditionals in Bizet/Verdi cases without recourse to a semantics that incorporates vagueness and that we ought to resist the temptation to introduce vagueness into the semantics of conditionals due to the principle of minimal mutilation. This solution will also allow us to avoid having to choose whether or not to incorporate CEM in the logic of conditionals on the basis of problems with Bizet/Verdi type cases alone and this is a good thing as that determination should probably not be entirely militated by conflicting intuitions or by appeals to what most speakers would affirm about Bizet/Verdi cases alone.

2. Quine's Example and CEM

Quine famously discussed the following pair of conditionals in his 1950 book:

(BV1) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

(BV2) If Bizet and Verdi had been compatriots, Verdi would have been French.

What this pair of conditionals is ultimately supposed to show is that there can be ties in terms of the closeness of counterfactual possible worlds and so Stalnaker's analysis of the logic of counterfactuals is supposed to fail. The basic idea is that while there is good reason to suppose that world where Bizet and Verdi are both French or are both Italian are more similar to the actual world than worlds where they are, for example, Nigerian, Australian or Sri Lankan it seems intuitively to be the case that there is no good reason to suppose either that the world where they are both Italian is closer to the actual world than the world where they are both French or that the world where they are both French is closer to the actual world than the world where they are both Italian. These two counterfactual worlds seem

³ Robert C. Stalnaker, "A Defense of Conditional Excluded Middle," in *Ifs*, eds. William Harper, Robert C. Stalnaker, and Glenn Pearce (Dordrecht: D. Reidel, 1981), 87-104.

⁴ Bas C. Van Fraassen, "Singular Terms, Truth-value Gaps and Free Logic," *Journal of Philosophy* 63 (1966): 481-495.

to be *equally* close to the actual world. As a result, there does not seem to be any reason to treat one conditional as more acceptable than the other. So, more controversially, there is supposed to be no reason to suppose that the first conditional is to be regarded as true and the second as false or vice versa. However, let us look more closely both at how this problem arises and why Stalnaker responds to the Bizet/Verdi case in the way that he does.

3. Stalnaker's and Lewis' Theories in a Nutshell

Stalnaker and Lewis independently proposed accounts of the logic of counterfactuals in the late 60s and early 70s. While these two theories are formally quite similar, they were presented on the basis of somewhat different semantic ideas. Nevertheless, these semantics differences are largely superficial when closely analyzed, with the exception of one major point of disagreement that in turn reflects a major difference in terms of the formal principles characterizing these two logics. Let us begin by looking at the semantics for these two accounts of counterfactuals.

Stalnaker's semantics for counterfactuals was presented in terms of possible worlds and the concept of a selection function.⁵ This selection function f takes proposition and possible world pairs into a possible world. More straightforwardly then, the truth conditions for counterfactuals are given as follows:

$A > B$ is true at world i , if and only if, B is true at $f(A, i)$.

Of course, f is governed by a number of well-known constraints.

Alternatively, Lewis' semantics for counterfactuals was presented in terms of a comparative similarity relation.⁶ Where $S(i, j, k)$ means that j is more similar to i than k is to i , Lewis gives the truth conditions for counterfactuals as follows:

$A > B$ is true, if and only if, there is a A -world j such that B is true at j and all in all A -worlds at least as similar to i as to j .

Stalnaker, however, showed that the choice of presenting semantics in terms of a selection function or in terms of a comparative similarity relation is really arbitrary.⁷ Nevertheless, the two theories of counterfactuals that arise from these semantic basis and the constraints imposed on them are not strictly equivalent. It turns out that when one looks at the details, Stalnaker's theory is a special more-restricted case of Lewis' theory. Lewis' theory involves a well-ordering of all

⁵ Robert C. Stalnaker, "A Theory of Conditionals," in *Studies in Logical Theory*, ed. James W. Cornman (Oxford: Blackwell, 1968), 98-112.

⁶ David Lewis, *Counterfactuals* (Cambridge: Harvard University Press, 1973).

⁷ Stalnaker, "A Defense of Conditional," 87-104.

possible worlds while Stalnaker's theory involves only a weak total ordering of possible worlds. This then gives rise to the crucial point where the theories differ. Stalnaker's theory assumes what Lewis called the limit and uniqueness assumptions. The details of the limit assumption are not important here, but acceptance of it and the uniqueness assumption is what gives rise to the problems associated with CEM noted above.⁸ The uniqueness assumption can be stated as follows:

(uniqueness) for every world i and proposition A there is at most one A -world minimally different from i .

Accepting both of these assumptions amounts to the acceptance of CEM, but the uniqueness assumption is what effectively rules out ties in the similarity of worlds. There cannot be two worlds that are equally similar to a given possible world.

Stalnaker admits that this is an idealization that he has made with respect to the semantics of counterfactuals, specifically with respect to the selection function.⁹ He defends this view on the basis of "...unreflective linguistic intuition,"¹⁰ and argues essentially that treating both of the Bizet/Verdi counterfactuals as indeterminate in truth value better reflects such semantic intuitions than Lewis' view, where they both turn out to be false.

4. Coherence as a Guide to Counterfactual Acceptance

Stalnaker and Lewis developed their semantic views of counterfactuals in terms of truth conditions and this was framed in terms of possible worlds. However, as argued in the previous section, the issue of the acceptability of CEM should not be driven by semantic considerations. Rather, what is needed is a clear account of the acceptability conditions for counterfactuals that explains the resistance to CEM and Bizet/Verdi type cases. Fortunately, there has been considerable discussion of this matter in the debate about the Ramsey test for conditional acceptance that is so-named because of Ramsey's brief footnote comment made in a paper in 1929.

In this vein, Carlos Alchourrón, Peter Gärdenfors, and David Makinson developed the AGM theory of belief revision in the 1980s and a number of related theories have arisen as a consequence.¹¹ Here we will specifically focus on the

⁸ See Charles B. Cross, "Conditional Excluded Middle," *Erkenntnis* 70 (2009): 173-188 for discussion of the relationship between the limit assumption, the uniqueness assumption and the principle of counterfactual consistency.

⁹ Stalnaker, "A Defense of Conditional," 89.

¹⁰ Stalnaker, "A Defense of Conditional," 92.

¹¹ See Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson, "On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision," *Journal of Symbolic Logic* 50

version of this view as presented by Gärdenfors.¹² These theories are fundamentally based on the concept of a belief state, belief set or a corpus of beliefs, K , typically satisfying the following minimal conditions (where it is assumed that belief states are given a representation in some language L):

(BS) A set of sentences, K , is a belief state if and only if (i) K is consistent, and (ii) K is objectively closed under logical implication.

The content of a belief state is then defined as the set of logical consequences of K (so $\{b: K \in b\} =_{df.} Cn(K)$). Given this basic form of epistemic representation, the AGM-type theories are intended to be a normative theory about how a given belief state which satisfies the definition of a belief state is related to other belief states satisfying that definition relative to: (1) the addition of a new belief b to K_i , or (2) the retraction of a belief b from K_i , where $b \in K_i$. Belief changes of the latter kind are termed *contractions*, but belief changes of the former kind must be further sub-divided into those that require giving up some elements of K_i and those that do not. Additions of beliefs that do not require giving up previously held beliefs are termed *expansions*, and those that do are termed *revisions*.¹³ Specifically, for our purposes here it is the concept of a revision that is of crucial importance to the issue of providing an account of rational commitment for conditionals. In any case, given AGM-style theories the dynamics of beliefs will then simply be the epistemically normative rules that govern rational cases of contraction, revision and expansion of belief states.

The fundamental insight behind these theories is then that belief changes that are contractions should be fundamentally conservative in nature. In other words, in belief changes one ought to make the minimal alterations necessary to incorporate new information and to maintain or restore logical consistency. This fundamental assumption is supposed to be justified in virtue of a principle of informational economy. This principle holds that information is intrinsically and practically valuable and so we should retain it at all costs unless we are forced to do otherwise. So, while the details are not important here, the revision operations on belief states are restricted so as to obey a principle of minimal mutilation.

(1985): 510-30, Peter Gärdenfors, *Knowledge in Flux. Modeling the Dynamics of Epistemic States* (Cambridge: MIT Press, 1988) and Isaac Levi, *For the Sake of the Argument: Ramsey test Conditionals, Inductive Inference, and Nonmonotonic Reasoning* (Cambridge: Cambridge University Press, 1996).

¹² In Gärdenfors, *Knowledge in Flux*.

¹³ In point of fact the AGM theory really only holds that there are two dynamical operations on belief states, because revision is defined in terms of expansion and contraction.

What is important to the topic of this paper is that on the basis of such theories of belief revision, the defenders of this approach to belief dynamics have also proposed that one could also give a theory of rational conditional commitment.¹⁴ The core concept of this theory is the Ramsey Test:¹⁵

(RT) Accept a sentence of the form $A > C$ in the state of belief K if and only if the minimal change of K needed to accept A also requires accepting C .¹⁶

Even in this quasi-formal form we can see what the AGM and other theorists have in mind. The Ramsey Test requires that we modify our beliefs by accepting A into our standing system of beliefs and then see what the result is.¹⁷ This view is typically framed in terms of a version of the epistemological coherence theory of justification and this seems natural given BS.¹⁸ The idea is that one's beliefs are justified to the degree that they hang together or are mutually supportive. The idea then is that our belief system is justified in virtue of this feature of the system as a whole and there are several extant version of

¹⁴ See Peter Gärdenfors, "An Epistemic Approach to Conditionals," *American Philosophical Quarterly* 18 (1981): 203-211, Gärdenfors, *Knowledge in Flux*, and Peter Gärdenfors, "Imaging and Conditionalization," *The Journal of Philosophy* 79 (1982): 747-760.

¹⁵ See F. P. Ramsey, "Laws and Causality," reprinted in *F.P. Ramsey: Philosophical Papers*, ed. D. H. Mellor (Cambridge: Cambridge University Press, 1929/1990). See Michael Shaffer, "Three Problematic Theories of Conditional Acceptance," *Logos & Episteme* 1 (2011): 117-125 and Michael Shaffer, "Doxastic Voluntarism, Epistemic Deontology and Belief-Contravening Commitments," *American Philosophical Quarterly* 50 (2013): 73-82 for some discussions of problems for naïve formulations of the Ramsey test.

¹⁶ For a relatively recent discussion of RT and related views see Isaac Levi, *Mild Contraction* (Oxford: Oxford University Press, 2004).

¹⁷ David H. Sanford, *If P, then Q: Conditionals and the Foundations of Reasoning*, 2nd ed. (New York: Routledge, 2003) contains the objection that in many cases where the antecedent of such a conditional is a radical departure from what we believe to be the case, we cannot in fact employ the Ramsey test because we do not know what would be the case if we believed such an antecedent. So, he claims that many conditions are simply void, rather than true or false. It is worth pointing out here that Sanford's criticism is weak at best. It simply does not follow that because we cannot always clearly determine what would be the case if we were to believe some claim, a conditional with such an antecedent has no truth value. See chapters 5 and 6 of Timothy Williamson, *The Philosophy of Philosophy* (Blackwell, Oxford, 2007) for discussion of one suggestion for how such knowledge might be obtained.

¹⁸ See Peter Gärdenfors, "The Dynamics of Belief Systems: Foundations Versus Coherence Theories," in *Knowledge, Belief and Strategic Interaction*, eds. Cristina Bicchieri, Maria Luisa Dalla Chiara (Cambridge: Cambridge University Press, 1992) for the most thorough defense of the AGM theory in terms of coherentism. See Michael Shaffer, "Coherence, Justification, and the AGM Theory of Belief Revision," in *Perspectives on Coherentism*, ed. Yves Bouchard (Ontario: Aylmer-Éditions du Scribe, 2002) for some worries about this view.

coherence theory that are plausible views of justification.¹⁹ The most famous are of course those of Bonjour and Lehrer, but Thagard's version is also a well-regarded and more recent version of coherentism.²⁰ In any case, we need not get bogged down in the debate about the particular details of coherentism here and we can simply adopt a basic, largely unanalyzed and broadly intuitive conception of that view for the purposes of this paper. This is also desirable because the results here are then not dependent on any particular version of coherence theory. So we shall simply accept that a belief corpus is coherent to the degree that its elements fit together and are mutually supportive. Once we accept this interpretation of RT and the notion of a belief state on which it is based, there is a natural way to extend RT to cases of comparative acceptance for conditionals.

First, it is important to note that it is not at all clear that on RT either BV1 or BV2 is acceptable. This is because the minimal change of belief needed to incorporate the claim that Bizet and Verdi are compatriots does not obviously *require* accepting either that Bizet would have been Italian or that Verdi would have been French. But, both BV1 and BV2 seem to be acceptable conditionals nonetheless because accepting the shared antecedent permits one to accept either that Bizet would have been Italian or that Verdi would have been French. What is also important in the case of BV1 and BV2 is that they in an important sense compete. We then need to introduce the appropriate concept of a competitor as it applies to counterfactual conditionals. For the purpose of this paper we can simply adopt the following concept of the competition of conditionals:

(COMP) A counterfactual conditional $A \rightarrow C$ competes with all other counterfactual conditionals that have A as an antecedent.

So, in the case of the Bizet/Verdi conditionals, we have a case of two competing conditionals and this should be no surprise. As we have seen there is something important about the relationship between those two conditionals that ties them together intimately. Given COMP we can then replace RT with an appropriate concept of comparative acceptance given the coherentist interpretation of belief states as follows:

¹⁹ There is of course some controversy about such views, especially those that are framed in terms of probabilistic notions of coherence. See Luc Bovens and Stephan Hartmann, *Bayesian Epistemology* (Oxford: Oxford University Press, 2003) and Erik J. Olsson, *Against Coherence*, (Oxford: Oxford University Press, 2005) for discussion of this matter.

²⁰ See Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985), Keith Lehrer, *Theory of Knowledge* (Boulder: Westview Press, 1990), and Paul Thagard, *Coherence in Thought and Action* (Cambridge: MIT Press, 2000).

(CCA) Accept a sentence $A > C$ in the state of belief K rather than $A > B$ if and only if the minimal change of K needed to accept A , K' , permits accepting C , the minimal change of K needed to accept A , K'' , also permits accepting B and the changes necessary to maintain the coherence of K' are less extensive than those necessary to maintain the coherence of K'' .

So defined, the principle of comparative conditional acceptance allows us to introduce a differential notion of conditional acceptance that is normative because it is based on the coherence theory of justification. Moreover, as we shall see in the next section, it allows us to explain Bizet/Verdi cases without having to depend entirely on suspicious appeals to semantic intuitions and without having to introduce vagueness into the semantics for those conditionals.²¹

5. Explaining Bizet/Verdi Cases.

So why are our two conditionals so problematic and how does CCA make sense of the apparently problematic nature of them? Recall the Bizet/Verdi conditionals:

(BV1) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

(BV2) If Bizet and Verdi had been compatriots, Verdi would have been French.

By COMP BV1 and BV2 are competing counterfactual conditionals. Now if we apply CCA to our dual of sentences we should see that the revision of our state of belief K by the addition of the shared antecedent of BV1 and BV2 permits the acceptance both of the claim that (I) Bizet would have been Italian and it also permits the acceptance of the claim that (F) Verdi would have been French.²² This can be made more apparent by comparing the case of BV1 and BV2 with the cases where BV1 and BV2 are compared in terms of CCA with the following conditional:

(BV3) If Bizet and Verdi had been compatriots, Bizet would have been Dutch.

The changes necessary to accept BV3 are clearly more extensive than those needed to maintain consistency given the acceptance of BV1 or BV2. Moreover, given the relevant parts of our belief corpus and our intuitive understanding of coherence it is also reasonable to suppose that the revision of K by I, K' , and the revision of K by F, K'' , are *equally extensive*. Both resultant belief states hang

²¹ The reliability of semantic intuitions has recently been questioned in Edouard Machery. Ron Mallon. Shaun Nichols, and Stephen P. Stich, "Semantics, Cross-cultural Style," *Cognition* 92 (2004): B1-B12 and the reliability of intuitions in general have been more generally question in Jonathan M. Weinberg, Shaun Nichols, and Stephen Stich, "Normativity and Epistemic Intuitions," *Philosophical Topics* 29 (2001): 429-460..

²²This can be seen also in that both BV1 and BV2 satisfy RT.

together or are mutually supportive to the same degree – or to a very similar degree – given what we know about Bizet, Verdi and the world in general, and the degree of change necessary to incorporate the antecedent and consequent of both is not noticeably different. It is just as coherent and requires the same sorts of changes of the same degree to suppose that, if the two men were compatriots, Bizet would be French as it is to suppose that, if the two men were compatriots, Verdi would be Italian. But the changes necessary to pursue either of these options in a coherent manner are clearly less extensive than the changes necessary to entertain the supposition that if the two men were compatriots, Bizet (or Verdi) would have been Dutch. Importantly, this means that while both BV1 and BV2 are acceptable there is no reason to accept BV1 over BV2 and no reason to accept BV2 over BV1 as per CCA. This then straightforwardly *explains* our inability to determine which is true and it explains this without any appeal to semantic vagueness and without any unsupported appeals to semantic intuition. As a result, we do not need to take Stalnaker's radical semantic steps in order to deal with these sorts of cases. If the theory of counterfactual acceptance presented here is even broadly correct, then that the Bizet/Verdi cases are odd may well just be a reflection of a purely epistemic phenomenon and nothing deeper. This recognition in turn then shows that the Bizet/Verdi type cases do not decide the issue of CEM one way or the other. The *metaphysical/semantic* matter about whether there can be ties in terms of the similarities of worlds is not decided simply because we cannot *epistemically* distinguish conditionals in Bizet/Verdi type cases, and in deference to the principle of minimal mutilation we ought to resist the move to introduce vagueness into the semantics of conditionals *pace* Stalnaker.

DISCUSSION NOTES/DEBATE

TWO NON-COUNTEREXAMPLES TO TRUTH-TRACKING THEORIES OF KNOWLEDGE

Fred ADAMS, Murray CLARKE

ABSTRACT: In a recent paper, Tristan Haze offers two examples that, he claims, are counterexamples to Nozick's Theory of Knowledge. Haze claims his examples work against Nozick's theory understood as relativized to belief forming methods *M*. We believe that they fail to be counterexamples to Nozick's theory. Since he aims the examples at tracking theories generally, we will also explain why they are not counterexamples to Dretske's Conclusive Reasons Theory of Knowledge.

KEYWORDS: Robert Nozick, tracking theories, Fred Dretske, conclusive reasons

In a recent paper, Tristan Haze¹ offers two examples that, he claims, are counterexamples to Nozick's Theory of Knowledge.² Haze claims his examples work against Nozick's theory understood as relativized to belief forming methods *M*. We believe that they fail to be counterexamples to Nozick's theory. Since he aims the examples at tracking theories generally, we will also explain why they are not counterexamples to Dretske's Conclusive Reasons Theory of Knowledge.³

As Haze rightly points out, we maintain that to fully understand Nozick's Tracking Theory one must know that Nozick relativizes tracking to the knower's belief-forming method *M*.⁴ Nozick explains that a subject might know something by one method, but not by another because one method enables one to track the truth and the other doesn't.

We will use Haze's own formulation of Nozick's tracking conditions (though this is not exactly Nozick's wording of his conditions). We will explain below why this matters.

S knows, via method (or way of knowing) *M*, that *p* iff

¹ Tristan Haze, "Two New Counterexamples to the Truth-Tracking Theory of Knowledge," *Logos and Episteme* VI, 3 (2015): 309-311.

² Robert Nozick, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).

³ Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1-22.

⁴ Fred Adams and Murray Clarke, "Resurrecting the Tracking Theories," *Australasian Journal of Philosophy* 83, 2 (2005): 207-221.

1. p is true
2. S believes, via method M , that p
3. If p weren't true, and S were to use M to arrive at a belief whether (or not) p , then S wouldn't believe, via M , that p
4. If p were true, and S were to use M to arrive at a belief whether (or not) p , S would believe, via M , that p .⁵

Here is Haze's first example:

I have a deep-seated, counterfactually robust, delusional belief that my neighbor is a divine oracle. He is actually a very reliable and truthful tax-lawyer. There is a point about tax law he has always wanted to tell me, p . One day, he tells me that p , and I believe him, because I believe he is a divine oracle. I would never believe him if I knew he was a lawyer, because I am very distrustful of lawyers.

Haze claims that he does not know that p because his belief rests upon a delusion (though counterfactually robust). We take it that Haze thinks the reason he doesn't know that p is that part of the explanation of his trusting the lawyer is that he delusionally believes his neighbor is a divine oracle and not a lawyer. If he thought the neighbor was a lawyer, he would not believe anything he tells him.

But what is his belief-forming method M ? Suppose that his delusion infects his belief-forming methods. If so, we don't see why this would be a counterexample to Nozick. After all, delusional belief forming methods fail to track the truth. That is partly what makes them delusions. A deluded person may fail to satisfy either condition 3 or 4, and thus not know that p . Deluded people may believe false things or fail to believe true ones (actually and counterfactually). If this is why Haze claims that he fails to know that p in this example, then he is mistaken to think this is a counterexample to Nozick's account. Nozick's theory would give the same result.

Of course, if the delusion is only about whether or not the neighbor is a lawyer, and not about anything the neighbor says to Haze about tax law, then the delusion does not infect Haze's belief-forming methods about propositions uttered by the neighbor. In that case, given the reliable testimony of the neighbor and the reliability of Haze's hearing and understanding what the neighbor says and his belief forming method of trusting what the neighbor says about tax law, we fail to see why Haze would not know that p . His belief forming methods about what the neighbor says about tax law are delusion-free. So his beliefs about tax law track the truth and Nozick's theory yields the result that Haze knows that p . We see this as the right result and not a counterexample to Nozick. So we think Haze draws

⁵ Haze, "Two New Counterexamples," 310.

the wrong conclusion in this example and he is mistaken to claim he doesn't know. Hence, in example one, either his delusion does spread and infects his belief-forming methods or not. But in neither case is it a counterexample to Nozick's theory.

Here is Haze's second example:

My neighbor is a tax lawyer. Here, unlike in the previous counterexample, I have no delusional belief. It is my neighbor who is the strange one: for years, he has intently nurtured an eccentric plan to get me to believe the truth about whether p , where p is a true proposition of tax law, along with five false propositions about tax law. His intention to do this is very counterfactually robust. He moves in next door to me and slowly wins my trust. One day, he begins to regale me with points of tax law. He asserts six propositions: p and five false ones. I believe them all.

Haze claims that he does not know the true proposition of tax law p , but that Nozick's theory would claim that he does know that p . This is not the case. Nozick's theory implies no such thing. Nozick's theory implies the opposite. We think the reason Haze believes this is a counterexample is because he relativizes the method M to the neighbor and the neighbor's dispensing of information and not to Haze's own belief-forming methods. Haze seems to think the method here is that with respect to the true proposition p , the neighbor would not say " p " unless p . This causes Haze to think Nozick's tracking conditions are satisfied and that Nozick's theory implies that Haze knows that p . However, this is not the case. And it is not the case even if, with respect to p only, the neighbor wouldn't say " p " unless p .

Nozick is very clear that methods are the belief-forming methods of the cognizer. In this case, the relevant method M has to be something Haze uses or has some control over. He has no direct control over what the neighbor chooses to tell him, so the neighbor's method of dispensing information is not Haze's method of consuming information. Haze's method M in the example is to trust what the neighbor says. And this method clearly does not track the truth because it is not restricted to " p " alone, but freely ranges over the other five falsehoods the neighbor utters and Haze believes. So this too, when properly understood, does not constitute a counterexample to Nozick's tracking theory. Contrary to Haze's claim that these two examples are different than the kind of examples we discussed,⁶ they in fact are importantly the same in so far as they fail to be counterexamples to Nozick's theory.

⁶ In Adams and Clarke, "Resurrecting the Tracking Theories."

Furthermore, they are not counterexamples to Dretske's Conclusive Reasons Theory either.⁷ In example 1, the reason R that Haze believes that p is that the neighbor said " p ." And the neighbor, being a very reliable and truthful tax lawyer, would not have said " p " unless p . So, on Dretske's theory, Haze would know that p , via conclusive reason R . Of course, if the delusion were affecting Haze's formation of beliefs about things the tax lawyer says, then Haze would not believe p solely based on R (what the lawyer said). And so if his delusion spread, he would not know according to Dretske's Conclusive Reasons Theory of Knowledge.

And in example 2, Haze would not know that p on the basis of R , where R is 'believing that p because the lawyer said " p ."' It is false that the lawyer would not have said something about tax law unless it were true. This counterfactual is false because the lawyer utters five other false propositions, which Haze believes. Of the six things the neighbor tells Haze, Haze has no way of discriminating which are true and which are false. When Haze believes that p and the neighbor says " p ," saying " p " sounds to Haze indiscriminately the same in truth value to the neighbor's saying " q ," " r ," " s ," " t ," and " v ." However, Haze cannot tell which are true and which are false just by the neighbor's utterances. So while he believes the truth with respect to p , he does not know that p is true because R (the neighbor said it) is not a conclusive reason for p .

What both of these purported counterexamples have in common is that a particular belief with a strange epistemic pedigree is advanced. In the first case, the method M involved, according to the cognizer, is 'believe the oracle.' However, the 'oracle' is just a very reliable and truthful tax lawyer and so the method is, in fact, reliable and the cognizer knows that p . In the second case, the method M , 'believe my neighbor,' is unreliable most of the time and so the cognizer fails to know that p . Haze thinks the cognizer fails to know in both cases because he focuses only on the specific belief and fails to consider whether the method is reliable or not. We think he knows in the first case, but not the second, because the method is reliable in case one, but not case two. But notice that Haze treats these examples as if they were single-case problems. The problem is that they are not single-case methods at all. That is, they mimic the single-case horn of the Generality Problem that Goldman faced concerning the reliable-process theory of epistemic justification.⁸ The Generality Problem concerns the issue of how to individuate the width of process types in a principled, or non-question-

⁷ Fred Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1-22.

⁸ Alvin Goldman, "What is Justified Belief?" in *Justification and Knowledge*, ed. George Pappas (Dordrecht: D. Reidel, 1979): 1-23.

begging, way.⁹ The issue can be posed as a dilemma. If we individuate process types too narrowly then a reliable process might have just one instance. But process types cannot be tokens or all true beliefs would be reliably produced and all false beliefs would be unreliably produced. This is the 'single-case' horn of the dilemma. On the other hand if we construe process types too widely then the 'no distinction' problem awaits us. For instance, visual perception might count as a process type such that all beliefs formed on that basis are equally justified. But beliefs about mountain goats where the percipient is 300 yards away from the goat are clearly not as reliable a belief-forming method as the same belief arrived at from viewing the goat at 30 feet. The problem is that Goldman's reliable process account of epistemic justification and knowledge has no way to draw this distinction in a principled way. This is the 'no distinction' horn of the generality problem. The solution to the Generality Problem consists in providing a principled, i.e., non *Ad Hoc*, account of process types that is neither too narrow nor too wide. The Generality Problem remains an important, but unresolved, issue for the reliable process account of justification.¹⁰

One way to think about the difference between Dretske's and Nozick's tracking theories of knowledge (DTK and NTK) and a Goldman-style reliability theory of knowledge (GTK), is that the latter is simply offering a weaker tracking account. Where GTK requires that the process type be .9 reliable (or thereabouts) in near possible worlds for a reliable true belief to count as knowledge, DTK and NTK require complete reliability.¹¹ But for all of these accounts, the process type or method *M* must be generally reliable. In other words, the process type or method *M* must be a type, not a token. Process types cannot be individuated too narrowly or the single-case horn of the generality problem will be in play. We think that this is exactly where Haze goes wrong since both of his examples are treated as if they were single-cases where the process type is a token, not a type. If the lawyer in case one is reliable and truthful as Haze asserts such that he would not have said "*p*" unless *p*, then the method of 'believing the Oracle' will produce knowledge that *p* and would produce knowledge in every near possible world, not just in this case. The method *M* is reliable, so this is not a single-case. In case two, the method *M* is not reliable concerning five of the six beliefs. Hence, the method

⁹ Murray Clarke, *Reconstructing Reason and Representation* (Cambridge: The MIT Press, 2004), 85.

¹⁰ Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," *Philosophical Studies* 89 (1998): 1-29.

¹¹ In Goldman's "What is Justified Belief," he says a reliable process is one that generates more true beliefs than false. Here we say a process is reliable if it generates true beliefs approximately 90 percent of the time because Goldman said this in conversation with Clarke.

M is unreliable and the cognizer does not know that *p* even when *p* is true. Again, the method *M* is not a single-case. Haze could only make his case if the method *M* in these examples was a single-case, but it isn't. In both cases, the method *M* is a type, in the first case that type of method is reliable and in the second, it is not reliable. Haze neglects the fact that reliability accounts of knowledge require that the method *M* must be nomically reliable, it must be a reliable process type that the cognizer employs. In Nozick's case, those process types (methods *M*) are individuated from the standpoint of the cognizer. For Dretske, this work is done in the Conclusive Reasons Theory of Knowledge where, holding certain conditions *C* fixed, the cognizer's reason *R* is his reason for believing that *p*. The result is that the following condition must be satisfied in order to know that *p*: Given your reason or evidence *R* and fixed circumstances *C*, it must be the case that it is not physically or circumstantially possible that not-*p*. If it is physically or circumstantially possible that not-*p*, then one cannot know that *p* in those circumstances.

So, for instance, in Dretske's famous thermometer example, *that* the thermometer is working properly must be held fixed when considering counterfactuals concerning a child's temperature in near possible worlds.¹² Hence, the thought that the thermometer might be broken cannot be used as the basis of an objection to his account since, as Dretske says: 'if it is *that* kind of thermometer' then, of course, one cannot know the child's temperature is *p*.¹³ Ultimately, what is held fixed and what is allowed to vary depends on the laws of nature (or other law-like circumstantial conditions) operating at the time and the "conclusiveness" of the reason *R*. If one is using a defective thermometer then one cannot know that the child's temperature is normal even if it is normal. Dretske's point is that the thermometer must be such that in the circumstances *C* it would not read "*p*" unless *p*, in order to have conclusive empirical reasons for believing that *p*.

The upshot of all of this for Haze is that his purported counterexamples ignore the role that 'holding the method *M* fixed' plays in Nozick's account of knowledge and its equivalent, in Dretske's account of knowledge. That role requires that the method *M* must be reliable in the sense that the process type employed must be generally reliable – it must be a genuine 'type,' not a 'token,' in order to avoid the single-case problem. In the case of reliable indicator accounts of knowledge such as Dretske and Nozick defend, the effect of imposing counterfactual requirements has the same effect: the reasons *R*, or method *M*, must be reliable in counterfactual situations or near possible worlds. All reliability

¹³ Dretske, "Conclusive Reasons," 2.

theorists share this conviction. But the method *M* in both of his examples actually is a type (pace Haze), not a token, and so the tracking theories would get the correct result in these cases. For this reason, these examples are failed single-case objections to tracking theories.¹⁴

¹⁴ Haze is not alone in his misunderstanding of tracking theories of knowledge. John Williams and Neil Sinhababu make the same type of mistake in their recent paper. See John Williams and Neil Sinhababu, "The Backward Clock, Truth-Tracking, and Safety," *Journal of Philosophy* 112, 1 (2015): 46-55. As a result, a reply to them is currently in progress. Finally, we would like to thank John Barker for his comments on this paper.

A FAILED TWIST TO AN OLD PROBLEM: A REPLY TO JOHN N. WILLIAMS

Rodrigo BORGES

ABSTRACT: John N. Williams argued that Peter Klein's defeasibility theory of knowledge excludes the possibility of one knowing that one has (first-order) a posteriori knowledge. He does that by way of adding a new twist to an objection Klein himself answered more than forty years ago. In this paper I argue that Williams' objection misses its target because of this new twist.

KEYWORDS: defeasible reasoning, Peter Klein, second-order knowledge,
John N. Williams

This is a reply to John N. Williams' paper "Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge."¹ That paper argues that Peter Klein's defeasibility theory of knowledge excludes the possibility of one knowing that one has (first-order) a *posteriori* knowledge. Klein himself answered a version of this objection in "A Proposed Definition of Propositional Knowledge."² Williams' paper adds a new twist to the objection Klein answered more than forty years ago. I will argue that Williams' objection misses its target because of this new twist.

1. The Old Problem and the Old Solution

When fully spelled out, Klein's analysis of knowledge comes down to this:

(Defeasibility) *S* knows that α iff (1) α ; (2) *S* believes that α ; (3) *S* is justified in believing that α ; (4) there is no truth, *d*, such that the conjunction of *d* and *S*'s justification, *j*, fails to justify *S* in believing that α .³

¹ John N. Williams, "Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge," *Analysis* 75 (2015): 213-17.

² Peter Klein, "A Proposed Definition of Propositional Knowledge," *The Journal of Philosophy* (1971): 471-82.

³ Since a truth may misleadingly suggest the falsehood of something one is justified in believing truly (as in the Grabit Case introduced in the literature by Keith Lehrer and Thomas Paxson, "Knowledge: Undefeated Justified True Belief," *The Journal of Philosophy* 66 (1969): 225-37. In that case the truth "Tom's mother said that Tom has an identical twin who is also in the library" misleadingly suggests that "Tom stole the book" is false.), Klein's view incorporates a distinction

Towards the end of his paper,⁴ Klein considered the following objection to Defeasibility:⁵

If the definition were accepted, it would never be true that *S* knows that she knows that *x* because she could never know that the fourth condition held.

In reply to this objection Klein points out that, given Defeasibility, *S* knows that she knows *x* if and only if *S* knows “*S* knows that *x*” satisfies each of the necessary conditions in Defeasibility. In other words, *S* knows that she knows that *x* if and only if each of the following statements is true:

- (I) *S* knows that *x*
- (II) *S* believes that *S* knows that *x*
- (III) *S* is justified in believing that she knows that *x*
- (IV) There is no truth, *d*, such that the conjunction of *d* and one’s justification, *j*, fails to justify *S* in believing that *S* knows that *x*.

As Klein⁶ points out, because knowing entails that there is no defeater of one’s justification, *S* is justified in believing she knows that *x* only if she is justified in believing there is no defeater of her justification for believing that *x*. In other words, III is true only if *S* is justified in believing there is no defeater of her justification for believing *x*. In the same paper Klein argued that there is no reason to think that *S* is never justified in believing there is no defeater of the justification she has for her first-order belief.

This, in a nutshell, is Klein’s solution to the old problem. Before we look at John Williams’ new version of this objection, let me substantiate Klein’s reply by

between truths that actually defeat one’s justification (i.e., *genuine defeaters*) and truths that only appear to defeat one’s justification (i.e., *misleading defeaters*). Only the former truly defeats. In this paper I will refer only to genuine defeaters, but will drop the qualifier “genuine” for ease of exposition. Nothing in my exchange with Williams depends on this issue. See Peter Klein, *Certainty: A Refutation of Scepticism* (Minnesota: University of Minnesota Press, 1981), 148–166 for his treatment of the distinction.

⁴ Klein, “A Proposed Definition,” 480.

⁵ Even though I follow the argument in Klein, “A Proposed Definition” here, I have updated the nomenclature he used in that paper to a more current one, in line not only with Klein’s later work (e.g., Klein, *Certainty* and Peter Klein, “Useful False Beliefs,” in *New Essays in Epistemology*, ed. Quentin Smith (New York: Oxford University Press, 2008)) but also with more widespread use in current epistemology. The nomenclature in Klein, “A Proposed Definition” followed closely the nomenclature in Roderick Chisholm, *Theory of Knowledge* (Upper Saddle River: Prentice Hall, 1966). Nothing of substance hinges on these changes.

⁶ Klein, “A Proposed Definition,” 481.

providing a logically possible case in which I through IV are all true. This should establish that Defeasibility does not exclude second-order knowledge.

As I look up I undergo the experience as of something being a computer screen in front of me. I thereby form the belief that

(p) there is a computer screen in front of me.

Since this is a normal case of perceptual experience, I satisfy all conditions in Defeasibility, i.e.,

(I*) I know that p .⁷

Suppose further that I reflect on whether I know that p , realize that it is a normal case of perceptual experience, and come to believe I do know it. That is, the following is true:

(II*) I believe I know that p .

I* and II* entail that I have a true second-order belief. Now, according to Klein, S is justified in believing that α if and only if, given S 's evidence, S 's belief in α satisfies some (perhaps contextually determined) threshold for knowledge-grade justification.⁸ This means that I know I have knowledge-grade justification for believing there is a computer screen in front of me only if I know that my justification for believing that there is one is not defeated. But my total evidence bearing on the issue of whether I am justified in believing that p includes not only my knowledge that p , but also my knowledge that this is a normal case of perceptual experience, that I am not drugged or otherwise visually impaired, and so on. Thus, we may plausibly argue that, given my evidence, I am in a position to know that there is no defeater of my justification for believing that p . Defeaters prevent one from knowing by preventing one's justification from satisfying the (perhaps contextually determined) threshold for knowledge-grade justification. They prevent S 's justification from satisfying this threshold by either undermining the support her evidence provides to her belief, or by making probable the denial of what she believes given her evidence.⁹ In the case at hand, there would be a

⁷ Although this is a case of *non-inferential* knowledge, the same could be said, *mutatis mutandis*, about *inferential* knowledge.

⁸ This is, roughly, what Klein means by his notion of *confirmation*, which is the centerpiece of his account of justification. See Klein, *Certainty*, 61-7.

⁹ According to the nomenclature popularized by John Pollock, the first kind of defeater is an *undermining defeater*, while the latter kind of defeater is a *rebutting defeater*. See John Pollock, "Defeasible Reasoning," in *Reasoning: Studies of Human Inference and Its Foundation*, eds. Jonathan Adler and Lance J. Rips (Cambridge: Cambridge University Press, 2008) for a recent statement of Pollock's view.

defeater of my justification for believing that there is a computer screen in front of me if, for example, I had taken a drug which causes hallucinations 80 percent of the time, or if $\neg p$ were true. But, by assumption, nothing like that is true in this situation. In other words, both III* and IV* are true:

(III*) I am justified in believing I know that p .

(IV*) There is no truth, d , such that the conjunction of d and my justification, j , for believing that I know that p fails to justify me in believing that I know that p .

Claims I* through IV* all seem to be true in this case; so, it is plausible to think that I know that I know that p . The upshot is that Defeasibility does not make it impossible for there to be second-order knowledge. I conclude, then, that contrary to what Williams would have us believe it is logically possible for Klein's Defeasibility to be true and for one to know that one has first-order *a posteriori* knowledge.

2. Williams' New Twist

Williams' new twist to the old objection comes in the form of a principle about concepts he finds "plausible:"¹⁰

(CLAIM) If the satisfaction of a condition at least partly constitutes an instance of a concept, then knowing that such an instance obtains requires you to know *a priori* that the condition is satisfied.¹¹

¹⁰ Williams, "Not Knowing," 215.

¹¹ Although Williams does not explicitly formulate CLAIM as requiring *a priori* knowledge, one must read CLAIM in this way lest his argument against Klein be made invalid (see below), for Williams explicitly requires that S know *a priori* that she satisfies the no-defeater condition in order for her to know that she knows. If I am wrong about this and Williams' argument is invalid, then so much the worse for his argument. More precisely, this is what I take to be Williams' argument:

1. If the satisfaction of a condition at least partly constitutes an instance of a concept, then knowing that such an instance obtains requires you to know *a priori* that the condition is satisfied. [CLAIM/Assumption]
2. If the satisfaction of a condition at least partly constitutes an instance of *knowledge*, then knowing that such an instance obtains requires you to know *a priori* that the condition is satisfied. [CLAIM/ from 1]
3. The satisfaction of the no-defeater condition partly constitutes instances of *knowledge*. [from Defeasibility]
4. For any instance k of *knowledge*, if you know that k obtains in case C , then you know *a priori* that the no-defeater condition is satisfied in case C . [from 2 and 3]

To get a feel for how CLAIM works, consider Williams' own example:¹² since x being three-sided partially constitutes x being a triangle, I know that x is a triangle only if I know that x is three-sided. Now, CLAIM and Defeasibility together entail that one knows that one knows α only if one knows *a priori* that one's justification satisfies the no-defeater condition. Williams then argues that, since one cannot know *a priori* that one's knowledge that α satisfies the no-defeater condition, one cannot know that one knows that α . This is Williams' *new twist* to the old objection: it is not enough that S knows that her first-order knowledge satisfies all conditions on knowledge, if she wants to know that she knows, she must know *a priori* that her first-order *a posteriori* knowledge satisfies all the conditions on knowledge.

Let us look more closely at CLAIM and at Williams' new twist. Our assessment will reveal that CLAIM and the instance of this principle Williams applies to *knowledge* are both false.

Suppose that satisfying the condition

(*) S can prove (some) mathematical theorems

partially constitutes the concept *mathematician*. The assumption is plausible because we commonly think of mathematicians as people who can prove *at least one* mathematical theorem. Now, consider Timmy, who is a freshman in college and not particularly math-savvy. If Timmy were confronted with a proof of a mathematical theorem he would not be able to follow it; he would not even be able to grasp any of the concepts in the proof. Now, suppose Timmy's Calculus professor, a skillful mathematician, satisfies condition (*), and that on the first day of class she tells Timmy and all the other students in Timmy's class that she can prove many mathematical theorems. Intuitively, Timmy knows his teacher is a mathematician even though this concept is partially constituted by condition (*) and his knowledge that the professor satisfies (*) is *a posteriori*, for it is based on his experience as of something being his calculus professor telling him she satisfies (*). But if that is the case, then CLAIM is false on account of the fact that Timmy knows the concept *mathematician* is instantiated by his professor, even though he does not know *a priori* that the professor satisfies a condition that partially constitutes that concept. As a matter of fact, it seems to me that Timmy would know *a posteriori* that his professor is a mathematician even if she had not told the class that she satisfies (*),

5. You cannot know *a priori* that the no-defeater condition is satisfied in C .
[Assumption]

You do not know that k obtains in C . [from 4 and 5 by *modus tollens*]

¹² Williams, "Not Knowing," 215.

but told them only that she is a mathematician. Either version of the case counterexemplifies CLAIM.

Now, consider CLAIM as it applies to *knowledge*:

(KLAIM) If the satisfaction of a condition at least partly constitutes an instance of *knowledge*, then knowing that such an instance obtains requires you to know *a priori* that the condition is satisfied.

KLAIM is false because of Williams' new twist. To see that, let us look at what happens when we apply KLAIM to the other traditional conditions on knowledge (i.e., the justification, belief, and truth conditions).

Take justification and belief first. If KLAIM is true, then one cannot know *a posteriori* that those conditions are satisfied. This is a bad result because our second-order knowledge that those conditions are satisfied *is* sometimes justified *a posteriori*. I am completely ignorant of quantum mechanics, but if Stephen Hawking were to tell me that *q* is a testable prediction of the theory, then, assuming this is a normal case of transmission of knowledge via testimony, I not only come to know that *q* is a testable prediction of quantum mechanics, but I am also in a position to know both that I believe that *q* and that I am justified in believing that *q*. The problem for KLAIM is that my justification for believing that I believe that *q* with justification is arguably *a posteriori*, for it includes the justification that emerges from my undergoing a particular experience: if I had not experienced Stephen Hawking, the celebrated physicist, asserting to me that *q*, I would not have believed that *q*, nor would I have been justified in believing that *q*.

Things get worse when we apply KLAIM to the truth condition on knowledge. Williams faces a dilemma: if KLAIM is true, then, necessarily, either there is no second-order knowledge or no first-order *a posteriori* knowledge. That there is such a dilemma should be reason enough to reject KLAIM and Williams' argument, which relies on it. No epistemology that accepts either (or both) of those horns should be deemed satisfactory.

Here is how KLAIM forces this dilemma on Williams. As before, let "*p*" stand for the claim that there is a computer screen in front of me. Also as before, suppose that I know that *p* and that I know that *p* in virtue of my true belief being suitably related to my experience as of something being a computer screen in front of me. As a result, the justification for my knowledge that *p* is *a posteriori*. Suppose I reflect on the question of whether I know that *p* and come to believe I do know it in virtue of reliably assessing my perceptual experience as veridical. Now, a condition on knowledge is that the known proposition be true. Because knowledge entails truth, it follows from KLAIM that I know that I know that *p* only if I know *a priori* that my belief that *p* satisfies this condition; that is, given KLAIM, I know that I

know that p only if I know *a priori* that my belief that p is true. But if one knows that a belief is true, then one knows the truth the belief is about. So, given KLAIM, I know that I know that p only if *I know a priori that p* . But, by assumption, I know that p *a posteriori*. We have derived a contradiction from KLAIM by applying it to a seemingly innocent case. Something's gotta give. I think KLAIM has got to go. If KLAIM is true, then either my knowledge that p is not *a posteriori* or I can't know that I know that p . The first horn of this dilemma seems false on its face, and the second one leads to a curious form of skepticism: considering that there is nothing special about this case, the result of this argument generalizes to all cases of first-order *a posteriori* knowledge.

3. Conclusion

In sum, John Williams' new twist on the old problem for Defeasibility fails. His problem for Defeasibility arises only when the requirements for iterative knowledge are made too high. What is more, this lesson applies to a number of other views that also incorporate a no-defeater clause in their definition of *knowledge*.^{13,14}

¹³ e.g., Keith Lehrer, *Theory of Knowledge 2nd ed.* (Boulder: Westview Press, 1990), John Pollock and Joseph Cruz, *Contemporary Theories of Knowledge* (Lanham: Rowman and Littlefield, 1999), and Marshall Swain, *Reasons and Knowledge* (Ithaca: Cornell University Press, 1981).

¹⁴ I am very grateful to Cherie Braden, Peter Klein, and John N. Williams for discussion and feedback on different drafts of this paper. I am happy to acknowledge that the research in this paper was partly funded by the CAPES/Fulbright Commission. I am also grateful for the partial support my research received from the São Paulo Research Foundation (FAPESP) through grant 2015/02419-4.

STILL A NEW PROBLEM FOR DEFEASIBILITY: A REJOINDER TO BORGES

John N. WILLIAMS

ABSTRACT: I objected that the defeasibility theory of knowledge prohibits you from knowing that you know that p if your knowledge that p is *a posteriori*. Rodrigo Borges claims that Peter Klein has already satisfactorily answered a version of my objection. He attempts to defend Klein's reply and argues that my objection fails because a principle on which it is based is false. I will show that my objection is not a version of the old one that Klein attempts (unsuccessfully) to address, that Borges' defence of Klein's reply fails and that his argument against my new objection leaves it untouched.

KEYWORDS: defeasibility theory of knowledge, *a posteriori*, Peter Klein

I objected that the defeasibility theory of knowledge prohibits you from knowing that you know that p if your knowledge that p is *a posteriori*.¹ Rodrigo Borges claims that Peter Klein has already satisfactorily answered a version of my objection.² He attempts to defend Klein's reply and argues that my objection fails because a principle on which it is based is false.

I will show that my objection is not a version of the old one that Klein attempts (unsuccessfully) to address, that Borges' defence of Klein's reply fails and that his argument against my new objection leaves it untouched.

1. The Old Objection, Klein's Unsatisfactory Reply and Borges' Defence of Klein

What might be called the *early* defeasibility theory of propositional knowledge may be formulated as follows.

¹ John N. Williams, "Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge," *Analysis* 75 (2015): 214.

² Rodrigo Borges, "A Failed Twist to an Old Problem: A Reply to John Williams," *Logos and Episteme* VII, 1 (2016): 75-81, citing Peter Klein, "A Proposed Definition of Propositional Knowledge," *Journal of Philosophy* 68 (1971): 471-82. This is a more charitable description of Borges' strategy than his own. What he actually says is that "Klein himself answered a version of this objection in 'A Proposed Definition of Propositional Knowledge.' Williams' paper adds a new twist to the objection Klein answered more than forty years ago. I will argue that Williams' objection misses its target because of this new twist." Surely my objection cannot fail simply because it is a *different* objection from the one Klein addresses.

John N. Williams

You know that p just in case you have a justified true belief that p and there is no defeater D of your justification for believing that p

where

D is a defeater of your justification for believing that p just in case it is a truth such that believing it would render your belief that p unjustified.

This is essentially the theory that Klein proposes, calling a defeater, “a disqualifying proposition,”³ and taking “you have a justified true belief that p ” as synonymous with “ p is evident to you.”⁴ We may represent this definition of knowledge as:

S knows that p just in case

- (1) p
- (2) S believes that p
- (3) S is justified in believing that p
- (4) There is no defeater of S 's justification for believing that p .

As Borges observes,⁵ Klein considers the objection that “If the definition were accepted, it would never be true that S knows that he knows that p because he could never know that the fourth condition held.”⁶ This is the old objection. Now, a *first way* to analyse S 's knowledge that she knows that p in terms of the early defeasibility theory is to substitute “ S knows that p ” for “ p ” in each of the four conditions above, yielding

³ Klein, “A Proposed Definition,” 475.

⁴ Klein (“A Proposed Definition,” 475) proposes that “ S knows that p at t if and only if

- (i) p is true;
- (ii) S believes p at t ;
- (iii) p is evident to S at t ;
- (iv) there is no true proposition such that if it became evident to S at t , p would no longer be evident to S .”

Roderick M. Chisholm, in *Theory of Knowledge* (Englewood Cliffs, N.J.: Prentice Hall, 1966), 22, has it that sufficient conditions for p to be evident to S are that it is more reasonable for S to believe p than to withhold belief in p and that there is no proposition such that it is more reasonable for S to believe it than it is for him to believe p . To these Klein adds a third, that S has no reason to believe that the situation is abnormal (“A Proposed Definition,” 473). Nonetheless, he takes “evident” as interchangeable with “justified” (“A Proposed Definition,” 473).

⁵ Borges, “A Failed Twist,” 76.

⁶ Klein, “A Proposed Definition,” 480.

- (1') *S* knows that *p*
- (2') *S* believes that she knows that *p*
- (3') *S* is justified in believing that she knows that *p*
- (4') There is no defeater of *S*'s justification for believing that she knows that *p*.⁷

This is precisely what Klein does in response to the old objection. He then says that "It seems quite clear that these conditions could be fulfilled; or rather, the definition itself does not rule out the possibility that these conditions are fulfilled."⁸ This misses the point of the objection, which, to concentrate on the fourth condition, was *not* that *there can be no defeater of S's justification for believing that she knows that p*, but rather that she could not know (4), in other words that *she could not know that there is no defeater of her justification for believing that p*. Klein could try claiming that the definition does not rule out the possibility of *S* knowing (4), but that would merely beg the question. To be fair to Klein however, the objection itself gives no reason why *S* cannot know (4). It seems likely that Klein fails to answer the objection he anticipates because it is easy to confuse this first way of analysing *S*'s knowledge that she knows that *p* in terms of the early defeasibility theory with a *second way*. This is to analyse it as *S*'s knowledge of each of the four conditions, yielding:

- (1'') *S* knows that *p*
- (2'') *S* knows that she believes that *p*
- (3'') *S* knows that she is justified in believing that *p*
- (4'') *S* knows that there is no defeater of her justification for believing that *p*

The difference between (4') and (4'') is crucial, because the objection was not that (4') cannot be satisfied but that (4'') cannot be satisfied. That objection is premised upon this second way of analysis, not the first. This enables us to see that the objection relies upon the *knowing you know principle*:

If you know that you know that *p*, then you know the content of each necessary condition of your knowing that *p*.

We can now give the form of the *argument for the old objection*:

⁷ Klein ("A Proposed Definition," 481) puts this result as

"1' *S* knows that *p*, ...

2' *S* believes that he knows that *p*;

3' '*S* knows that *p*' is evident to *S*...

4' There is no disqualifying proposition for '*S* knows that *p*.'"

⁸ Klein, "A Proposed Definition," 481.

- I. Given the early defeasibility theory, if you know that p then (4) there is no defeater of your justification for believing that p .
- II. Given the early defeasibility theory, if you know that you know that p , then (4'') you know that there is no defeater of your justification for believing that p . (from I plus the knowing you know principle)
- III. You cannot know that there is no defeater of your justification for believing that p .
- IV. Given the early defeasibility theory, you cannot know that you know that p . (from II and III)

We should note that no reason is given for III. Borges then attempts to “substantiate Klein’s reply by providing a logically possible case” in which (1')-(4') are all true.⁹ This strongly suggests that Borges has in mind the second way of analysing second-order knowledge, which explains why he follows Klein in missing the point of the old objection. If he were to succeed in giving a case in which (1')-(4') are all true then this would show that that IV, namely the conclusion of the old objection, is false. But it would not show what goes wrong with the argument for that conclusion.

Borges then attempts to give the following case in which (1')-(4') are all true, namely *Computer Screen*. You appear to see a computer screen in front of you, which is at least part of your justification for believing that there is one in front of you. He stipulates that this is a normal case of perceptual experience, that you know that there is a computer screen in front of you and that each of (1)-(4), namely the conditions of the early defeasibility theory, are met. I have no quarrel with this. I agree that the early defeasibility theory may allow you first-order knowledge.¹⁰

Setting external-world scepticism aside,¹¹ there are many different sorts of possible truths that, if you believed them, would render you unjustified in believing that there is a computer screen in front of you. These include the possible truth that you are prone to hallucination when near electrical equipment, that there is a trick of the light that causes you to mistake a printer for a computer screen or that makes the screen appear to be in front of you when it is in fact to one side, that what you are looking at is a realistic facsimile of a computer screen, and so on. Since Borges has stipulated that this is a normal case of perception, these are likewise stipulated not to be truths, in other words that there is no

⁹ Borges, “A Failed Twist,” 76-77.

¹⁰ Williams “Not Knowing You Know,” 215.

¹¹ As I did in “Not Knowing You Know,” 214.

defeater of your justification for believing that there is a computer screen in front of you.

So far so good. Borges continues, "Suppose further that I reflect on whether I know [that there is a computer screen in front of me], realize that it is a normal case of perceptual experience, and come to believe I do know it."¹² He goes on to claim that it is clear that there is no "defeater of my justification for *believing that I know* that there is a computer screen in front of me" because "*by assumption*, nothing like that is true in this situation."¹³

This does not follow. Borges has stipulated that there is no defeater of your justification for *believing that there is a computer screen in front of you*, with the result that the early defeasibility theory allows you to know that there is one in front of you. But this is not to stipulate that there is no defeater of your justification for *believing that you know that there is a computer screen in front of you*. For it is possible that although there is no defeater of your justification for believing that *p*, there is nonetheless a defeater of your justification for believing that you *know* that *p*. Such a case arises in *Biology Teacher*.

Someone who looks exactly like your biology teacher tells you that the insect that you have captured is a Brazilian wandering spider. You know your teacher to be reliable, sincere and an expert on spiders and so you come to believe that she has transmitted her knowledge to you, with the result that you now know that the insect is a Brazilian wandering spider. But unbeknownst to you, (*D*) the person who has just told you this is the identical twin of your teacher who is no expert on spiders, but can still classify them fairly accurately (or to a degree of statistical accuracy that warrants belief without constituting knowledge-grade justification).

If you were to come to believe (*D*) then you would cease to be justified in thinking that you *know* that the insect is a Brazilian wandering spider, but you would still be justified in thinking that it *is* one.

A second problem that afflicts Borges' defence of Klein is that although your perceptual experience of what appears to be a computer screen in front of you might be normal, it is far from obvious how you can *realize* that it is normal merely by reflection. I will return to this point below in section 3.

Borges concludes his defence of Klein by saying that "Defeasibility does not make it impossible for there to be second-order knowledge."¹⁴ Borges means that the early defeasibility theory does not prohibit you from knowing that you know

¹² Borges, "A Failed Twist," 77.

¹³ Borges, "A Failed Twist," 78 (my italics).

¹⁴ Borges, "A Failed Twist," 78.

John N. Williams

that *p*. But I nowhere claim that it does. Instead my new objection is only that it prohibits you from knowing that you know that *p* if your knowledge that *p* is a *posteriori*.¹⁵ I now turn to my new objection.

2. My New Objection

My new objection starts with the *conceptual principle*.

If the satisfaction of a condition at least partly constitutes an instance of a concept, then knowing that such an instance obtains requires you to know that the condition is satisfied.¹⁶

For example, since the concept of a triangle is at least partly constituted by it being three-sided, you know that a figure is a triangle only if you know that it is three-sided. The early defeasibility theory is intended to be an analysis of the concept of knowledge, with the result that the satisfaction of condition (4) is supposed to at least partly constitute an instance of the concept of knowledge. So given the early defeasibility theory, by the conceptual principle, your knowing that there is an instance of your knowledge requires you to know that (4) is satisfied. To take Borges's own example, on the early defeasibility theory, knowing that you know that there is a computer screen in front of you requires you to know that

(A) There is no defeater of your justification for believing that there is a computer screen in front of you.

This is equivalent to

(B) There is no truth such that believing it would render you unjustified in believing that there is a computer screen in front of you.

This in turn is equivalent to

¹⁵ Williams, "Not Knowing You Know," 214.

¹⁶ Williams, "Not Knowing You Know," 215. Strictly speaking, you might not have the concept of the satisfaction of a condition, and so given the plausible principle that you can have a belief only if you have the ability to think the thought of its content, you could not believe, nor therefore know, anything about the satisfaction of conditions. A more cumbersome but more accurate formulation of the principle is as follows.

If the satisfaction of a condition that *q* at least partly constitutes an instance of a concept *C*, then knowing that *p*, where *p* reports an instance of *C*, requires you to know that *q*.

For ease of exposition I will stick with the less cumbersome formulation. Nothing turns upon this.

- (C) Every truth is not such that if you were to believe it then this would render you unjustified in believing that there is a computer screen in front of you.

But how could you possibly know (C)? You cannot know it *a priori*. You cannot tell in advance that (C) is true. Mere reflection will not allow you to foresee all threats to your justification of your belief that there is a computer screen in front of you. As we saw above in section 1, there are many different sorts of possible truths that, if you believed them, would render you unjustified in believing that there is a computer screen in front of you. Who is to say what all of these truths are? The only other way for you to know (C) is to know every truth and ascertain that each is not such that if you were to believe it then this would render you unjustified in believing that there is a computer screen in front of you. This is not a way that you can follow, for the simple reason that being less than omniscient, you cannot know every truth.

This problem for the early defeasibility theory arises when your first-order knowledge that *p* is *a posteriori*. This is because the justification for your belief that *p* comes from experience, with the result that reflection alone will not enable you to verify that there are no other empirical truths that would defeat your experiential justification should you believe them. Nor will experience enable you to verify this, because you cannot consider every empirical truth.

We can now give the form of the *argument for my new objection*:

- I'. Given the early defeasibility theory, the satisfaction of condition (4) at least partly constitutes an instance of the concept of knowledge.
- II. Given the early defeasibility theory, if you know that you know that *p*, then (4'') you know that there is no defeater of your justification for believing that *p*. (from I' plus the conceptual principle)
- III'. There are only two ways in which you could know that there is no defeater of your justification for believing that *p*, namely by knowing *a priori* that every truth is not such that if you were to believe it then this would render you unjustified in believing that *p*, or by knowing every truth and ascertaining that each is not such that if you were to believe it then this would render you unjustified in believing that *p*.
- IV'. If your knowledge that *p* is *a posteriori*, then you cannot know *a priori* that every truth is not such that if you were to believe it then this would render you unjustified in believing that *p*, nor can you know every truth and ascertain that each is not such that if you were to believe it then this would render you unjustified in believing that *p*.
- V'. Given the early defeasibility theory, you cannot know that you know that *p* if your knowledge that *p* is *a posteriori*. (from II', III' and IV')

But any satisfactory theory must allow you not only to have *a posteriori* knowledge, but also to know that you have it. The early defeasibility theory does not allow this. So the theory is unsatisfactory.

3. Borges' Failed Reply to My New Objection

As in the argument for the old objection, one step in this argument is II. However, there are three important differences. First, my argument for my new objection appeals to the conceptual principle, whereas that for the old objection appeals to the knowing you know principle. Second, the argument for the old objection contains an important premise that is unsupported, namely III. In contrast, my argument for my new objection provides reasons why you cannot know that there is no defeater of your justification for believing that *p* when that justification is *a posteriori*. Third, the conclusion of the argument of the old objection is that given the early defeasibility theory, you cannot know that you know that *p*. In contrast, that of my new objection is only that given the early defeasibility theory, then you cannot know that you know that *p* if your knowledge that *p* is *a posteriori*.

Given these differences, it seems quite a stretch for Borges to describe my objection as a "new twist to the old objection."¹⁷ My objection is substantially different from the old objection that both Klein and Borges fail to address.

Nor is it true, contrary to Borges, that my objection is that it is impossible for the early defeasibility theory to be true and also for one to acquire second-order knowledge. That is the old objection, not mine.¹⁸

In response to my objection Borges again appeals to *Computer Screen*. He argues that since you may realize by reflection that this is a normal case of perceptual experience, you may realize in advance that there is no defeater of your justification for believing that there is one in front of you. This is supposed to follow because he has already stipulated a normal case of perceptual experience to

¹⁷ Borges, "A Failed Twist," 78.

¹⁸ In fact it seems to me that the early defeasibility theory allows you to know that you know that *p* if your knowledge that *p* is *a priori*. Suppose that you reflect upon the concepts of 2 and addition and come to recognize, and hence believe, that $2 + 2 = 4$, because you realize that this could not possibly be otherwise. By introspection, you may realize that you believe this and so come to recognize that what you yourself now believe is a conceptual truth (your justification for what you believe). You may now reason that since $2 + 2$ could not possibly be other than 4, there could be no truth that would deprive you of your justification for thinking that $2 + 2 = 4$, were you to believe it, because no truth could change the fact that $2 + 2$ could not possibly be other than 4, as you could still realize. You can indeed know in advance that there is no defeater of your justification for believing that $2 + 2 = 4$.

be one in which there are no truths such as that you are prone to hallucination when near electrical equipment and so on.¹⁹

There are at least two problems with this response. First, it is difficult to see how mere reflection on your apparent experience of a computer screen in front of you could tell you that this is a normal case of perceptual experience and not one in which, for example, you are looking at a realistic facsimile of a computer screen. You might know that your past experiences have overwhelmingly turned out to be veridical, thus providing you with inductive justification for thinking that your present experience is veridical too. But this would at least involve your memory of how past cases have turned out, not just your reflection on the present case.

Second and most importantly, *realizing* that this is a normal case of perception in which you appear to see a computer screen in front of you, involves *knowing* that there is one in front of you, so Borges' response assumes that if you know that there is one in front of you, then you may know that there is no defeater of your justification for believing that there is one in front of you. This begs the question of whether (4) is a necessary condition of knowledge. I argue that it can't be, for if it were a necessary condition, then you could not know that you have *a posteriori* knowledge, yet you could indeed know this.

We should also note that the conceptual principle makes no mention of *a priori* knowledge. However Borges represents it as CLAIM:

If the satisfaction of a condition at least partly constitutes an instance of a concept, then knowing that such an instance obtains requires you to know *a priori* that the condition is satisfied.²⁰

Borges writes

Although Williams does not explicitly formulate CLAIM as requiring *a priori* knowledge, one must read CLAIM in this way lest his argument against Klein be made invalid, for Williams explicitly requires that *S* know *a priori* that she satisfies the no-defeater condition in order for her to know that she knows.²¹

But I do not even *implicitly* require that *S* knows *a priori* that she satisfies the no-defeater condition in order for her to know that she knows! In my

¹⁹ Borges says that "Suppose that I reflect on whether I know ... realize that it is a normal case of perceptual experience ..." ("A Failed Twist," 77) and that "my total evidence bearing on the issue of whether I am justified in believing ... includes ... also my knowledge that this is a normal case of perceptual experience, that I am not drugged or otherwise visually impaired, and so on." ("A Failed Twist," 77).

²⁰ Borges, "A Failed Twist," 78.

²¹ Borges, "A Failed Twist," 78, note 11.

argument, II makes no mention of *a priori* knowledge. Nonetheless my argument is valid as it stands.

This misunderstanding undermines Borges' two remaining criticisms of my argument. The first of these is that there is a counterexample to CLAIM, namely *Mathematician*:

Suppose that *S*'s ability to prove some mathematical theorems at least partly constitutes the instance of the concept *S is a mathematician*. Timmy is unable to grasp any concepts involved by a mathematical proof. His calculus professor tells him that she can prove many mathematical theorems.²²

Borges observes that intuitively, Timmy knows *a posteriori* that the professor is a mathematician, since his knowledge is based on his experience of what she tells him. But he does not know *a priori* that she has the ability to prove some mathematical theorems, for knowing that is also based on what she tells him. This falsifies CLAIM. So it does, but my argument relies upon the conceptual principle, not CLAIM.²³

Borges' remaining criticism is that applying CLAIM to the concept of knowledge results in two false predictions. He assumes that a condition of your knowing that *p* is that you have a justified belief that *p*. Satisfying this condition at least partly constitutes the concept of knowledge.²⁴ So CLAIM predicts that if you know that you know that *p*, then you know *a priori* that you have a justified belief that *p*. Given that you know this *a posteriori*, this falsifies CLAIM.²⁵ So it does, but my claim is not CLAIM, only the conceptual principle.

Likewise, a condition of your knowing that *p* is that *p*, and satisfying this truth-condition at least partly constitutes the concept of knowledge. So by CLAIM, knowing that an instance of your knowledge obtains (in other words, knowing that you know that *p*) requires you to know *a priori* that the truth-condition is satisfied (in other words, you know *a priori* that *p*). Thus if you know

²² Borges, "A Failed Twist," 79.

²³ Another problem with this example is that it is not enough to just *suppose* that *S*'s ability to prove some mathematical theorems at least partly constitutes the instance of the concept *S is a mathematician*. To falsify the conceptual principle, or even CLAIM for that matter, Borges needs an *actual* case in which the satisfaction of a condition at least partly constitutes an instance of a concept. I also note in passing that there is a kind of circularity in the supposition, since the satisfaction of the concept *S is a mathematician* already involves the concept of mathematics. However I am unsure what to make of this.

²⁴ Of course there are non-defeasibilists who will argue that the concept of knowledge is in no way constituted by justified belief.

²⁵ Borges, "A Failed Twist," 80.

that you know that p , then you know *a priori* that p .²⁶ Given that you may know that you know that p in a case in which your knowledge that p is *a posteriori*, this again falsifies CLAIM.²⁷ Yet again it does, but once again, my objection to the early defeasibility theory in no way relies upon CLAIM.

4. Concluding Remarks

Both Klein and Borges fail to address the old objection to the early defeasibility theory and Borges' attempt to defend Klein against it fails. In any case my new objection is substantially different from the old one. Borges' attempt to fault my objection fails largely because he attributes to me a principle that I neither hold nor need.

In my "Not Knowing You Know," I argue that my new objection counts not only against the early defeasibility theory, but also against what might be called the *later* defeasibility theory originating from Klein:

You know that p just in case you have a justified true belief that p and there is no undefeated defeater D of your justification for believing that p

where

D is defeated by D^* just in case D^* is a truth such that believing it, in addition to believing D , preserves your original justification for believing that p .²⁸

²⁶ This may be put equivalently as follows. Either you do not know that you know that p or your knowledge that p is *a posteriori*. This is why Borges thinks that I face a dilemma, since for some instances of " p ," I would not want to accept either disjunct. See Borges, "A Failed Twist," 80.

²⁷ Borges, "A Failed Twist," 80-81.

²⁸ See Peter Klein, *Certainty: A Refutation of Scepticism* (Minneapolis: University of Minnesota Press, 1981). The early theory incorrectly excludes cases of knowledge, as shown by the case of *Tom Grabit*, in Keith Lehrer and Thomas Paxson, "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* 66 (1969): 225-37. Here is an adapted version.

Using your reliable vision and memory, you see someone who looks just like Tom Grabit stealing a book at the library, and on this basis believe that he stole a book. Unbeknownst to you, Tom's mother claims that he is away on a trip and has an identical twin who is in the library. But still unbeknownst to you, she is demented. Tom did steal a book.

Surely you know that Tom stole a book, but there is a truth that would render you unjustified in believing that Tom stole one were you to believe it, namely that Tom's mother claims that he is away on a trip and has an identical twin who is in the library. In contrast, the later defeasibility theory explains why you know that Tom stole a book. There is also another truth, namely that she is demented. If you were to believe both truths, then your original justification for believing that Tom stole a book would be preserved. Your justification is ultimately undefeated.

John N. Williams

In this rejoinder to Borges I have only discussed the early defeasibility theory, since as he observes, nothing in his reply to me depends on this distinction.²⁹ Everything I have said in this rejoinder will apply equally to the later theory.³⁰

²⁹ Borges, "A Failed Twist," 76, note 3.

³⁰ I am most grateful to Claudio de Almeida for his insightful discussion of previous drafts.

JUSTIFICATION AND THE UNIQUENESS THESIS AGAIN – A RESPONSE TO ANANTHARAMAN

Luis ROSA

ABSTRACT: I reinforce my defense of permissivism about the rationality of doxastic attitudes on the face of a certain body of evidence against criticism published in this journal by Anantharaman. After making some conceptual clarifications, I manage to show that at least one of my original arguments pro-permissivism is left unscathed by Anantharaman's points.

KEYWORDS: uniqueness thesis, epistemic rationality, permissiveness

In “Justification and the Uniqueness Thesis” (published in this journal),¹ I argued that the following claim is false: for any proposition p and body of evidence E , there is at most one type of doxastic attitude that is rational for one to take toward p on the basis of E . This is the so-called ‘uniqueness thesis.’²

In the original arguments, I used considerations about cognitive limitation (possession or lack of inferential abilities) in order to motivate the following claims: (i) there are situations in which the attitude of belief and the attitude of suspension of judgment would both be rational attitudes for one to take towards p given one’s evidence E , and (ii) there are situations in which the attitude of belief and the attitude of disbelief would both be rational attitudes for one to take towards p given one’s evidence E . Call the former the ‘moderate permissiveness thesis,’ and the latter the ‘extreme permissiveness thesis.’ Both the moderate and the extreme permissiveness theses are inconsistent with the uniqueness thesis.

My case against the uniqueness thesis was recently criticized by Muralidharan Anantharaman (also in this journal),³ who purports to defend a particular version of that thesis against my objections. He does so by appealing to what he calls a “non-deontic account of rationality.”⁴ Before I address

¹ Luis Rosa, “Justification and the Uniqueness Thesis,” *Logos & Episteme* III, 4 (2012): 572-574.

² See Richard Feldman, “Reasonable Religious Disagreements,” in *Philosophers Without Gods*, ed. Louise M. Antony (New York: Oxford University Press, 2007), 205.

³ Muralidharan Anantharaman, “Defending the Uniqueness Thesis: A Reply to Luis Rosa,” *Logos & Episteme* VI, 1 (2015): 129-139.

⁴ Anantharaman, “Defending the Uniqueness Thesis,” 130.

Anantharaman's objections, however, let me make some important preliminary points – some of which I believe were not sufficiently clear in the original paper.

First and foremost, the uniqueness thesis I take to be false is not a thesis about sheer evidential support. It is rather a thesis about *ex ante* justification/rationality (as opposed to *ex post* or doxastic justification/rationality): it is about the range of doxastic attitudes that one is entitled to form given one's available evidence (whether or not the subject has already formed the relevant doxastic attitudes). To be sure, it may be false that there is at most one type of doxastic attitude that is rational for one to take toward *p* given one's evidence *E* while it is still true that *E* *either* gives sufficient support to *p*, *or* to not-*p*, *or* it is neutral with respect to *p* – where neither of these disjuncts is compatible with the others.

Second, I take it that the truth-conditions for attributions of *ex ante* justification or rationality should not abstract away from human's cognitive capacities. Believing that *p* is *ex ante* justified or rational for *S* when and only when *S* is epistemically entitled to believe that *p*. But a subject is not entitled to believe that *p* on the basis of certain body of evidence *E* when she is not able to do so competently. That is: no one is epistemically entitled to form a belief *in the wrong way*, even if one has evidence that gives overall support to the propositional content of that belief.⁵ E.g., even if a certain complicated theorem *T* follows from one's available reasons, it might not be rational for one to believe that *T*: it may be the case that the *only* ways one can come to believe that *T* using one's available reasons are epistemically bad ways. In that case, given one's situation, one is not entitled to believe that *T*, for if one were to believe that *T* one would do so in the wrong way.

So I submit that it is rational or justified for a subject *S* to believe that *p*, given *S*'s evidence *E*, only when *S* is able to believe that *p* on the basis of *E* in the right way (or *S* is able to *competently* believe that *p* on the basis of *E*). What is the right way of forming a belief on the basis of certain reasons? There are at least two things we can briefly say about this here.

First, a belief is formed on the basis of certain reasons in the right way only when the relevant process is an instantiation of a reliable type of process. I.e., if one's belief *Bq* is based on one's beliefs *Bp*₁,..., *Bp*_{*n*} in the right way, then the process that leads one from *Bp*₁,..., *Bp*_{*n*} to *Bq* is a token of a reliable type of process – a type of process that is truth-conducive. This means that all or almost all tokens

⁵ For examples of beliefs formed on the basis of good reasons in the wrong way, see John Turri, "On the Relationship Between Propositional and Doxastic Justification," *Philosophy and Phenomenological Research* LXXX, 2 (2010): 312-326.

of that type will be such that, *if* the propositional contents of their input-states are true, then the propositional contents of their output-states are necessarily or most likely true. Second, a belief Bq is formed on the basis of Bp_1, \dots, Bp_n in the right way only when the truth of p_1, \dots, p_n is *relevant* to the truth of q . E.g., the process of forming a belief that $\sqrt{25} = 5$ on the basis of the belief that *It is raining in New York* is a 100% reliable process, but the ‘premise’ here is irrelevant to the ‘conclusion.’ So a criterion of relevance must also be satisfied here.

With these preliminary considerations in place, let me now address Anantharaman’s criticism. Here I will concentrate only on the moderate permissiveness thesis. My arguments for the extreme version of the permissiveness thesis relied on the idea that we can rationally believe things on the basis of inconsistent reasons, and it is my understanding that I would need to deal with that issue in a more detailed manner first. The thesis that we can rationally believe things on the basis of inconsistent reasons deserves special attention itself. So this is not the place where I will try to make a case for it. Still, if I can successfully defend the moderate version of the permissiveness thesis against Anantharaman’s criticism, my point will still hold that the uniqueness thesis is false.

Anantharaman takes the concept of justification that I originally used to be a ‘deontic’ concept of justification, that is, as a concept that obeys to an ‘ought implies can’ principle. As it was pointed out above, I do indeed defend that *ex ante* justification or rationality is not only a matter of what evidence is available to the subject, but also of whether the subject is able to form the relevant doxastic attitude on the basis of her evidence in the right way. There is actually no ‘ought implies can’ principle involved here, however. Rather, what is involved is a ‘can implies can’ principle, where the former ‘can’ expresses *epistemic* permissibility and the latter one expresses cognitive ability.

Even so, that concept of justification still seems to fit Anantharaman’s bill. His point seems to be this: there is an alternative concept of rationality or justification, ‘procedural rationality’ or ‘p-rationality’, according to which only beliefs that were formed in an impeccable way count as justified. The uniqueness thesis – so we are told – would then be true when formulated with this concept of rationality.

This is how Anantharaman explicates the notion of p-rationality: “We may say that an inference is p-rational if and only if no performance errors were made.”⁶ What should count as a performance error here? Two types of error are pointed out by Anantharaman: first, you make a performance error in reasoning

⁶ Anantharaman, “Defending the Uniqueness Thesis,” 133.

about p when you reason in an invalid way about p ,⁷ second, you make a performance error in reasoning about p when you *fail* to make a valid inference that bears on the truth of p .⁸

Now, when it comes to the *ex post* status of justification that pairs with the *ex ante* status that I mentioned above, the claim that S 's belief is *ex post* justified or justifiably held is actually supposed to imply that S 's belief is also 'procedurally rational', in the sense that it was produced through good reasoning. Believing that p is *ex ante* rational for S only when S has good reasons R to believe that p and S is able to competently form a belief toward p on the basis of R . Accordingly, S 's belief that p is *ex post* justified (or S justifiably believes that p) only when S *competently* believes or holds the belief that p on the basis of good reasons R . So there is no invalid reasoning that confers *ex post* justification upon any belief: this type of performance error is excluded.

My example in favor of the moderate permissiveness thesis (*ergo*, against the uniqueness thesis) was supposed to be one in which (i) the subject is *ex post* justified in believing a proposition p (i.e., she justifiably believes that p) on the basis of R , but (ii) she could be *ex post* justified in suspending judgment about p (i.e., she could justifiably suspend judgment about p) on the basis of R as well. But the fact that a subject is *ex post* justified in holding a certain doxastic attitude entails that the relevant attitude is *ex ante* justified for that subject. So if the example really satisfies the description I just gave, it is also an example in which (iii) one is *ex ante* justified in believing that p given one's reasons R , but (iv) one could also be *ex ante* justified in suspending judgment about p given one's reasons R .

So let us look at the original example. I submit that no invalid reasoning is performed by the subject in the example – neither in the (actual) scenario of type (i), nor in the (counterfactual) scenario of type (ii). Consider the scenario of type (i) first. Here we have Amanda, in the actual world w_1 , performing a *modus tollens* type of inference from her reasons:

(1) If I clicked the wrong link, my computer has a virus now,

and

(2) My computer has no virus now,

to a belief in the conclusion:

⁷ In this context, we should understand the concepts of validity/invalidity in a broad sense, in such a way as to allow for inductive or probabilistic validity as well.

⁸ Anantharaman, "Defending the Uniqueness Thesis," 133.

(3) I did not click the wrong link.

Amanda's inference is competent in w_1 , and her premises are justified. That should lead us to conclude that Amanda's belief is rationally held in w_1 , and that there is no invalid reasoning at all going on here.

Now consider the scenario of type (ii), in a counterfactual situation w_2 . Here, Amanda has (1) and (2) available as reasons again – but now she lacks the competence to perform the relevant inference to the conclusion (3). She even considers whether (3) is the case. But as she cannot properly infer that (3) is true from (1) and (2), she cannot see why (3) must be true if (1) and (2) are. So she suspends judgment about (3). Again, there is no invalid reasoning going on here (in w_2). There would be invalid reasoning if Amanda were to infer that (3) is false – but this is not the case in w_2 . It is not like Amanda is drawing a conclusion that does not follow from/is not made probable by her available reasons in w_2 .

Notice also that Amanda's type of situation in w_2 is not at all a far-fetched type of situation. Right now there are some complicated things that follow from what you and I believe (e.g., theorems of arithmetic) that we are not able to infer to be true in a competent manner. If we were to consider them, we would suspend judgment about them, and it would be rational for us to do so. But our suspension of judgment about such complicated truths would not constitute any use of invalid reasoning: we would not be reaching conclusions that do not follow from our reasons (we would not be reaching conclusions at all!).

So the first type of performance error that was pointed out by Anantharaman – making an invalid inference – is avoided by Amanda in both the situation of type (i) (the w_1 situation) and the situation of type (ii) (the w_2 situation) of our example.

That leaves us with what would be a second type of performance error, according to Anantharaman: failing to make a valid inference. The idea here seems to be that you are in a sense irrational when (a) you consider p , (b) *there is* a piece of good reasoning that could lead you from your reasons to a belief that p , but (c) you do not go through that piece of reasoning and you fail to competently form the belief that p .

Let T again be a very complicated mathematical theorem, one that follows from certain basic axioms. Suppose you have that basic knowledge and you have all the concepts that are necessary to entertain T (they may be very simple ones). Suppose, further, that the shorter inferential path from those basic axioms to T is a very long and cunning path (although it is a finite path). Maybe only a very small percentage of humans could go through that path – or maybe even *none*! If certain theorem-proving algorithms could somehow be artificially implemented in your

brain, or maybe if you were exposed to a certain type of intense mathematical training, then you would be able to derive T from your basic mathematical knowledge. But no such algorithm is actually implemented in your brain, and you have not been exposed to any such intense training. Now I ask you to consider T , and I ask you whether T is true or false. But you have no clue as to whether T is true or false, so you suspend judgment about T .

Is your reasoning performance faulty in this case? Are you performing wrongly when you refrain from believing that T when you cannot see any reason to believe that T ? Or: is it faulty to refrain from believing that T when you just cannot tell whether T is true?

I submit that your reasoning performance *is not* faulty in this case. To be sure, you are *failing to perform* in a certain way – but that by itself is not sufficient in order to constitute a performance error. There is a sort of performance error in the vicinity, however: you are in error when you fail to perform a valid inference after considering its conclusion if, given your actual cognitive state, you could have performed that inference and reached that conclusion (but you suspended your judgment about that conclusion instead). Using the notions of possible-worlds semantics in order to interpret the modal term ‘could’ here, the qualification ‘given your actual cognitive state’ is supposed to put a constraint on the set of possible worlds that count as accessible from the actual one. The relevant relation of accessibility at play here – call it ‘cognitive accessibility’ – is such that a world v counts as accessible from a world w only when the reasoner’s inferential abilities in v and in w are the same.

But since we are assuming you could not have performed an inference from your basic mathematical knowledge to T in that sense, we cannot attribute to you that sort of performance error. The same applies to Amanda in the situation of type (ii) (the w_2 situation). Amanda does not have the ability to perform a *modus tollens* type of inference – she is ‘blind’ to conclusions of inferences of that type.

So failing to perform a valid inference constitutes a performance error only when there is a cognitively accessible situation (accessible from the actual one) in which the subject performs that inference. The cognitive resources available to Amanda in w_2 , however, will not allow her to perform the relevant *modus tollens* inference in any situation that is cognitively accessible to her. It would follow, then, that she does not make a performance error in w_2 . So she would also be ‘procedurally rational’ in w_2 . So, contrary to Anantharaman’s suggestion, the example also shows that the uniqueness thesis is false even when it is formulated with a ‘procedural’ notion of rationality.

KNOWLEDGE BASED ON SEEING

Mark SCHROEDER

ABSTRACT: In *Epistemological Disjunctivism*, Duncan Prichard defends his brand of epistemological disjunctivism from three worries. In this paper I argue that his responses to two of these worries are in tension with one another.

KEYWORDS: Duncan Prichard, epistemological disjunctivism, knowledge, basing

In his wide-ranging and ambitious *Epistemological Disjunctivism*, Duncan Pritchard outlines and defends what he calls the “holy grail” of epistemology – a view that aims to combine the virtues of both internalist and externalist approaches in epistemology, and which claims to offer a novel, robust, and “satisfying” response to the problem of radical skepticism. The only problem with this view, Pritchard notes, is that it “occupies a region of logical space in epistemology that many hold is simply unavailable.”¹ According to Pritchard’s diagnosis, there are three chief *prima facie* problems for his brand of epistemological disjunctivism that have seemed to make it unavailable, and his defense is constituted by replies to these three problems, in addition to an elaboration of its virtues. But unfortunately for the search for epistemology’s holy grail, in this paper I will show that Pritchard’s responses to two of the three problems facing his preferred form of epistemological disjunctivism are in tension.

1. The Setup: What Is Epistemological Disjunctivism?

The view that Prichard calls epistemological disjunctivism, he formulates as follows:

Epistemological Disjunctivism: The Core Thesis

In paradigmatic cases of perceptual knowledge an agent, S, has perceptual knowledge that ϕ in virtue of being in possession of rational support, R, for her belief that ϕ which is both factive (i.e., R’s obtaining entails ϕ) and reflectively accessible to S.²

¹ Duncan Prichard, *Epistemological Disjunctivism* (Oxford: Oxford University Press, 2013), 18.

² Pritchard, *Epistemological Disjunctivism*, 13. After stating this principle using ‘ ϕ ’ as a schematic letter for propositions on the first page of his first chapter, Pritchard goes on to never use ‘ ϕ ’ again, instead using ‘*p*.’ I’ll follow him in this.

What Pritchard means by this is best illustrated by considering a specific case of perceptual knowledge, such as that in which you know that this paper is in front of you. In the paradigmatic case, you start by seeing that this paper is in front of you, and at least partially in virtue of that fact, the fact that you see that this paper is in front of you is reflectively accessible to you – meaning that you can know it by reflection alone. Call this fact *R*.

So in virtue of being in a position to know that you see that this paper is in front of you, you possess *R* as rational support – as your *reason* for your perceptual knowledge that this paper is in front of you. And it is a particularly excellent reason, because it actually *entails* that this paper is in front of you (since you cannot count as seeing that this is true unless it really is true). So in paradigmatic cases of perceptual knowledge, your knowledge is supported by reasons of the very best sort, and these reasons are not merely some external facts about reliability or safety that are epistemically inaccessible to you and hence unavailable in responding to radical skepticism, but rather they are your very own reasons – accessible to you on the basis of reflection alone. Hence, Pritchard's view has much in common with classical forms of internalist foundationalism such as those of C.I. Lewis and Roderick Chisholm, on which the basic perceptual reasons are facts about your subjective psychological experience that are presumed to be luminous, incorrigible, or otherwise knowable on the basis of reflection alone, except that Pritchard's view claims that in paradigmatic cases, what happens is that there is a *world-implicating* fact – that you factively *see that p* – that is knowable by reflection alone. And this gives Pritchard's view a whole set of advantages that are unavailable to classical forms of internalism.

What makes this view count as a kind of *disjunctivist* view about perceptual knowledge is that as Pritchard interprets the view, the kind of rational support that the *core thesis* describes is not available in cases in which you are faced with an illusion or a hallucination. Although this isn't obvious from the formulation of the *core thesis* by itself, it follows from Pritchard's interpretation that for a piece of rational support to be *reflectively available* to you, it must be possible for you to know it by reflection alone. Since knowledge is factive, knowability by reflection alone is presumably factive,³ and so you cannot have this rational support for your belief that *p* unless you *see that p*. In the philosophy of perception, disjunctivism about perceptual experience is the view that seeing that *p* is a state that is different in kind from states that are subjectively indiscriminable from it, such as suffering a visual illusion as of *p* or having a hallucination as of *p*. So by analogy, the

³ Perhaps some things that are not true can be known by reflection alone, such as *that you are reflecting*, but *that you see that ϕ* is not among them.

consequence of Pritchard's view that the rational support available for perceptual beliefs is different in kind in these cases is naturally termed a kind of *epistemological* disjunctivism, and as Pritchard notes, though there is no entailment between perceptual and epistemological disjunctivism in either direction, the two views may be natural allies.

I'll note in passing that it is probably not good practice in nomenclature to name views by their consequences rather than by their core theses. Pritchard's view earns the name of *epistemological disjunctivism* because it holds that the form of rational support available for perceptual beliefs differs between the good and bad cases, but he accepts this thesis for very particular reasons that need not be shared by everyone who holds that the form of rational support available in the good and bad cases differ. For example, on one sort of view, in virtue of seeing that *p*, you count as having *that p* available as your reason to believe that *p*.⁴ This view allows a form of rational support that is only available in the good case, but it does not carry Pritchard's commitment that the fact that you see that *p* is reflectively available. Similarly, on another sort of view, when you see that *p*, the fact that you see that *p* counts as part of the rational support that you have for *p* regardless of whether it is reflectively accessible to you – just because it is true.⁵ This view again allows a form of rational support that is only available in the good case, but again, it does not carry Pritchard's commitments about reflective availability. So the term, 'epistemological disjunctivism' is probably better reserved as a name for what these different views have in common. Hence I'll continue in what follows to refer to *Pritchard's view* and to the *core thesis*, or to *Pritchard's form of epistemological disjunctivism*.

⁴ This appears to be Timothy Williamson's view; he holds that your evidence is what you know and that seeing that *P* entails knowing that *P*, so in virtue of seeing that *P*, you come to have *P* as part of your evidence. Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press: 2000).

⁵ This is how Matthew McGrath and Juan Comesaña interpret John McDowell's view, especially in John McDowell, "Avoiding the Myth of the Given," in *Experience, Norm, and Nature*, ed. John McDowell and J. Lindgaard (Oxford: Blackwell Publishing, 2009). See especially Matthew McGrath and Juan Comesaña, "Perceptual Reasons," *Philosophical Studies* Online First, <http://link.springer.com/article/10.1007%2Fs11098-015-0542-x>. In his book, Pritchard distinguishes on pages 36–37 between *accessibilist* and *mentalist* forms of epistemological internalism, and defends the *core thesis* as satisfying a form of accessibilism. In that context, it is natural to interpret the view McGrath and Comesaña attribute to McDowell as the mentalist analogue of Pritchard's core thesis.

2. The Basing Problem

The simplest problem that Pritchard considers for his core thesis is posed by the idea that seeing that *p* entails knowing that *p*. This idea is, Pritchard suggests, accepted by “most views,”⁶ and it is certainly a consequence of Timothy Williamson’s familiar thesis that knowledge is the most general factive stative attitude.⁷ But if seeing entails knowing, then it is hard to see how the fact that you see that this paper is in front of you could be your *basis* for believing that this paper is in front of you.

Pritchard’s solution to this problem is to reject – rightly, I believe – the principle that *seeing* entails *knowing*. His chief counterexample to this entailment involves a case in which you have *misleading evidence*. For example, if you are driving through normal countryside but have rationally come to believe that you are in fake barn country, then when you look at the barn directly in front of you, if you are rational, then you will see that there is a barn, but you will not *know* that there is a barn, because you will not *believe* that there is a barn. And indeed, even if you irrationally do believe that there is a barn, still the fact that this is an irrational belief will prevent it from being knowledge.

I have no issue with this counterexample; I myself have actually offered the same sort of case as a counterexample to the thesis that seeing entails knowing. And indeed, I believe that there are other sorts of counterexamples. For example, your perceptual experience represents far more things than you actually form beliefs about. And in the good case, your visual relationship to all of these things that are represented is factive – a kind of visual success. So you see them to be the case, but since you don’t believe them to be the case, you don’t know them to be the case. And similarly, even when you do form beliefs on the basis of your perceptual experiences, those beliefs are always formed some time *after* the perceptual experience. So in the intervening time, while you are in the process of forming the belief on the basis of what you see, you see that *p* without believing that *p*, and hence without knowing that *p*.

Indeed – and this will be important later – I believe that it is a conceptual point that if *R* is the *original basis* of one’s belief that *p*, then one must have been in possession of *R* *before* forming the belief that *p*. It is possible, of course, to shift

⁶ Pritchard, *Epistemological Disjunctivism*, 21.

⁷ Though see John Turri, “Does Perceiving Entail Knowing?” *Theoria* 76 (2010), 197–206 for a dissenting opinion. Turri’s counterexamples to the *sees entails knows* principle are inconsistent with the judgments about seeing in fake barn cases that Pritchard relies on, and I myself prefer Pritchard’s reasons for rejecting the principle (see Mark Schroeder, “Knowledge is Not the Most General Factive Stative Attitude,” unpublished manuscript).

the basis for one's belief that p – for example, you may start by believing that p for the reason that S , and then come to learn that R , which also supports p . If you then learn that you were wrong about S , or simply forget S , R may then be the basis for your belief that p even if you were not in possession of R before you formed the belief that p . But this is possible only because you had some other original basis for your belief, which has shifted. The conceptual point is that if R is the *original* basis of your belief that p , then you must have been in possession of R *before* forming the belief that p .

Pritchard does not explicitly endorse this conceptual principle about basing. But in the absence of such a principle, it is hard to see why he should think that it needs to be possible to see that p without knowing that p , in order to make sense of how it is possible for the fact that one sees that p to be one's *basis* for one's knowledge that p . The conceptual point about temporal priority of original basing explains *why* the basing problem is even a *prima facie* problem for Pritchard's thesis, *and* it explains why rejecting the principle that seeing entails knowing would help to address this problem. And it follows from this principle that it is possible to see without knowing even in cases in which you have no misleading evidence and are in a position to know – but simply don't know, yet. This turns out to be in direct tension with Pritchard's solution to a second *prima facie* problem for his core thesis.

3. The Access Problem

The second serious problem for his core thesis that Pritchard takes up is what he calls the *access* problem. The access problem is a worry to the effect that if it is really possible to know by reflection alone that one sees that p , as Pritchard's core thesis claims, then it must be possible to know by reflection alone that p . But p is just a paradigmatic ordinary empirical proposition that is directly perceptually observable. Such propositions are not knowable on the basis of reflection alone, and so that is good reason to think that the core thesis must be false – it can't be possible to know that one sees that p on the basis of reflection alone.

Pritchard first imagines that this problem is posed for an agent who already has empirical knowledge. But he notes that if it is only possible for an agent who already has empirical knowledge that p to know that p on the basis of reflection, that does not really show that it is possible to know that p on the basis of reflection *alone*. And so in order to constitute a serious objection, the access problem must focus on cases in which the agent does *not* have knowledge that p . Pritchard formulates this more formidable version of this objection as follows:

The Access Problem

(AP1'') *S* can know by reflection alone that she is in possession of the factive reason *R* for believing the specific empirical proposition *p* (although she does not believe that *p* on this basis, or on any other basis). [Premise]

(AP2'') *S* can know by reflection alone that *R* entails *p*. [Premise]

(APC'') *S* can know by reflection alone the specific empirical proposition *p*.
[From (AP1''), (AP2'')]

As Pritchard notes, this argument is valid, and the second premise is exceedingly hard to deny. Moreover, as he further notes, on his own view it is possible to see that *p* without knowing that *p*. And the core thesis says that in paradigmatic cases, when an agent sees that *p*, the fact that she sees that *p* is *reflectively available* to her. So this is what leads to the impression that Pritchard must accept the first premise.

But Pritchard claims that the first premise is false. This, he claims, is because the *only* cases in which it is possible to see that *p* without already knowing that *p* are cases in which one has misleading evidence – for example, that one justifiedly believes that one is in fake barn country even though one is not. But in these cases, he claims, obviously you cannot know by reflection alone that one sees that *p* – for example, that there is a barn in front of you. So, he claims, the *very cases* in which the access problem could possibly pose a threat are cases in which misleading evidence makes the first premise false. Hence, he concludes, the only cases in which an agent has a reflective justification available for an empirical proposition *p* are ones in which she knows that *p* on independent empirical grounds. So the fact that there is a reflective justification available in these cases does not, he holds, undermine the truism that such truths can only be known empirically.

4. The Clash

In order to maintain his answer to the access problem, Pritchard must claim that the *only* counterexamples to the principle that seeing entails knowing are cases in which the subject is in possession of misleading evidence – that is, that they are all cases in which the subject sees but is not even in a *position* to know. But as we saw earlier, the conceptual point that lies behind the basing problem *presupposes* that there are *other* cases in which one sees without knowing. And that is because, according to the conceptual point, nothing can be the original basis for one's belief unless one had it *before* one formed the belief. So in particular, that one sees that

p cannot be the basis for one's belief that p unless one sees that p before one knows that p . But in all such cases of successfully basing belief on what one sees, there is a time period during which one sees that p and is in a *position* to know that p , but does not yet know that p . And during this time, both premises of the argument formulating the access problem are *true*.

As I've argued, in the absence of the conceptual point about priority in basing, it is very hard to see why the claim that seeing entails knowing would pose even a *prima facie* problem for Pritchard's core thesis. So in principle, Pritchard could reject the conceptual point and take it as a point in his favor that it turns out that the basing problem is not, after all, even a *prima facie* problem for his view. But I think this would be a mistake. The claim that original basing requires temporal priority is central to any plausible understanding of what basing is or why it is epistemologically significant. Without this conceptual point, I worry that we would lose enough grip on what makes basing important to make it very unclear why Pritchard's core thesis has anything enlightening to say about perceptual knowledge or about skepticism. So I conclude, instead, that Pritchard is on the right track to solving the basing problem, but on the wrong track to solving the access problem. And this should be no surprise; plausibly the access problem is the most central reason why the view that Pritchard describes in his book has been widely perceived as occupying an "unavailable" portion of logical space.⁸

⁸ Special thanks to Conor McHugh, Clayton Littlejohn, Hagit Benbaji, and Ram Neta.

NOTES ON THE CONTRIBUTORS

Fred Adams is Professor of Linguistics & Cognitive Science and Professor of Philosophy at the University of Delaware. He has published well over 100 articles and chapters on cognitive science, epistemology, philosophy of language, philosophy of action, and related topics. He is also co-author of *The Bounds of Cognition* (Wiley-Blackwell, 2008, with Ken Aizawa) and of *An Introduction to the Philosophy of Psychology* (Cambridge University Press, 2015, with Daniel Weiskopf. Contact: fa@udel.edu.

Rodrigo Borges got his PhD in philosophy from Rutgers in 2015 and has recently started a faculty position at Pontifical Catholic University of Rio Grande do Sul (PUCRS). His work focuses mainly in epistemological issues. He is the co-editor, with Peter Klein and Claudio de Almeida, of *Explaining Knowledge: New Essays on the Gettier Problem* (Oxford University Press, forthcoming). Contact: epistemen@gmail.com.

Murray Clarke is Professor in the Department of Philosophy at Concordia University in Montreal, Quebec. His research interests span contemporary epistemology, cognitive science and philosophy of science. He is the author of *Reconstructing Reason and Representation* (MIT, 2004). He has published numerous articles including “Resurrecting the Tracking Theories” (*Australasian Journal of Philosophy*, 2005), “Concepts, Intuitions and Epistemic Norms” (*Logos & Episteme*, 2010), and “Critical Notice: Jose Zalabardo’s Scepticism and Reliable Belief” (*Teorema*, 2014). Contact: Murray.Clarke@concordia.ca.

Mihai Hîncu is a Postdoctoral Research Fellow (SOP HRD/159/1.5/S/133675 Project) at the Romanian Academy, Iași Branch. He received his PhD in Philosophy from the University of Bucharest in 2012, with a thesis on the unarticulated constituents and the semantics of belief reports. His main areas of specialization are logic, formal semantics, and philosophy of language. Also, his research interests include formal epistemology, decision theory, game theory, and metaethics. Contact: mihaihincu@gmail.com.

Moti Mizrahi is an Assistant Professor of Philosophy at the Florida Institute of Technology. He received his PhD from the Graduate Center of the City University of New York. He has teaching and research interests in argumentation theory,

epistemology, ethics, metaphysics, philosophy of religion, and philosophy of science. His most recent publications include “Three Arguments against the Expertise Defense” (*Metaphilosophy*, 2015), “Does Conceivability Entail Metaphysical Possibility?” (*Ratio*, 2015, with David R. Morrow), “Kuhn’s Incommensurability Thesis: What’s the Argument?” (*Social Epistemology*, 2015). He has forthcoming papers in *Argumentation*, *International Journal for the Study of Skepticism*, *International Studies in Philosophy of Science*, and *Journal for General Philosophy of Science*. Contact: motimizra@gmail.com.

Alexander Pruss is Professor of Philosophy at Baylor University in Waco, Texas, and works in metaphysics, philosophy of religion, formal epistemology, applied ethics and philosophy of mathematics. Contact: Alexander_Pruss@baylor.edu.

Luis Rosa is a Humboldt post-doctoral fellow at the Munich Center for Mathematical Philosophy (LMU Munich). His work concentrates on issues concerning epistemic rationality, epistemic logic and philosophy of psychology/AI. Contact: Luis.Rosa@lrz.uni-muenchen.de.

Mark Schroeder is Professor of Philosophy at the University of Southern California. He works widely on topics connected with epistemology, metaethics, normative ethics, and the philosophy of language, and is the author of *Slaves of the Passions* (Oxford University Press, 2007), *Being For: Evaluating the Semantic Program of Expressivism* (Oxford University Press, 2008), *Noncognitivism in Ethics* (Routledge, 2010), *Explaining the Reasons We Share* (Oxford University Press, 2014), and *Expressing Our Attitudes* (Oxford University Press, 2015). He is currently at work on a book, *Reasons First*, exploring, *inter alia*, how knowledge might be grounded in factive relations to the world without a form of epistemological disjunctivism being true. Contact: maschroe@usc.edu.

Michael J. Shaffer is a professor of philosophy at St. Cloud State University. He is the author of *Counterfactuals and Scientific Realism* (Palgrave, 2012) and *The Experimental Turn and the Methods of Philosophy* (Routledge, forthcoming). He has published numerous articles on epistemology, logic and the philosophy of science. Contact: mjshaffer@stcloudstate.edu.

John N. Williams is an associate professor in philosophy in the School of Social Sciences at Singapore Management University. His research interests include paradoxes, epistemology and applied ethics. His most recent articles include: “The Backward Clock, Truth-Tracking, and Safety” (with Neil Sinhababu, *Journal of*

Philosophy, 2015), “Not Knowing You Know: A New Objection to the Defeasibility Theory of Knowledge” (*Analysis*, 2015), “Moore’s Paradox in Speech: A Critical Survey” (*Philosophy Compass*, 2015), and “Moore’s Paradox in Thought: A Critical Survey” (*Philosophy Compass*, 2015). Contact: johnwilliams@smu.edu.sg.

***LOGOS & EPISTEME*: AIMS & SCOPE**

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

NOTES TO CONTRIBUTORS

1. Accepted Papers

The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

2. Submission Address

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*.rtf) or Microsoft Word document (*.docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author's affiliation and professional address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it. The review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers. The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are

received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address:

Eugen Huzum

Institutul de Cercetări Economice și Sociale „Gh. Zane”

Academia Română, Filiala Iași

Str. Teodor Codrescu, Nr. 2, 700481

Iași, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Property & Royalties

Articles accepted for publication will become the property of Logos & Episteme and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay

royalties. Authors of accepted manuscripts will receive free of charge one copy of the issue containing their papers.

14. Permissions

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.