2022, 8 (1)

ARGUMENTA

The Journal of the Italian Society for Analytic Philosophy

First published 2022 by the University of Sassari

Produced and designed for digital publication by the *Argumenta* Staff

Argumenta is the official journal of the Italian Society for Analytic Philosophy (SIFA). It was founded in 2014 in response to a common demand for the creation of an Italian journal explicitly devoted to the publication of high quality research in analytic philosophy. From the beginning *Argumenta* was conceived as an international journal, and has benefitted from the cooperation of some of the most distinguished Italian and non-Italian scholars in all areas of analytic philosophy.

Contents

Editorial

While we might easily understand that the word "context" can be said in many ways, it is much more difficult to discern which of these many ways are appropriate. In the Introduction to the Special Issue entitled *Context*, Diana Mazzarella, Antonio Negro, and Carlo Penco come to our aid. Indeed, under the witty title "Everything You Always Wanted to Know about Context (But Were Afraid to Ask)"—calling to mind Woody Allen's acumen—they provide a broad and accurate overview of the multiple areas of study in which the notion of context is crucial.

Above all, the eight essays that make up the Special Issue present clear and innovative ways of addressing the notion of context in some of the areas outlined in the Introduction, thus demonstrating how vital and valuable this notion can be for philosophical analysis.

The present number also includes three articles that have already appeared in 'early view' (by Alberto Barbieri, Costanza Larese, and Daniel Rönnedal), and that have already made and will continue to make significant contributions to discussion in their respective fields.

The number is then rounded off by the section of Book Reviews. We are proud to offer readers three new thoughtful reviews of as many interesting books.

Finally, I would like to thank all the colleagues who have acted as external referees, the members of the Editorial Board, the Editors of the Special Issue, the Editors of the Book Reviews, and the Assistant Editors. All of them have been very generous with their work, advice, and suggestions. In particular, the

Argumenta 8, 1 (2022): 3-4
ISSN 2465-2334

team of assistant editors proved to be super helpful and tireless as always. I thank them from the bottom of my heart because they, above all, embody the vitality of the journal.

As usual, the articles appearing in *Argumenta* are freely accessible and freely downloadable, therefore it only remains to wish you:

Buona lettura!

Massimo Dell'Utri
Editor-in-Chief

Argumenta 8, 1 (2022)
Special Issue

Contexts

Edited by

Diana Mazzarella, Antonio Negro,
Carlo Penco

The Journal of the Italian Society for Analytic Philosophy

Contents

Contexts: Everything You Always Wanted to Know about Context (But Were Afraid to Ask)

Diana Mazzarella, Antonio Negro,** Carlo Penco****

* *University of Neuchatel*

** *FINO North Western Doctoral Consortium*

*** *University of Genoa*

Abstract

We give a short recap of how the notion of context has been developed in the philosophy of language since its introduction by Frege. We introduce various aspects of the concept of context: context of utterance, context at the semantics-pragmatics boundary, and social and cognitive context. We thereby offer to readers not accustomed to the distinctions used in the philosophy of language a framework to better understand the papers enclosed in this issue (and of which we provide summaries at the end of this introduction).

Keywords: Context of utterance, Common ground, Semantics-pragmatics boundary, Social philosophy of language, Cognitive context.

1. Introduction

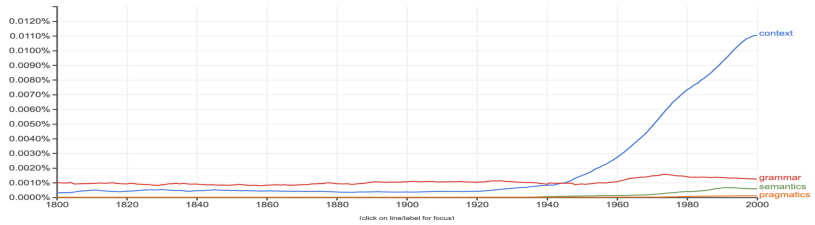
As Domaneschi and Penco (2013) say, context is said "in many ways", like being for Aristotle. The topic of context in philosophy, as it is discussed today, was born with Frege's principle of contextuality, extensively discussed by Michael Dummett (1973, 1981), and developed with the seminal works by David Kaplan (1972, 1989a, b) and Robert Stalnaker (2002, 2014), who used different meanings and formalisations of the notion of context. The series of conferences on Modelling and Using Context, started by Patrick Brezillon in 1997, showed the many sides of the notion of context, which affects philosophy as much as artificial intelligence, psychology, or sociology.[1] Certainly, the term "context" has had a great

[1] For an early analysis of different concepts of context from a computer scientist's viewpoint, see Hayes 1997 or Brezillon and Turner 2021. For a list of the CONTEXT conferences

Argumenta 8, 1 (2022): 9-33
ISSN 2465-2334

DOI 10.14275/2465-2334/202215.maz

development in many areas of study in the second half of the XX century. The diagram below from the Ngram viewer shows the incredible increase in the use of the term "context" when compared to terms like "semantics", "pragmatics", or "grammar".



It would be difficult to speak of "context" without any specification of the kind of cultural or academic environment in which the term is used. For example, we frequently hear that artificial intelligence is "all about context",[2] but here too we can refer to a variety of applications. Even if we focus on the discussions of philosophers, we cannot come to a consensus about which notion of context is best suited for which theoretical purpose. Furthermore, the explanatory power of the notion of context is often questioned. It is obvious that when we say, "I have nothing to put on", the term "nothing" is to be understood under a contextual restriction. At first sight, one may think that the idea of universal quantification has no meaning without contextual restrictions. However, Williamson (2003) challenges this proposal and shows the viability of a general notion of non-contextualized quantifiers, leaving to pragmatics the problem of defining the contextual aspect. The debate on which role the notion of context should have (in syntax, in pragmatics, in semantics) is still alive. The notion of context has immediate applications in ethics, epistemology, and philosophy of language, and we already have many reference works (e.g., Preyer & Peter 2005, 2007, Pynn 2016, Penco & Vignolo 2020, Ciecierski & Grabarczyk 2020). While there is much discussion in ethics about the contextual dependence of ethical statements, a clear thematization of the concept of context remains lacking, as does the abundant discussion of contextualist epistemology. It seems that a clearer definition of the concept of context has been mainly developed in the philosophy of language and in linguistics, from which it is applied elsewhere. The first important distinction is the one between linguistic context and extralinguistic context.

Linguistic Context: Linguistic context refers to what precedes or follows a string of text. It is what helps understand the meaning of words in the context of a sentence or the meaning of a sentence in the context of a wider part of the text, with its anaphoric links. We speak in these cases of sentence context or discourse context. The work of Hans Kamp on *Discourse Representation Semantics* (Kamp & Reyle 1993) brought discourse context and the resolution of anaphora to the attention of both linguists and philosophers.

Extralinguistic Context: Extralinguistic context typically refers to both physical and cognitive features of a given situation. It is, first of all, what helps interpret the referents of indexicals and demonstratives. It also helps to understand the

from 1997 to 2019 see http://context19.disi.unitn.it/index.php/context-conferences-and-journal/

[2] See, e.g., Singer 2022.

various presuppositions that underpin a speech, as well as the various moves that allow the discourse to develop and update. The two main concepts of (extralinguistic) contexts derive from Kaplan's logic of demonstratives and Stalnaker's theory of presuppositions. In the former, we speak of the context of utterance, basically defined by time, place, speaker, and possible world. In the latter, we speak of cognitive context, or common ground, which is the set of presuppositions or propositions accepted as true by participants in a conversation (the context set).

Besides the broad distinction between linguistic and extralinguistic contexts, it is worth looking at some finer-grained notions of context discussed in the literature: opaque contexts, dynamic contexts and passive/active contexts.

Opaque Contexts: Opaque contexts are those linguistic contexts in which substitutivity breaks down. They include modal contexts, belief contexts, and other kinds of indirect contexts. Since Leibniz and Frege, the rule of substitutivity of identicals (the semantic counterpart of Leibniz's law of indiscernibles) was taken as a central rule in logic: we may substitute an expression with another with the same denotation and the truth value of the sentence does not change. From

(1) Hesperus is a planet

and

(2) Hesperus = Phosphorus,

we may derive

(3) Phosphorus is a planet.

Consider, though, how things change if we embed (1) inside a belief context like

(4) Pia believes that Hesperus is a planet.

I cannot conclude from (2) and (4) that Pia believes that Phosphorus is a planet because she might ignore that Hesperus = Phosphorus, and therefore she may rationally believe (1) and at the same time deny (3). Hundreds of papers discussed the problem of what happens when the golden rule of substitutivity fails, from Carnap 1956 to Evans 1982, or Kripke 1979, either challenging or developing Frege's view according to whom substitutivity in doxastic indirect contexts works by substituting expressions with the same sense and not just with the same reference (see also Kripke 2008). This topic also has a counterpart in epistemic logics, with an interplay of philosophical problems and artificial intelligence (on which see Frixione 1994 and Wang 2018).

Dynamic Contexts: Following Stalnaker, Irene Heim (1992) developed *Update Semantics*, which is based on the idea of meaning as *context change potential*. The meaning of a sentence has the power to change the context set, selecting which possible worlds must be cut off from the set of shared presuppositions. Semantics, therefore, must explain the dynamic of changes in discourse. Developments in dynamic semantics propose different formalisms with respect to both the context of utterance and the discourse context.

*Passive and Active Context*s: Kripke (2011) suggests a general distinction between passive and active contexts. The passive context, also referred to as the global context, consists of general background information available to the speakers. The active context, or local context, includes salient questions or topics, as well as assertions, relevant to the specific conversation in which the speaker and

the hearer are engaged.[3] It is the active context that helps the hearer understand presuppositions triggered by expressions like "again", "too", and so on.[4]

Given the breadth of the subject, we have opted to provide some opening notes to assist the reader in understanding the theoretical and historical background of the papers in this issue, with reference to three main areas: (i) Utterance context and demonstratives, (ii) Contextual dependence and the semantic-pragmatics divide, (iii) Cognitive and social context.

2. Context of Utterance: Indexicals and Demonstratives

The concept of context was only informally present in Frege's writings and had different connotations. One particular development of the notion of context found its first assessment in the philosophical environment of the University of California, Los Angeles (UCLA), where Hans Reichenbach was appointed in 1938 (with the help of Charles W. Morris) and Carnap arrived from the University of Chicago in 1954, the year after Reichenbach died, to be joined by Alonzo Church in 1967. The three philosophers and logicians extensively debated Fregean notions, although only Reichenbach (1947) attempted to present a formalisation of indexicals to which Frege had devoted much attention in his late paper "The Thought" (1918).

2.1. The Origin of the Formalization of Context in Logic

We find the first attempt to formalise the notion of context with Richard Montague (1930-1971), who studied with Alfred Tarski in Berkeley and later arrived at UCLA.[5] Under the influence of Carnap and Church,[6] Montague fully developed the first instance of intensional semantics: the semantic values of the expressions (e. g. names, predicates and sentences) are functions from possible worlds to extensions of the appropriate kinds (objects, classes and truth values). Crucially, though, Montague remarked that the semantic values of some expressions, like temporal or spatial adverbs ("now" or "here") or pronouns ("I", "he", "she" ...), are dependent not only on possible worlds, but also on other contextual factors like the speaker or the time and location of the utterance. Therefore, he further elaborated his original idea and proposed to consider intensions as functions from an *index* to an extension, where the index can be considered as the first formal representation of the idea of context. The index is a sequence of all the factors on which the semantic value of expressions depends (which for Montague included possible worlds, times, individuals and their role as speaker or addressee).

Montague intensional semantics had two problems: (a) a function from an index to a truth value underestimates the contribution of lexical meaning. Meanings cannot be reduced to intensions; (b) compositionality becomes impossible

[3] See Kripke 2011: Chapter 12.

[4] On the general discussion of the relations between global and local contexts, see Schlenker 2010, where the distinction concerns the problem of presupposition projection. Apparently, the terms "local context" and "global context" may be used in different settings, like social theories or artificial intelligence systems, and we do not consider the variety of these applications here.

[5] See the Introduction by Richmond Thomason of Montague (1974: 41).

[6] In his 1970 paper "Pragmatics and Intensional logic" Montague (1974: 145) recalls Carnap suggesting in conversation that "intensional object should be identified with function from possible worlds to extensions of the appropriate sources".

with quantification in modal or temporal contexts. Hans Kamp, one of Montague's students, and David Kaplan, Carnap's last PhD student, proposed double indexing as a solution to these problems. In what follows, we discuss them in turn.

To gain an intuitive understanding of the first problem, Kaplan (1976: 82) points out the distinction between the following utterances:

(5) I am here now.

(6) David Kaplan is in Los Angeles on April 21, 1976.

According to Kaplan, (5) is logically true, since, for any context, it is true in the world of the context, while (6) is not logically true. Gillian Russell (2008: 56) claims that (5) is a typical case of an analytic sentence that can be said to be "true in virtue of meaning", while this is not the case for (6). This distinction cannot be made with the index introduced by Montague. How to address this problem?

The answer—called "double indexing"—was to divide Montague's index into two different components: on the one hand, the context of utterance (made up of the relevant parameters like speaker, time, location and possible world), and on the other hand, a circumstance of evaluation (a possible world and a time). We need to distinguish between character and content. The character is the meaning or linguistic rule attached to linguistic expressions. In the case of indexicals like "I", "here" and "now" (and so on) the character is fixed, differently from the meaning of ambiguous expressions like "bank". However, the content of indexicals depends on the context of utterance, namely, content is always taken with respect to a given context. Contents are functions from possible worlds to extensions (the content of "I" is the function that for each possible world points to the individual who is the speaker in the context in which the sentence is uttered). To give an interpretation of "I am here now", then, I need first to understand speaker, time and location of the utterance. Only after the context provides the content of "I", "here", and "now", I can use the standard evaluation function from possible worlds to extensions. This formal aspect of Kaplan's theory is often connected with the direct reference theory, for which indexicals are direct reference devices like proper names (on this aspect see Martì 2022).

The second problem (about compositionality and quantification) concerns the use of modal or temporal operators. Think of

(7) Once, everyone now alive hadn't yet been born.[7]

The restriction "now alive" is inside the scope of the quantifier, but at the same time it takes the property of its subjects (being now alive) outside the scope of the temporal operator "once". The restriction concerns people alive at the time of the context of utterance but not alive in the context of the temporal operator (at the time when they were not yet born). We cannot make these distinctions with a single index à la Montague, where possible worlds, time, location, and speaker are all mixed together. We need to distinguish the context of the utterance from the context of evaluation given by the temporal operator, or, in other cases, by modal operators, like the following:

(8) Possibly, everybody now writing their papers could be at the beach swimming.

[7] We use here an example from Stalnaker 2014: 19. Kamp (1971), who introduced the idea of double indexing, used "now" as a sentential operator, before adhering to the most common view of using "now" as an indexical constant (see Kamp 2013: 7). Kamp gives abundant reference to the influence of Reichenbach on his view of temporal logic.

In this case, the restriction "now writing their papers" is inside the scope of the quantifier "everybody" but outside the scope of the modal operator "possibly", because—in the possible worlds where they are at the beach—they are not writing but swimming. The solution, again, is to distinguish character and content, where the character of "now" lets the property of *writing their papers* apply to the individuals in the possible world of the context of utterance, while the property of *being at the beach* is applied to the same individuals in the possible world of the circumstance of evaluation.

2.2. Kaplan's Standard View of the Context of Utterance

Kaplan (1989a, b) developed the logic of demonstratives where he gave both an intuitive presentation of the concept of context and a formal one. The intuitive idea is that utterance context is a representation of a situation, listing as parameters the central aspects of a situation in which a speech happens, mapping the traditional idea of the deictic centre given by Karl Bühler. For Bühler (2011: 102-103) the deictic field has its centre in the three deictic words, I, here and now, that form the coordinate system in which "all partners in communication are and remain caught up". The deictic centre maps what standard view calls "automatic" indexicals ("I", "here", "now") because uttering them automatically fix the reference.

Kaplan distinguishes pure indexicals and demonstratives, and this distinction is well represented in a schema proposed by John Perry 1989b[8] (before presenting a theory alternative to Kaplan's):

	Narrow Context <s,l,t>	Wide Context
Automatic	I, Now*, Here*	Tomorrow, Yesterday, …
Intentional	Now, Here	This man, There, This, That

We have here a perspicuous representation of the context-character-content (CCC) theory of indexicals, running between the two extremes of the spectrum. With pure indexicals ("I", "now", "here"), narrow context (time, location, and speaker) is sufficient for the individuation of the referent. However, with "here" and "now", we may also have different conceptions of the time and place ("here where I am", "here in this room", "here in this town", and so on—or "now at this moment", "now at this time of the year", "now in this historical period", and so on), as well as demonstrative uses, as when I point on a map and say "we will

[8] John Perry (1979) brought new arguments for the essentiality, or ineliminability, of indexicals. Simple examples show that beliefs or knowledge obtained through indexical expressions are not reducible to beliefs or knowledge of objective states of affairs: an amnesiac reading a biography of himself may learn a lot without realising that the biography is about him. If JP is in a supermarket and sees a trail of sugar on the ground and thinks "he, who is pouring sugar, is very stupid", he will go around looking for that idiot. But when he realises "I am pouring sugar", he will change the position of the sugar box to avoid pouring sugar idiotically from his cart. Although the proposition expressed in the context is the same in all cases—that is the ordered pair <JP, pouring sugar>—having thought in terms of "I" instead of "he" changes the content of the belief and the consequent behaviour.

arrive here…". On the other end of the spectrum, bare demonstratives ("this", "that") require a wide context (everything useful to fix the content, like a demonstration) on whose background to understand the referential intention of the speaker. Kaplan (1989b: 583) considered demonstrations as "a mere externalisation of a perceptual intention": the intention of the speaker is what determines the referent of a demonstrative.

The definition of context of utterance given by Kaplan has been a standpoint in the development of the concept of context. Different notions of context are often defined with respect to their similarity or difference from Kaplan's idea of the context of utterance. We hint here at some alternatives. To begin with, David Lewis (1980) accepted the idea of double indexing but thought that we do not need a two-stage procedure to pass from context to content. He agreed that we need a context (speaker, time, location, world of utterance) and an index (possible world and time) but considered Kaplan's idea of what is said redundant and suggested instead having a function from context-index pairs to extensions. Stefano Predelli (2005, 2013) partly follows Lewis' ideas.

Stalnaker (2014) identified some problems with Lewis' solutions and at the same time discussed the relation between Kaplan's context and his view of common ground. On the one hand, he claims, common ground or the context set of presuppositions taken for granted by the interlocutors in a dialogue may be part of the context, together with the standard parameters, given that—according to Stalnaker—presuppositions are part of the extralinguistic context. But the contrary holds too, and Kaplan's context may be considered, from a logical point of view, as part of the common ground or context set. For a discussion of these two possibilities, see Stalnaker 2014: § 1.5.

Partial alternatives to Kaplan's view on the context of utterance are often developed under different interpretations of Fregean ideas. Among these, we find Wolfgang Künne (1992), who relied on some Fregean remarks on hybrid proper names to reject Kaplan's "intentionalist" viewpoint of demonstrations and suggested that demonstrations should be treated as parameters of the Kaplanian context (see also Predelli 2006; Kripke 2008; Künne 2010; Penco 2013; Textor 2015; Stojnic 2021). We should also mention here Manuel Garcia Carpintero (1998, 2015), who followed Reichenbach's token reflexive theory, critically discussed by Kaplan (1989a: 519), François Recanati (2012, 2013), who recovered the Fregean idea of "modes of presentations" under a new perspective, and Korta and Perry (2020), on which we will elaborate in the following section.

3. Contextual Dependence and the Pragmatics-Semantics Divide

Are truth conditions the only fundamental aspect of the meaning or sense of a sentence? After the original Frege-Wittgenstein idea that the sense of a sentence (a Fregean thought) is given by its truth conditions, there have been many attempts to find different and subtler notions of the content of an assertion. Sentences with the same truth conditions (same assertoric content) may have different constituents and therefore different meanings. Indexicals may help: If we say, on Monday, "Today it is raining in Stokholm" and, on Tuesday, "Yesterday it was raining in Stokholm", we may claim that the two sentences have the same truth conditions: they are true if and only if on Monday it was raining. However, Kripke (2008: 204-206) insisted, commenting on Burge's interpretation of Frege,

that the two sentences have expressions with different meanings, and therefore, given compositionality, they should express different thoughts. To address this concern, much subsequent literature distinguished a logical or truth-conditional sense and an epistemic or cognitive sense (see, e.g., Garavaso 1991; Beaney 1997; Künne 2007; Penco 2003, 2013; Perry 2021). This literature includes Dummett's distinction between assertoric content and ingredient sense,[9] as well as different two-dimensional semantics projects (on which see Davies and Stoljar 2004). Korta and Perry (2011) suggested distinguishing different levels of truth conditions: reflexive truth conditions concern the truth conditions when the indexicals are not saturated ("I am tired" is true iff the speaker of the utterance is tired) and subject matter truth conditions (I am tired said by Diana is true iff Diana is tired). In Kaplan's theory, utterances are modelled as pairs of expressions and contexts, and the context of utterance is just the set of parameters that permits the interpretation of a sentence in context as a proposition.[10] Instead, in the reflexive and referential approach by Korta and Perry (2011), utterances are explicitly what the theory is about. Although context is intended à la Kaplan as the situation or circumstance in which a sentence is uttered, the parameters of the context (speaker, time and location) are considered properties of utterances. The proposed advantage of this theory is that it can help represent different cognitive significances of the utterance, depending on the relation different people have to the utterance itself. For this analysis, see de Ponte, Korta and Perry 2020.[11]

3.1. What is Said and the Problems with Standard Semantic Analysis

Central to this discussion is the question of how to define the intuitive notion of "what is said". In the traditional Gricean view, what is said is the truth-conditional content that is the result of a process of linguistic decoding, supplemented by disambiguation, reference assignment, and saturation of the indexicals (Grice, 1989). However, this is not generally agreed upon. Even when accepting the intuitive idea that "what is said" maps the conception of truth-conditional content, many authors argue for a more robust idea of context dependence and claim that

[9] Dummett claimed that "It is raining here" has the same assertoric content of "It is raining where I am", but the two utterances have difference ingredient senses as they behave differently in modal contexts or contexts with a temporal adverb like "always". "It is always raining here" has a different assertoric content than "It is always raining where I am", because "here" is temporally rigid, while "where I am" is temporally variable. Dummett's distinction has been developed by Gareth Evans and Jason Stanley and criticised by Tsohatzidis (2015) and Stojnić (2017). Rabern (2017) shows that Dummett's distinction is coherent with Lewis' principle of coherence.

[10] Kaplan (1989b: 584) distinguishes between utterances and occurrences of expressions in a context: the former belongs to the theory of speech acts and the latter belong to semantics.

[11] Perry (1989a) distinguishes pre-semantic, semantic, and post semantic uses of context. An example will do. Take the sentence "I saw her duck under the table". Before deciding the interpretation of the indexical, I need to understand which context we are referring to, whether the person in question had lost an animal or was seeking security in a worrying situation (earthquake, war…). We will therefore decide whether "duck" stands for an animal or for a verb. This is a pre-semantic use of context that helps determine what is said. The semantic use (saturation of indexicals) begins when syntactic ambiguity is resolved, given the knowledge of the situation. The post-semantic use is dependent on previous tacit assumptions. For example, if I say "it is raining" I assume that I am referring to the location where I am, so we may think that location is an unarticulated constituent of the utterance.

disambiguation and saturation are not enough to get to the truth conditions and that we need context to provide further elements contributing to what is said. We enter here into the debate on the boundaries between semantics and pragmatics, a debate that has resulted in a flourishing body of literature.[12] The contrast is between letting or banning pragmatic aspects enter into truth conditions. The following linguistic intuitions are typically discussed as challenging the standard Gricean analysis:

(i) *Context Shifting*: Depending on the context of the conversation, some utterances appear to have different truth conditions. If some parents say about their child:

"Diana is tall".

What they say may be true if we are comparing Diana to her schoolmates, but if we shift the context and we are speaking about basketball players, the conditions under which the sentence is evaluated change, and the utterance is false with respect to the average height of basketball players.

(ii) *Incompleteness*: If someone utters:

"Serena is ready".

What are they saying? They might say that she is ready for the last tennis game, or she is ready to begin a new life, or she is ready to climb Mount Everest, or maybe they are just saying that she is ready to go out. We cannot understand what is said, unless we know what Serena is ready for, deriving the information from the context.

(iii) *Domain Restrictions*: If somebody says in a classroom:

"Every boy is seated".

What are they saying? They are probably not saying that every boy in the world is seated, but that boys in the classroom are seated.

(iv) *Deferred Meaning*: If a waiter says:

"The ham sandwich is annoyed".

What are they saying? Not that a piece of bread and ham is annoyed! In this case, we need to rely on context to understand that the waiter refers to the person who ordered the ham sandwich.

(v) *Relativity of Truth*: of a plant whose leaves have been accurately painted green, somebody says

"The leaves are green"

in front of an artist wanting to photograph something green and in front of a botanist interested in the original structure of the plant. The same utterance seems to be true and false at the same time.

3.2. Answers to the Challenges

These challenges to the standard Gricean account of what is said have been addressed in different ways. Among many trends, we may identify indexicalism,

[12] Discussions on the topic began with Turner 1999, von Heusinger and Turner 2003, Bianchi 2004, Szabo 2004, and gave rise to different thematic series addressing the problem from different viewpoints, like the *Current Research in the Semantics/Pragmatics Interface* (CRISPI)(Brill) or the *Palgrave Studies in Pragmatics, Language and Cognition* and other collections.

minimalism, relativism, and contextualism as the main competitors.[13] In what follows, we provide a brief characterisation of each of them.

Indexicalists: In his paper "Semantics in context" and other contributions, Jason Stanley develops an original idea for treating domain restriction. When I say that I'll bring all the beers to the party or that all the dogs were barking, I don't usually mean a universal generalisation over all beer or all dogs, but rather I intend to limit the domain of quantification to some specific set, possibly assumed in advance in the conversation (the beers in my fridge or the dogs in the neighbourhood). Stanley and Szabo (2000) and Stanley (2005a, b) suggest incorporating domain restriction into semantics, associating indices to every nominal in a sentence. Indices bind the referent of the nominal to a subset of the domain (e.g., to a subset of the beers, the ones in my fridge). The question belongs to semantics because it implies assigning semantic values to constituents of a sentence, relative to a context that has an input in the logical form of the sentence. Not everybody agrees with this idea, called "indexicalism", and some alternative solutions do not require postulating bound variables in the logical form.[14]

Minimalists such as Cappelen and Lepore (2008) or Borg (2007, 2012) identify the truth conditional content of utterances with their minimal content, following the classical disquotational analysis (for which "*p*" is true iff *p*). For instance, the ordered pair <Serena, ready> or <Diana, tall> represent the minimal content expressed by the utterances "Serena is ready" or "Diana is tall". What "ready" or "tall" mean is a question of metaphysics or of speech act theory.

Relativists (such as McFarlane 2014) distinguish between the context of utterance and the context of assessment, where the same sentence can be true or false depending on the latter. A sentence, therefore, is true or false in the context of utterance depending on the context of assessment. For instance, "Diana is tall" is true if "tall" counts as relative to her schoolmates and false if "tall" is relative to the standard height of basketball players, or "the leaves are green" is true if "green" counts as relative to their appearance and false if "green" counts as the original colour of the leaves.

Contextualists such as Kent Bach, Robyn Carston, François Recanati, and others, rely on the notion of "intuitive truth conditions" or "pragmatic truth conditions" and, in general, insist on the general underdetermination of the meaning of the lexicon, which strongly depends on contextual clues. To face these problems, there are different strategies, mostly linked to different kinds of enrichment of the sentence in context.

Essentially, there are two main contrasts in the debate on semantics-pragmatics boundaries: on the one hand, the distinction between "implicit" and "explicit" approaches to context (Neale 1990), and on the other hand the relation between formal and informal notions in semantics. For the former question, contextualists take context as an element from which to find information for enriching the proposition (or the propositional template, assuming that meaning is underdetermined), while others tend to connect more strictly the logical form of the sentence to contextual restrictions, as exemplified above in Stanley's view of

[13] For a general assessment of the different trends on context dependence, including relativism, see the entry of *Internet Encyclopedia of Philosophy* written by C. Penco and Massimiliano Vignolo (2020).

[14] Against the binding argument presented by Stanley, see for instance Neale 2005, in defence of Perry's idea of "unarticulated constituents". See also Zeman 2017.

indexicalism.[15] For the latter, Predelli (2005, 2013) considers clause-index pairs as the real objects of semantic evaluation, strictly distinguishing the semantic apparatus from reference to real-life objects or events. He, therefore, claims that part of the debate on context in semantics relies on misunderstanding the difference between informal and formal notions. Contextualists, on the other hand, insist on the idea of giving the idea of "intuitive truth conditions" a proper place in semantics.

3.3. An Example of a Contextualist Answer

Recanati (2005, 2010), who is one of the main proponents of contextualism, proposed the idea of "truth-conditional pragmatics", trying to explain the difference between "literal" truth conditions and "intuitive truth conditions". He distinguishes between saturation and modulation. *Saturation* is a mandatory pragmatic process guided by the meaning of the expressions. Indexicals are a clear case of expressions that require saturation, because—to understand the referent of "I"—the hearer is *guided* by the meaning of the indexical to select the speaker in the context of utterance, and analogously for other indexicals. *Modulation*, on the other hand, is a pragmatic process which is not triggered by the linguistically-encoded meaning of the expressions and concerns those cases in which the meaning of an expression is underdetermined. We need to directly rely upon the context in which a sentence is uttered, where "context" is intended both as objective and cognitive context, linked to world knowledge or shared beliefs. According to Recanati, the main modulation processes are free enrichment, loosening, and semantic transfer:

Free Enrichment may concern words like "tall" or "ready" (see (i) and (ii) above), and restrict their domain of application to some specific class of people or of actions (e.g., "tall" with respect to baseball players, or "ready" with respect to having the ability to participate in a tennis match);

Loosening, on the contrary, may broaden the domain of application, following Austin's example of "France is hexagonal" (apparently France's borders are not a perfect hexagon, but approximate it). In Recanati's example, "the ATM swallowed my credit card", the term "swallowed" applies to entities without a digestive system, but which may let something go inside the interior of their apparatus.

[15] Neale (1990) labels "implicit" and "explicit" the two main approaches to incomplete descriptions (and quantified NPs in general), which pose a difficult problem to any theory, starting from Russell's theory of descriptions. The explicit approach requires completing the matrix of the description to make the description to denote just one thing. This may be done with free enrichment or with more constrained means linked to some syntactic aspects (Carston 2002, Hall 2008, Neale 2004). The implicit approach assumes a contextual restriction of the domain of quantification. The implicit approach has been differently defended and developed by many philosophers and linguists: a first "simple" view would restrict the domain to a part of the world, for instance a situation (Barwise and Perry 1983). Westerståhl (1985: 49) remarks that we are not bound to use a unique context set as domain, given that different quantifiers in the same sentences may be interpreted relative to different domains (an assumption developed by Staney and Williamson 1995). Stanley and Szabò (2000), Stanley (2005) and others insist on the necessity of linking any possible completion with elements in the syntax, while Elbourne (2013) rejects the simple view of the implicit approach.

Semantic Transfer allows an expression to be used for a different meaning related to the expression, as in case (iv) above: "The ham sandwich is annoyed", where an inanimate object cannot have feelings, but in the context of a restaurant the meaning of "ham sandwich" is easily mapped onto the meaning of "customer who ordered the ham sandwich".

These and similar proposals, connected with an even more radical meaning underdetermination (Carston 2013, 2020, Unsteinsson 2015), have their critics, like Michael Devitt (2013, 2021) for whom there are conventional aspects in the meaning of the lexicon that should provide a better account. Lexicon, and the recognition of conventional rules of polysemy, should guide saturation processes, avoiding modulation processes. However, even stressing conventionality of meaning and relying on the idea of convention-guided saturation, the reference to context is unavoidable. If there is a slot to be mandatory filled in "tall", "ready", "ham sandwich", and so on, the context will decide what needs to fill the slot.

Although the notion of what is said remains a challenge for philosophers, the basic Gricean distinction between "what is said" and "what is implicated" is central to any theory of communication (however rich the notion of what is said is taken to be). According to Grice, rational speakers observe a Cooperative principle ("Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged") and some conversational maxims in order to provide true, informative, relevant and clear contributions to the conversation (Grice 1989). Crucially, according to Grice, when what is said appears to fall short of these conversational standards, hearers are then licenced to infer that what the speaker meant to communicate must be different or richer than what she literally said, and thus preserve this underlying assumption of cooperativeness. For instance, when someone replies "I'm tired" to an invitation for dinner, we can legitimately assume that what they intended to communicate was richer than what they said and infer that the speaker implicated that they are declining our invitation (and thus preserve the assumption that their utterance is relevant to the conversation at issue). Furthermore, Grice considers metaphors and irony as implicatures derived by the violation of the maxim of quality (telling the truth): when Shakespeare put in Romeo's mouth "Juliet is the sun", he made him violate the maxim of quality (everybody knows that Juliet is not the sun), thus licencing an array of implicatures (Juliet is like the sun, Romeo's life revolves around her, she is the source of joy in Romeo's life, etc.), that preserve the underlying assumption that the Romeo is still providing a cooperative contribution to the conversation.

A special case of implicatures, known as "scalar implicatures", has come to occupy a central place in the philosophy of language and linguistics. This kind of implicatures can be exemplified by the following utterance:

(9) Some students passed the exam.

In many contexts, an utterance like (9) can be taken to implicate that not all the students passed the exam, although the linguistically-encoded meaning of "some" (*at least some and possibly all*) is compatible with a situation in which all students passed the exam. From a Gricean perspective, this implicature—although generalized—is still context-dependent and can thus be cancelled without contradiction ("Some students passed the exam. Indeed, all of them did"). The debate around scalar implicatures concerns the conditions for their derivation, as well as the

extent to which they are linguistically—and not only pragmatically—mandated. Post-Gricean accounts, such as Relevancy Theory (Sperber & Wilson 1986/1995, Carston 1998), argue that this inference is licenced only under certain circumstances (e.g., in so-called "upper-bound" contexts, in which it is relevant to know whether *all the students passed the exams* is true, and at the condition that the speaker is judged to be sufficiently knowledgeable, see Breheny et al. 2013). In contrast with this, Neo-Gricean accounts maintain that scalar implicatures are derived routinely and independently of context, i.e. by default (Levinson 2000). Finally, scholars such as Chierchia (2013) have put forth "grammatical accounts", which see implicatures as arising through a silent exhaustification operator, akin to *only*, which acts on scalar alternatives. For a review of this debate, as well as the experimental evidence in line with or against these accounts, see Noveck 2018.

Finally, it is worth noting that Grice's heritage extends well beyond the issue of how to distinguish and explains different levels of meaning. For instance, more formally-oriented approaches to conversations, such as the *Questions Under Discussion (QUDs) framework* introduced by Roberts (1996/2012), have tried to provide a formalisation of the Gricean-inspired notion of "relevance to the conversation".[16] For instance, in the context of this framework, relevance is defined in terms of contextually established QUDs, that make up a stack of ordered questions that are prioritised in the common ground. According to this framework, an assertion is relevant if and only if it answers, at least partially, the current QUD.

4. Social and Cognitive Context

The notion of "cognitive context" may be interpreted in different ways, and we give here two main directions of research: (I) on the one hand cognitive context may be intended as the set of presuppositions given in the common ground, following Stalnaker's view. (II) On the other hand, cognitive context may be conceived as a psychological construct, in accordance with the research direction of Relevance theory. We will present some applications of the two directions. Both have relevance to what can be called the "social" context, that is, the context of conversational interactions where elements of social difference, prejudices and stereotypes may play a relevant role.

4.1. Common Ground, Social Context and Expressives

As discussed in section 1, Stalnaker's notion of context of utterance, the "context set", is defined as the set of presuppositions, i.e., propositions taken for granted in the common ground. In the philosophical arena, a new trend known as "Social Philosophy of Language" has emerged, discussing how to better define common ground (or context set) concerning the use of expressives, and more generally offensive language and hate speech (Khoo & Sterken 2021; McGowan & Maitra forth.). This trend has its roots in two established directions of research: on the one hand, the research on pornography as a speech act of subordination started by Rae Langton (1993), relying on older ideas of McKinnon (1987); on the other hand, the research on expressives (expressions like "Ouch" and "Oops" or "Jerk")

[16] See also Schoubye and Stokke 2016 and, for a criticism, Picazo 2022.

suggested by Kaplan (1999), developing some ideas of Frege. The discussion. on these themes has become so rich that we can only give a few hints.[17]

A short reminder of the first line of research (on pornography) makes us consider the relevant role that context plays in this discussion in prompting the revision of some aspects of the old theories. For instance, Saul reminds us that a speech act is not given by a sentence (or a video) but by an utterance in a context (or the viewing of a video in a context). She, therefore, claims that it is not pornographic videos per se that perform the act of subordinating women, but only some viewings of those videos in specific contexts. If the aim of McKinnon was to condemn pornography, she "should abandon the speech act approach" (Saul 2006: 247).

The second line of thought sparked a heated debate about the role of expressives or expressions arguably involving an expressive component, such as slurs, in altering the context set of a conversation. In the following, we offer some short remarks on this point. Slurs have a distinctive expressive power: they are defined as being derogative of an individual because of the individual's belonging to a social group with racial, sexual, or cultural connotations. Under negation, slurs behave differently from their neutral counterparts. Consider the following utterances:

(10) "John is a faggot",

(11) "John is homosexual".

If I deny (11) I just deny a fact,[18] while, if I deny (10), the derogatory meaning appears to escape negation. This suggests that slurs involve two contents:

(10a) John is homosexual

and

(10b) homosexuals are despicable as such.

There are different explanations of this kind of offensive language, starting from *semantic strategies*, for which the derogatory aspect belongs to the truth-conditional content of the utterance, and therefore (10) and (11) have two different truth conditions.[19] On the other hand, *pragmatic strategies* distinguish between the truth-conditional content and the derogatory aspect, claiming that (10) and (11) have the same truth conditions, but (10) conveys a derogatory aspect that (11) does not. Pragmatic strategies are of two kinds: on the one hand, starting from some notorious Fregean examples discussed by Kaplan (1999), and developed by Picardi (2006) and Williamson (2009), (10b) is conceived as a conventional implicature, something expressed but not asserted; on the other hand, starting from Stalnaker's idea of common ground, (10b) is considered to be a presupposition.

The role of context is much discussed under the idea that slurs are a kind of presupposition trigger. Stalnaker's theory says that a presupposition is appropriate if it is shared or accepted in the common ground. This might explain the facility

[17] In addition to the anthology of Khoo and Sterken (2021), we can consider Penco 2018 to be a development of some ideas of Eva Picardi (2006), also relying on Carpintero 2015, 2017.

[18] Apparently, tone of voice and social setting may be relevant: I may express the worst attitude using "neutral" words. However, in normal conversation, the difference between (10) and (11) is highly relevant, exactly for the reasons discussed in the text concerning the context set, that is the changes in common ground of interlocutors.

[19] For arguments supporting the semantic interpretation of pejoratives, see Hom 2012.

with which the slur is accepted in certain environments (where the presupposition is appropriate). However, as David Lewis (1979: 339) suggested, defining the process of *accommodation*, if you say something that requires a missing presupposition (e.g., using a slur requiring the presupposition of derogation towards a certain type of person), often "straightaway the presupposition comes into existence, making what you say acceptable after all". Therefore, presuppositions are a simple way to introduce into the conversational common ground a prejudice or a derogatory attitude, putting those who do not share the presupposition in an uneasy situation: if they don't say anything, it is as if they accept the presupposition. However, if they use the neutral term and not the slur, they may draw attention to what is actually the topic under discussion, partly cancelling or dismissing the presupposition. From a formal point of view, Bonomi (2006) introduces a new notion of *discommodation*, correlative to Lewis's accommodation, to explain what may happen in conversations with incompatible views. A participant in a conversation can open a "presuppositional slot" that allows a "local context shift". This lets the participant talk about a person without committing herself to the speaker's presuppositions.

The need for classification of the semantic and pragmatic approaches to slurs most likely conceals the complexity of the problem, which cannot be easily constrained within a precise definition and may benefit from different theories.[20] For instance, Cepollaro (2015, 2020) suggests a compromise between the conventional and presuppositional accounts, still relying on the conception of common ground.

As this brief overview shows, philosophers have become more attentive to the problem of social and ideological contrasts, where different presuppositions clash, and thus try to explain how the use of slurs may be a means to implicitly insert into the common ground derogatory attitudes that tend to be shared unless challenged.

4.2 Cognitive Context and the Relevance Theory Approach

A distinct approach to the notion of cognitive context was brought about by the cognitive turn in the field of pragmatics led by Post-Gricean frameworks such as Relevance Theory (Sperber & Wilson 1982; 1986/1995). This turn introduced a new characterization of the notion of context, conceived as a psychological construct:

> a subset of the hearer's assumptions about the world. It is these assumptions, of course, rather than the actual state of the world, that affect the interpretation of an utterance. A context in this sense is not limited to information about the immediate physical environment or the immediately preceding utterances: expectations about the future, scientific hypotheses or religious beliefs, anecdotal memories, general cultural assumptions, beliefs about the mental state of the speaker, may all play a role in interpretation" (Sperber & Wilson 1986/1995: 15-16).

[20] Differently from semantic and pragmatic approaches, Lepore and Stone (2018) claim that derogatory words are just prohibited words. For a general presentation of the topic of pejoratives, see the paper by Robin Jeshion (2021), who distinguishes canonical, descriptive, gendered slurs, and stereotyping expressions. On impositions of social roles in the use of slurs, see Popa-Wyatt 2018.

The scope of context is thus very broad, at least potentially. Crucially, though, psychological plausibility imposes some constraints: although the range of assumptions that can be brought to bear on the interpretation process is virtually unlimited, comprehension relies on a subset of these assumptions. How is the actual context selected among them?

According to Relevance Theory, the context for the comprehension is itself selected or chosen during the comprehension process so that it allows, together with the linguistic input, to achieve an overall relevant interpretation of a given utterance. Typically, only those contextual assumptions that, combined with the incoming information, satisfy the hearer's expectations of relevance become part of the context for comprehension.

This marks an important shift of perspective on the role of context in utterance comprehension: context is not given or uniquely determined but is the result of the very same (non-demonstrative) inferential process that underpins utterance comprehension: "Communication requires some degree of co-ordination between communicator and audience on the choice of a code and a context" (Sperber & Wilson 1986/1995: 43; see also Assimapokoulous 2017).

The idea that context is chosen during the process of interpretation has found interesting applications in different domains. For instance, Mazzarella and Domaneschi (2018) rely on this idea to explain the context sensitivity of presuppositions. Going beyond purely semantic approaches, they suggest that presuppositions are the output of an inferential process of pragmatic interpretation, in which their propositional content is constructed through a process of "mutual parallel adjustment" with the explicit content of the utterance, or what is said, and its implicatures. Consider, for instance, the following example (adapted from Mazzarella & Domaneschi 2018):

(12) Peter failed again

in which the semantics of the presupposition trigger "again" leads to the recovery of the proposition

(12b) Peter failed at (at least one) time $t < t'$.

Crucially, in different conversational settings, hearers can derive pragmatically enriched presuppositions, which specify not only the domain of Peter's failure but also the number of relevant times t at which Peter failed prior to t'. For instance, if Peter's parents were discussing whether Peter should change school, (12) would be relevant only if it provided good evidence in favour of the decision to move Peter to another school, and would thus be interpreted as presupposing that

(12c) Peter failed some school tests *many* times t_1-t_n before t'.

This is because, in such a conversational exchange, it is this proposition and not (12b) that allows the hearer to derive a relevant implicature. This example illustrates that the contextual assumptions exploited to establish what speakers intend to communicate are not given before the interpretation process starts. They are chosen or constructed on the fly, as part of the interpretation itself.

Given that the choice of context requires some degree of coordination between the speaker and the hearer, and that this involves some risk, some interesting questions arise: How can successful communication be achieved? Who is responsible for potential misunderstandings? Can the risky nature of communication be strategically exploited, especially in cases of manipulative or dishonest communication? We will focus on the latter question and briefly mention two

relevant directions of research, concerning respectively the notion of hearer manipulation and that of speaker accountability.

Maillat and Oswald (2009) and Maillat (2013) have proposed that the notion of (chosen) cognitive context can shed light on the phenomenon of manipulation. Specifically, they suggested that it is possible to define some forms of social manipulation as specific attempts to interfere with the process of context selection, by affecting the accessibility of potential contextual assumptions and preventing the hearer from accessing a dissonant, although optimal, set of these assumptions. For instance, in the case of the so-called "ad populum fallacy" (van Eemeren & Houtlosser 2008), which involves judging something acceptable because it is considered acceptable by many people, this interference can be exercised in two distinct ways. The manipulator can increase the degree of accessibility of the intended conclusion *p* through repetition, or she can use an expression such as "Everybody says *p*" to force the inferential process through a cognitive shortcut for strengthening-by-repetition, thus achieving an analogous effect.

As far as speaker accountability is concerned, Mazzarella (2021) has argued that the notion of cognitive context can address the long-standing issue of what makes speakers more or less likely to successfully deny the intention to communicate some risky content (for a discussion on "plausible deniability", see Pinker 2007). Much literature in the philosophy of language has suggested that, in virtue of their cancellability, conversational implicatures can be strategically denied, thus allowing the speaker to reduce their accountability for implicitly communicated messages (Fricker 2012). Crucially, if we interpret speaker denials as strategic moves to prompt a reconstruction of the cognitive context (in line with Camp 2018), we can better understand why some denials are judged as clearly implausible, while others may allow the speaker the benefit of the doubt. According to Mazzarella (2021), this depends on the cognitive utility of the re-constructed context put forth by the speaker: the greater the cognitive effects licenced by the reconstructed context and the lower the cognitive efforts involved, the higher the degree of plausibility of the denial.

5. Papers in this Issue

We present the papers featured in this special issue following the main sections presented above and provide a short overview of their respective contributions. They approach different topics in the philosophy of language by focusing on the relevance of context—in its many forms—in their analyses.

5.1. Papers on Context of Utterance and Indexicals

The first two papers develop in different ways Künne's notion of "hybrid proper names". A general problem about indexicals and demonstratives concerns their treatment in functional role semantics. Functional role semantics tend to be exclusively linked to linguistic features, therefore avoiding notions of "reference" or of "context of utterance". Adding indexicals and demonstratives in a functional role semantics may take two directions: one, à la Robert Brandom, treating indexicals as anaphoric initiators; the other inserting a theory of reference into the functional role semantics system. **Tadeusz Ciecierski and Paweł Grabarczyk's** paper analyses the difficulty of this second strategy and works out in detail their solution in the setting of a particular functional role semantics, an update of Kazimierz

Ajdukiewicz's directival theory of meaning, where the meaning of a linguistic expression is defined in terms of the role of the expressions in specific sets of sentences. In the theory under discussion, "directives" are kinds of sentences intended as defining lexical meaning, on the ground of "acceptance" from competent speakers. To insert indexicals in the theory Ciecierski and Grabarczyk treat indexicals as hybrid expressions, construed as ordered pairs composed of a linguistic expression and a referent. Expressions like "I", "you", "today", as such are not well-formed expressions and require to be considered as hybrid symbols. Therefore, this solution follows the strict requirement for which every different expression will refer to distinct objects: hybrid expressions involving the same linguistic component, but with different non-linguistic components, will refer to distinct objects.

Maciej Tarnowski relies on Kaplan's idea that demonstratives are direct reference devices like proper names. Kaplan (1989a: 506-507) said that the theory of indexicals "may prove useful" to face the problems associated with a theory of reference for proper names, and remarked the following difference: while the character (linguistic meaning) of an indexical is a function from contexts to extensions, proper names have no particular linguistic rule that fixes the content depending on context; therefore the meaning of proper names is identical with their content, a constant function from possible worlds to extensions. Against Kaplan's fundamental difference between proper names and indexicals, few authors tried to define proper names as kinds of indexicals, and therefore make them more context-dependent than previously thought. These attempts would have the advantage of creating a unifying theory of both indexicals and proper names. Tarnowski gives a summary of these attempts, mainly by Pelzcar, Rami and Recanati and shows their shortcomings. His paper tries then to give a solution based on treating proper names as simple demonstrative, in a hybrid view that should overcome the difficulties of the earlier treatments of proper names as indexicals.

5.2. Papers on the Pragmatics-Semantics Divide

Some papers in this issue touch upon the semantic/pragmatics boundary with different answers. **Ernesto Perini Santos** traces back the problem of the semantics/pragmatics divide to an old debate between Benson Mates and Stanley Cavell, He discusses a well-known example by Charles Travis, showing the different answers provided to it by minimalists and contextualists, and their related problems, as raised, for instance, by Claudia Picazo about the enrichment strategy. Perini Santos relies on Stefano Predelli's point of view, according to which a clause-index pair may be true at one point of evaluation and false at another, depending on the circumstance of evaluation (where the concept of "circumstance" is slightly different from Kaplan's). Eventually, this solution would transcend the dichotomy between minimalism and contextualism: going back to the old debate between Mates and Cavell may reveal new perspectives that were previously hidden due to the too-limited Gricean paradigm.

Giuseppe Varnier and Salvatore Pistoia-Reda discuss the claim that the language system contains a deductive inferential core (logicality of language hypothesis). Their paper concerns the acceptability or non-acceptability of certain structures, like, for instance, the sentence "it is raining, and it is not raining", which should be not acceptable in the logicality of the language hypothesis. However, the logicality hypothesis may be improved or rescued from a "pragmatic" or contextualist viewpoint, given that the sentence may be interpreted with a modulation or enrichment

of the nonlogical material, as with "it is raining, and it is not raining heavily". Therefore, we may accept analytical sentences as grammatical if their structures can be rescued in principle. This discussion is applied to belief ascriptions, which appear sensitive to logical consideration, against the standard view.

Aldair Díaz-Gómez discusses the topic of scalar implicatures, which has become a very central point in the debate on the pragmatics/semantics divide. He presents the alternative between a pragmatic and a grammatical account of scalar implicatures, where the former relies on the context of the dialogue (common ground), while the latter relies on a strictly lexically constrained context. The pragmatic account requiring access to contextual information should block inferences from SI where the presupposed proposition has an empty domain. On the other hand, a grammatical account would leave the SI work with an empty domain. He then provides an experiment on the processing of SIs, showing how the results are more coherent with the latter solution.

Ines Crespo, Andreas Heise, and Claudia Picazo touch upon a topic that has been much debated in the wake of Grice's theory of conversation, namely metaphorical interpretation. They present a view proposed by Asher and Lascarides who argue that discourse coherence holds the key to metaphorical interpretation. Crespo, Heise, and Picazo do not reject this basic idea, but they argue that discourse coherence is insufficient to deal with certain cases of metaphor and irony. In view of such considerations, they speak out in favour of a broader notion of contextual background that includes both world knowledge and perceptual information. On their account, elements from the extralinguistic context can play a role analogous to that of the linguistic context. Thus, they propose a notion of contextual coherence, eventually enriched via "Questions under Discussion", derived from both linguistic and extralinguistic elements of the context of utterance. Such a construal, they argue, may also help to make sense of interpretation in impoverished contexts, that is, where no previous context of discourse is available.

5.3. Papers on Social and Cognitive Context

Dealing with the much-debated problem of the distinction between lying and deceiving, **Palle Leth** invites us to reflect on the notion of "warrantability" of the hearer's contextual assumptions. Warrantability is meant to capture the extent to which hearers are justified in taking a certain assumption as part of the speaker's intended meaning. Crucially, the coordination between speakers and hearers on the choice of context takes place within the boundaries of what could have been plausibly intended. On the one hand, speakers cannot expect hearers to appeal to unforeseeable or highly unavailable assumptions. On the other hand, hearers are justified in relying on assumptions that are salient and not idiosyncratic. Despite its inevitably blurry boundaries, warrantability can shed new light on traditional distinctions such as the lying-misleading distinction. Indeed, Leth argues that speakers that communicate untruthful content to deceive the audience are deemed to be considered liars in so far as the inferential derivation of such untruthful content is warranted, independently of its degree of explicitness. Warrantability is thus tightened to speaker responsibility.

It is worth noting that, while the speaker-hearer coordination that underpins successful communication relies on the choice of both code and context, the weight of these two factors can vary. Speakers can try to be as explicit as possible in formulating what they intend to communicate, thus reducing the role of

contextual inferences. Or they can strategically play with implicit communication, as is the case in insinuations, and let contextual inferences carry controversial or risky content. The relative weight of code and context may be consciously manipulated by interlocutors and have important social—and even legal—consequences. As **Claudia Bianchi** reminds us, for instance, the public debate about speech acts in sexual contexts is dominated by discussions concerning what consent and refusal amount to. Crucially, much of this discussion focuses on the interplay between what is explicitly stated and what can be contextually inferred: "consent to intercourse cannot be inferred from contextual factors such as clothing, alcohol or drug consumption, flirting or engaging in some form of intimacy". Reducing the room for contextual inference appears the most important in what Bianchi calls "distorted communicative environments", in which dangerous gender stereotypes are present and reinforced. In this paper, Bianchi makes a case to anchor theoretical discussions on the linguistic practices that characterise (or should characterise) sexual negotiations to non-ideal, real-life communicative environments, "where the parties involved are not necessarily engaged in ethical undertakings, with more or less the same goals, and more or less the same price to pay for the activities performed". Bianchi focuses on theoretical models of sexual negotiation, specifically on the Collaborative Models advocated by Kukla (2018), Gardner (2018) and Caponetto (2021), which conceive initiations of sex in terms of invitations, offers and proposals. She argues that real-life cases show the limits of the Collaborative Models, which by *presupposing* women's sexual agency can run the risk of masking prevarication and abuse.[21]

References

Assimapokoulous, S. 2017, "Context in Relevance Theory", in Blochowiak, J., Grisot, C., Durrleman, S. and Laenzlinger, C. (eds.), *Formal Models in the Study of Language: Applications in Interdisciplinary Contexts*, Cham: Springer, 221-42.

Barwise, J. and Perry, J. 1983, *Situations and Attitudes*, Cambridge, MA: MIT Press.

Bianchi, C. (ed.) 2004, *The Semantics/Pragmatics Distinction*, Stanford: CSLI.

Bonomi, A. 2006, "Truth and Reference in Context", *Journal of Semantics*, 23, 2, 107-34.

Borg, E. 2007, "Minimalism vs. Contextualism in Semantics", in Preyer and Peter 2007, 546-71.

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

Breheny, R., Ferguson, H.J. and Katsos, N. 2013, "Taking the Epistemic Step: Toward a Model of On-Line Access to Conversational Implicatures", *Cognition*, 126, 3, 423-40.

Brézillon, P. and Turner, R. 2021, "Pragmatic Research on Context Modeling and Use", *Modeling and Using Context (MUC Review)*, 4, 1-6.

Bühler, K. 1934, *Sprachtheorie*, Jena: Fischer; Engl. tr. *Theory of Language*, Amsterdam: John Benjamins, 2011.

Camp, E. 2018, "Insinuation, Common Ground", in Fogal, D., Harris, D.W. and Moss, M. (eds.), *New Work on Speech Acts*, New York: Oxford University Press, 40-65.

[21] We thank Diego Marconi and Massimiliano Vignolo for suggestions on an earlier version of this introduction.

Cappelen, H. and Lepore, E. 2008, *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, London: John Wiley & Sons.

Carnap, R. 1956, *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago: University of Chicago Press.

Carston, R. 1998, "Informativeness, Relevance and Scalar Implicature", in Carston, R. and Uchida, S. (eds.), *Relevance Theory: Applications and Implications*, Amsterdam: John Benjamins, 179-236.

Carston, R. 2002, *Thoughts and Utterances*, Oxford: Blackwell.

Carston, R. 2013, "Word Meaning, What is Said and Explicature", in Domaneschi & Penco 2013, 175-204.

Carston, R. 2019, "Ad Hoc Concepts, Polysemy and the Lexicon", in Scott, K., Carston, R. and Clark, B. (eds.), *Relevance, Pragmatics and Interpretation*, Cambridge: Cambridge University Press, 150-62.

Carston, R. 2020, "Polysemy, Pragmatics and Sense Conventions", *Mind and Language*, 36, 1, 108-33.

Carrara, M. and Sacchi, E. 2006, "Propositions: An Introduction", *Grazer Philosophische Studien*, 72, 1, 1-28.

Caponetto, L. 2021, "Modeling Sex as a Joint Activity", *Open for Debate*, https://blogs.cardiff.ac.uk/openfordebate/modeling-sex-as-a-joint-activity/

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

Ciecierski, T. and Grabarczyk, P. (eds.) 2020, *The Architecture of Context and Context-Sensitivity*, Cham: Springer.

Davies, M. and Stoljar, D. (eds) 2004, "The Two-Dimensional Framework and Its Applications: Metaphysics, Language, and Mind", Special Issue of *Philosophical Studies*, 118, 1, 2.

De Ponte, M., Korta, K. and Perry, J. 2020, "Utterance and Context", in Ciecierski and Grabarczyk 2020, 15-28.

Devitt, M. 2013, "Three Methodological Flaws of Linguistic Pragmatism", in Domaneschi and Penco 2013, 285-300.

Devitt, M. 2021, *Overlooking Conventions. The Trouble with Linguistic Pragmatism*, Cham: Springer.

Domaneschi, F. and Penco, C. (eds.) 2013*, What is Said and What is Not (on Semantics/Pragmatics Interface)*, Stanford: CSLI.

Dummett, M. 1973, *Frege: Philosophy of Language*, London: Duckworth.

Dummett, M. 1981, "The Context Principle", in *The Interpretation of Frege's Philosophy*, London: Duckworth, 360-427.

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

van Eemeren, F. and Houtlosser, P. 2008, "Rhetoric in a Dialectical Framework: Fallacies as Derailments of Strategic Manoeuvring", in Weigand, E. (ed.), *Dialogue and Rhetoric*, Amsterdam: John Benjamins, 133-52.

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

Fricker, E. 2012, "Stating and Insinuating", *Proceedings of the Aristotelian Society*, Supplementary Volume LXXXVI, 61-94.

Frixione, M. 1994, *Logica, significato e intelligenza artificiale*, Milano: Angeli.

Garavaso, P. 1991, "Frege and the Analysis of Thought", *History and Philosophy of Logic*, 12, 195-210.

Garcia-Carpintero, M. 2015, "Contexts as Shared Commitments", *Frontiers of Psychology*, 6:1932, DOI: 10.3389/fpsyg.2015.01932

Garcia-Carpintero M. 2017, "Pejoratives, Contexts and Presuppositions", *Lecture Notes in Artificial Intelligence*, 10257, 15-24.

Garcia-Carpintero, M. and Kölbel, M. (eds.) 2008, *Relative Truth*, Oxford: Oxford University Press.

Gardner, J. 2018, "The Opposite of Rape", *Oxford Journal of Legal Studies*, 38, 1, 48-70.

Georgi, G. 2020, "Demonstratives in First-Order Logic", in Ciecierski and Grabarczyk 2020, 125-48.

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

Hall, A. 2008, "Free Enrichment or Hidden Indexicals?", *Mind & Language*, 23, 426-56.

Heim, I. 1992, "Presupposition Projection and the Semantics of Attitude Verbs", *Journal of Semantics*, 9, 3, 183-221.

Hom, C. 2012, "The Puzzle of Pejoratives", *Philosophical Studies*, 159, 383-405.

Kamp, H. 1971, "Formal Properties of 'now'", *Theoria*, 37/3, 227-73; repr. in Kamp 2013, 11-51.

Kamp, H. 2013, *Meaning and the Dynamics of Interpretation: Selected Papers of Hans Kamp*, ed. by K. von Heusinger and A. ter Meulen, Leiden: Brill.

Kamp, H. and Reyle, U. 1993, *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Collection: Studies in Linguistics and Philosophy, 42, Dordrecht: Kluwer.

Kaplan, D., 1979, "On the Logic of Demonstratives", *Journal of Philosophical Logic*, 8, 1, 81-98.

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

Kaplan, D. 1989b, "Afterthoughts", in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, 565-614.

Kaplan, D. 1999, "The Meaning of Ouch and Oops: Exploration in the Theory of Meaning as Use", Manuscript.

Khoo, J. and Sterken, R.K. (eds.) 2021, *The Routledge Handbook of Social and Political Philosophy of Language*, London: Routledge.

Korta, K. and Perry J. 2011, *Critical Pragmatics*, Cambridge: Cambridge University Press.

Kripke, S.A. 1979, "A Puzzle about Belief", in Margalit, A. (ed.) *Meaning and Use*, Dordrecht: Springer, 239-83.

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

Kripke, S. 2011, "Presupposition and Anaphora: Remarks on the Formulation of the Projection Problem", in *Philosophical Troubles*, Oxford: Oxford University Press, 351-72.

Kukla, R. 2018, "That's What She Said: The Language of Sexual Negotiation", *Ethics,* 129, 1, 70-97.

Künne, W. 1992, "Hybrid Proper Names", *Mind*, 101, 404, 721-31.

Künne, W. 2007, "A Dilemma in Frege's Philosophy of Thought and Language", *Rivista di Estetica*, 34, 1, 95-120.

Künne, W. 2010, "Sense, Reference and Hybridity: Reflections on Kripke's Recent Reading of Frege", *Dialectic*a, 64, 529-51.

Jeshion, R. 2021, "Varieties of Pejoratives", in Khoo and Sterken 2021, Chapter 13.

Langton R. 1993, "Speech Acts and Unspeakable Acts", *Philosophy & Public Affairs*, 22, 4, 293-330.

Lepore, E. and Stone, M. 2018, "Slurs and Tone", in Coliva, A. *et al.* (eds.), *Eva Picardi on Language, Analysis and History*, London: Palgrave, 205-18.

Lewis, D. 1979, "Scorekeeping in a Language Game", in Bäuerle, R., Egli, U. and Stechow, A. (eds.), *Semantics from Different Points of View*, Berlin: Springer, 127-87.

Lewis, D. 1980, "Index, Context, and Content", in Kanger, S. and Öhman, S (eds.), *Philosophy and Grammar*, Dordrecht: Reidel; repr. in David Lewis, *Papers in Philosophical Logic*, Cambridge: Cambridge University Press, 21-44.

Maillat, D. 2013, "Constraining Context Selection: On the Pragmatic Inevitability of Manipulation", *Journal of Pragmatics*, 59, 190-99.

Maillat, D. and Oswald, S. 2009, "Defining Manipulative Discourse: The Pragmatics of Cognitive Illusions", *International Review of Pragmatics*, 1, 2, 348-70.

Mazzarella, D. 2021, "I Didn't Mean to Suggest Anything Like That!: Deniability and Context Reconstruction", *Mind & Language*, DOI: 10.1111/mila.12377

Mazzarella, D. and Domaneschi, F. 2018, "Presuppositional Effects and Ostensive-Inferential Communication", *Journal of Pragmatics*, 138, 6, 17-29.

Montague, R. 1974, *Formal Philosophy: Selected Papers of Richard Montague*, edited and with an Introduction by R.H. Thomason, Yale: Yale University Press.

Neale, S. 1990, *Descriptions*, Cambridge, MA: MIT Press.

Neale, S. 2000, "On Being Explicit: Comments on Stanley and Szabo, and on Bach", *Mind & Language*, 15, 2/3, 284-94.

Neale, S. 2004, "This, That, and the Other", in Reimer, M. and Bezuidenhout, A. (eds.), *Descriptions and Beyond*, Oxford: Oxford University Press, 68-182.

Penco, C. 2003, "Frege, Sense and Limited Rationality", *Modern Logic*, 9, 53-65.

Penco, C. 2013, "Sense and Linguistic Meaning: A Solution to the Burge-Kripke Conflict", *Paradigmi* III, 75-89.

Penco, C. 2015, "Frege's Theory of Demonstratives as a Model for Indexical Thought", in Shott, D. (ed.), *Frege: Freund(e) und Feind(e)*, Berlin: Logos, 201-16.

Penco, C. 2018, "Refusing to Endorse: A Must Explanation for Pejoratives", in Coliva, A. *et al.* (eds.), *Eva Picardi on Language, Analysis and History*, London: Palgrave, 219-40.

Penco, C., 2021, "Indexicals and Essential Demonstrations", *Semiotica*, 240, 261-84.

Penco, C. and Vignolo, M. 2020, "Meaning and Context Sensitivity", *Internet Encyclopedia of Philosophy*, https://iep.utm.edu/mean-c-s/

Perry, J. 1979, "The Problem of the Essential Indexical", *Nous*, 13, 1, 3-21; repr. in *The Problem of the Essential Indexical and Other Essays*, Stanford: CSLI, 2nd ed., 2000, Chapter 2.

Perry, J. 1998a, "Indexicals, Contexts and Unarticulated Constituents", in *Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference*, Stanford: CSLI, 1-16.

Perry, J. 1998b, "Indexicals and Demonstratives", in Hale, B. and Wright, C (eds.), *A Companion to Philosophy of Language*, Oxford: Blackwell, 586-612.

Picardi, E. 2006, "Colouring, Multiple Propositions, and Assertoric Content", *Grazer Philosophische Studien*, 72, 1, 49-71.

Picazo, C. 2022, "Are Utterance Truth-conditions Systematically Determined*?", Inquiry*, 65, 8, 1020-41.

Pinker, S. 2007, "The Evolutionary Social Psychology of Off-Record Indirect Speech Acts", *Intercultural Pragmatics*, 4, 4, 437-61.

Ponte, M.D., Korta, K., and Perry, J. 2020, "Utterance and Context", in Ciecierski and Grabarczyk 2020, 15-28.

Popa-Wyatt, M. and Wyatt, J.L. 2018, "Slurs, Roles and Power", *Philosophical Studies*, 175, 2879-2906.

Predelli, S. 1998, "I Am Not Here Now", *Analysis*, 58, 2, 107-15.

Predelli, S. 2005, *Contexts: Meaning, Truth, and the Use of Language*, Oxford: Oxford University Press.

Predelli, S. 2013, *Meaning without Truth*, Oxford: Oxford University Press.

Preyer, G. and Peter, G. (eds.) 2005, *Contextualism in Philosophy: Knowledge, Meaning, and Truth*, Oxford: Clarendon Press.

Preyer, G. and Peter, G. (eds.) 2007, *Context-Sensitivity and Semantic Minimalism*, Oxford: Oxford University Press.

Pynn, G. 2016, *Contextualism in Epistemology*, Oxford: Oxford University Press.

Rabern, B. 2017, "A Bridge from Semantic Value to Content", *Philosophical Topics*, 45, 2, 181-207.

Recanati, F. 2005, "Literalism and Contextualism: Some Varieties", in Preyer and Peter 2005, 171-96.

Recanati, F. 2010, *Truth-Conditional Pragmatics*, Oxford: Clarendon Press.

Recanati, F. 2013, "Reference through Mental Files", in Domaneschi and Penco (eds.) 2013, 159-74.

Recanati, F., Stojanovic, I. and Villanueva, N. (eds.) 2010, *Context Dependence, Perspective, and Relativity*, Berlin: De Gruyter.

Reichenbach, H. 1947, *Elements of Symbolic Logic*, New York: Macmillan.

Roberts, C. 1996/2012, "Information Structure: Towards an Integrated Formal Theory of Pragmatics", *Semantics and Pragmatics*, 5, 6, 1-69.

Saul, J.M. 2006, "Pornography, Speech Acts and Context", *Proceedings of the Aristotelian Society*, 106, 229-48.

Saul, J.M. 2012, *Lying, Misleading, and What Is Said: An Exploration in Philosophy of Language and in Ethics*, Oxford: Oxford University Press.

Schlenker, P. (2010), "Local Contexts and Local Meanings", *Philosophical Studies*, 151, 1, 115-42.

Schoubye, A.J. and Stokke, A. (2016), "What is Said?", *Nous*, 50, 4, 759-93.

Singer, G. 2022, "Advancing Machine Intelligence: Why Context Is Everything", in *Towards Data Science*, May 10, https://towardsdatascience.com/advancing-machine-intelligence-why-context-is-everything-4bde90fb2d79

Sauerland, U. and Yatsushiro, K. 2009, *Semantics and Pragmatics: From Experiment to Theory*, Basingstoke: Palgrave-MacMillan.

Sperber, D. and Wilson, D. 1982, "Mutual Knowledge and Relevance in Theories of Comprehension", in Smith, N. (ed.), *Mutual Knowledge*, London: Academic Press, 61-87.

Sperber, D. and Wilson, D. 1986/1995, *Relevance: Communication and Cognition*, Oxford: Blackwell.

Stalnaker, R. 2002, "Common Ground", *Linguistics and Philosophy*, 25, 701-21.

Stalnaker, R. 2014, *Context*, Oxford: Oxford University Press.

Stanley, J. 2005a, "Semantics in Context", in Preyer and Peter 2005, 231-53.

Stanley, J. 2005b, *Language in Context*, Oxford: Oxford University Press.

Stanley, J. and Szabò, Z. 2000, "On Quantifier Domain Restriction", *Mind & Language*, 15, 219-61.

Stanley, J. and Williamson, T. 1995, "Quantifiers and Context-Dependence", *Analysis*, 55, 291-95.

Stojnić, U. 2017, "On the Connection between Semantic Content and the Objects of Assertion", *Philosophical Topics*, 452, 163-80.

Stojnić, U. 2021, *Context and Coherence: The Logic and Grammar of Prominence*, Oxford: Oxford University Press.

Szabo, Z.G. (ed.). 2004, *Semantics versus Pragmatics*, Oxford: Oxford University Press.

Travis, C. 2008, *Occasion-Sensitivity: Selected Essays*, Oxford: Oxford University Press.

Textor, M. 2015, "Frege's Theory of Hybrid Proper Names Extended", *Mind*, 124, 823-47.

Tsohatzidis, S.L. 2015, "The Distance Between 'Here' and 'Where I Am'", *Journal of Philosophical Research*, https://doi.org/10.5840/jpr2015101929

Turner, K. (ed.) 1999, *The Semantics/Pragmatics Interface from Different Points of View*, Oxford: Elsevier Science.

Unnsteinsson, E.G. 2014, "Compositionality and Sandbag Semantics", *Synthese*, 191, 3329-50.

von Heusinger, K., Peregrin, J., and Turner, K. (eds.) 2003, *Where Semantics Meets Pragmatics*, Conference Proceedings, http://gerlin.phil-fak.uni-koeln.de/kvh/konf/200303LSA/SemPrag03.Progr.pdf

Wang, Y. 2018, "Beyond Knowing That: A New Generation of Epistemic Logics", in van Ditmarsch, H. and Sandu, G. (eds.), *Jaakko Hintikka on Knowledge and Game-Theoretical Semantics*, Cham: Springer.

Westerståhl, D. 1985, "Determiners and Context Sets", in van Benthem, J. and A. ter Meulen (eds.), *Generalized Quantifiers in Natural Language*, Dordrecht: Foris, 45-71.

Williamson, T. 2003, "Everything", *Philosophical Perspectives*, 17, 415-65.

Williamson, T. 2006, "Reference, Inference and the Semantics of Pejoratives", in Almog, J. and Leonardi, P. (eds.), *The Philosophy of David Kaplan*, Oxford: Oxford University Press, 137-58.

Zeman, D. 2017, "The Use of the Binding Argument in the Debate about Location", in Conrad, S-J. and Petrus, K. (eds.), *Meaning, Context, and Methodology*, Boston: De Gruyter Mouton, 191-212.

Directives and Context

Tadeusz Ciecierski and Paweł Grabarczyk***

**University of Warsaw*

***IT University of Copenhagen and University of Lodz*

Abstract

The paper aims to add contextual dependence to the new directival theory of meaning, a functional role semantics based on Kazimierz Ajdukiewicz's directival theory of meaning. We show that the original formulation of the theory does not have a straight answer on how the meaning of indexicals and demonstratives is established. We illustrate it in the example of some problematic axiomatic and inferential directives containing indexicals. We show that the main reason why developing the new directival theory of meaning in this direction is difficult is that the theory focuses on the notion of a sentence (and not the notion of an utterance). To add the latter notion to the theory, we introduce the idea of admissible contextual distribution being an interpretation of the hybrid expression view on indexicals and demonstratives. We argue that this idea introduces a small but important modification to the concept of language matrix and gives way to define two distinct concepts of meaning: for an expression type and for a use of an expression type.

Keywords: Directival theory of meaning, Context, Indexicals, Demonstratives, Hybrid expressions.

1. New Directival Theory of Meaning

The new directival theory of meaning (henceforth the nDTM) is a functional role semantics based on a theory of meaning originally proposed by Kazimierz Ajdukiewicz in the 1930s.[1] The easiest way of explaining the theory is to start with a slogan that it defines the meaning of words based on a combination of syntax and pragmatics. Let us now see how both aspects contribute to the theory. The nDTM begins with an observation about a specific type of disputes language users sometimes engage in. Since this part of the theory deals with users and their

[1] Cf. Ajdukiewicz 1934, 1978. A detailed description of all differences between the original theory and the nDTM is beyond the scope of this paper. A reader interested in learning about these differences as well as learning about the details of the new take on Ajdukiewicz's theory can find both in Grabarczyk 2017, 2019. The summary provided below focuses on those aspects of the theory that are necessary for the understanding of the problem described in section 2 as well as the consequences to the theory that follow from the solution proposed in section 4.

Argumenta (2022): 35-53
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.cie

behavior, it can be classified as pragmatics. The nDTM points out that it is not uncommon for both sides of an argument to suspect that their dispute may be purely verbal—that it is only a result of using words with two different meanings. What the users do in these cases is shift their focus from the original dispute towards something the nDTM calls *semantic trials*. The trials confront the user with sentences that contain the disputed expression. The nDTM assumes that a competent language user knows that what they are expected to do during the trial is accept the invoked sentence, provided the circumstances are right. The titular 'directives' are rules of language that correlate specific sentences with correct circumstances and instruct the user to accept the sentence whenever it is uttered in these circumstances. The unusual nature of these sentences is that they are typically never used outside of the context of a semantic trial; they would have been seen as too trivial in a normal conversation. To make it clearer, imagine the following situation. Imagine two people who have a heated discussion about tables. At some point, one of them feels that the opinion of the other side is so outlandish, that they may be using the word "table" differently. To be sure, the person asks the following question: "Tables are a piece of furniture, right?". This is how a semantic trial can look like. It is aimed at checking if the interlocutors share the same directives. Directives function as cornerstones or boundaries of language because every competent user knows that once they fail to accept the sentence indicated by a directive, they will be treated by other language users as linguistically incompetent. It is important to point out that the directives function as a tacit linguistic mechanism; they do not have to be (and typically are not) known to the users. The users are trained to behave in a certain way and can formulate directives only as a *post-factum* observation of patterns of their behavior. In this respect, the directives function like the rules of syntax. The language user can gain full linguistic competence and not be able to formulate the rules of syntax.

Depending on the type of circumstances indicated by the directives, the nDTM differentiates between four types of directives: axiomatic, inferential, empirical, and promotive. The first category is the easiest to explain, as it is the one that does not specify any particular circumstances; the only thing the *axiomatic* directives ensure is that the user accepts a particular sentence, regardless of the circumstances they are in.[2] The table example we have just used falls under this category because the sentence "table is a piece of furniture" is expected to be accepted on all occasions. Building on this simple structure, it should now be easy to picture *inferential* directives as rules that instruct the user to accept a given sentence if they accepted some other, specified sentence—imagine the *modus ponens* rule as a model example of this. What it forces the user to do is that they must accept a given sentence if they already accepted an implication that has this sentence as a consequence and if they accepted the antecedent of this implication. In this case, the whole notion of circumstances correlated with the acceptance of a sentence boils down to the prior acceptance of other sentences.

Things get more interesting (and, admittedly, more complicated) once we proceed to the third and the fourth type of directive, which add extra-linguistic circumstances to the mix. *Empirical* directives instruct the language users to accept

[2] As we are going to see, the problems associated with contextual dependence challenge this simplified picture, as some unusual contexts combined with context-sensitive expressions may affect the reaction of the user.

a particular sentence if they happen to be in a certain perceptual state.[3] To see it in an example, imagine the sentence "This is red" accepted by users while they look at a vivid red surface. The last type of directive (one that we might call *promotive*) states that in situations where the users accept a certain sentence, they should perform a certain bodily action: enter a certain motor state. To see this type of directive in an example, imagine the sentence "Stop!" accepted by the user. Once they do that, it is required of them to accompany their acceptance with a particular reaction; to put it bluntly, they must stop moving.

Following the original theory, the nDTM assumes that for every simple expression in language, there is a set of directives that contains this expression. Roughly speaking, the main idea of the nDTM is that the meaning of a given simple expression is its placement within 'its' directives—the directives it appears in. For a more detailed picture, imagine the set of all directives of a given language encoded in some way that retains their structure, that indicates the circumstances and the sentence that is to be accepted. Once we have it, we end up with something that the nDTM calls a *language matrix*. This construct enables us to define the meaning of an expression as its distribution in the language matrix. Even though the language matrix reflects the behavior of the users, the distribution of terms is given in a purely formal manner; it refers only to the structure of sentences and the placement of the expression within them. Because of this, as mentioned at the beginning, the theory can be said to be a combination of syntax and pragmatics. What must be specifically noted is that the nDTM does not make any use of the notion of reference; it is a purely non-referential theory of meaning. Needless to say, this does not mean that the nDTM claims that language is non-referential. The only thing the nDTM claims is that it is possible to explain the notion of linguistic meaning without invoking the notion of reference. The nDTM is best seen as a "prohibitive semantics"—a theory that construes linguistic meaning as something that regulates only the boundaries of language, something that helps us avoid misuse of language (as opposed to determining its use). The point of such a minimalist approach to semantics is that it removes a lot of pressure from the notion of linguistic meaning. We should not expect the notion of meaning to explain all the mechanisms that are present in the language. As argued in Grabarczyk (2019) since prohibitive semantics can give us notions such as "synonymy" or "translation", there is no need to postulate a richer theory of meaning. Still, if we wanted to explain linguistic phenomena that go beyond the minimalistic meaning defined in the nDTM, the theory does not prevent us from doing so. For example, there is nothing that prevents us from adding theory of reference as an auxiliary theory because the fact the nDTM does not make use of this notion does not mean it cannot be useful for explaining other linguistic facts. The solution we present in this article should be treated as one such addition to the original theory—an extension that helps it cover cases it was not originally designed to cover.

Since the language matrix is a theoretical construct of the linguist—a description of patterns of acceptance that can be observed in the linguistic community—it can be represented in a variety of ways. One simple way of representing it is a table that contains two main sections: the input and the output parts. Both parts can contain linguistic and extra-linguistic parts, and the four types of meaning directives we described above can now be distinguished depending on which

[3] The way this state is understood is rather complex, but we do not have the space to discuss it in more detail. See Grabarczyk 2019 for a thorough discussion of its nature.

combination of input and output they contain. Axiomatic directives are the ones that result in linguistic output, regardless of the type of input. Inferential directives tie linguistic output with linguistic input; empirical directives present linguistic output when a specified extra-linguistic input is present; and, finally, promotive directives combine linguistic input and non-linguistic output.

To use a very crude simplification of the relation between a language matrix and the meaning of expressions, imagine that you observe a community of language users and discover sentences they always accept in certain circumstances. Let us now say that you fill in a spreadsheet table that indicates circumstances and the sentence to be accepted. The meaning of a given expression could then be given via a complete list of all the cells it appears in. Let us illustrate it on a very simple example that uses only two types of directives: axiomatic and inferential. Consider the following toy language that contains the following directives (assume that the symbol + signifies that a given sentence has been accepted by the users and the symbol ++ signifies that it should be accepted).

Axiomatic directives:

1.1 ++P(a).

1.2 ++P(a) & Q(b).

Inferential directives:

2.1 +P(a); ++ Q(b).

2.2 +P(a) & Q(b); ++ Q(c).

2.3 +Q(b); ++ (Qd).

According to the nDTM, the language matrix for this toy language is any systematic way that enables us to dismantle the language syntax and divide the directives into circumstances and a sentence that needs to be accepted. The method used by Ajdukiewicz, that I copy below is that for every expression we start with the expression followed by its main connective, its first argument, its first connective, etc. We decompose expressions as long as we get to the level of atomic expressions. An example of the result of this procedure for the above toy language could look like this:

	Circumstances							Response							
1.1								P(a)	P	a					
1.2								P(a)&Q(b)	&	P(a)	P	a	Q(b)	Q	b
2.1															
2.2	P(a)&Q(b)	&	P(a)	P	a	Q(b)	Q	Q(c)	Q	c					
2.3	Q(b)	Q	b					Q(d)	Q	d					

Now, if we wanted to define the meaning of a given simple expression, for example, Q, it would be identified with the distribution of this expression in the matrix. To make it easier to grasp we can now represent this distribution graphically. According to the nDTM, any expression that happens to have the same distribution as Q is synonymous with Q and any expression that has an identical distribution in some other language is a translation of Q.

	Circumstances								Response							
1.1									P(a)	P	a					
1.2									P(a)&Q(b)	&	P(a)	P	a	Q(b)	Q	b
2.1																
2.2	P(a)&Q(b)	&	P(a)	P	a	Q(b)	Q	b	Q(c)	Q	c					
2.3	Q(b)	Q	b						Q(d)	Q	d					

A definition of this type may disappoint people who expect the theory of meaning to provide something akin to a dictionary entry, but, as argued by Grabarczyk (2019), it can still deliver many other features of meaning, such as an explanation of synonymy, translation, or the difference between meaningful and nonsensical expressions.

Two additional explanations are in order. First, the language users are not expected to know the directives. As we already said, the directives are a theoretical construct of the linguist. What exists in the community is a certain regularity of reactions: the users recognize that they are challenged to the semantic trial and that if they fail to accept the sentence enclosed in the directive, they will not be treated as competent users of a given expression. Second, the axiomatic and inferential directives differ from the two other types of directives in that they contain schemas of sentences (and not particular substitutions of schemas). Once again, this will be most easily explained in the case of axiomatic directives. Imagine a simple case of identity. It is plausible to assume many languages to expect their users to accept sentences of the form A=A.[4] What it means in practice is that the users know that any sentence of this form where the variable A has been replaced with an expression of a correct syntactical category should be accepted.

What is crucial for our purposes is that neither the nDTM nor the original Ajdukiewiczian theory engages with the notion of *context*, even though it touches upon it indirectly at least twice. Firstly, when the notion of the semantic trial is introduced. Recognizing semantic trials may be difficult, as they can be signaled in different ways in different linguistic communities. The most conspicuous way of signaling them is that the disputants engage in the semantic discourse and start to talk about 'meanings' or 'senses' of the word. These expressions function as markers of a contextual switch from a normal conversation to a semantic trial. One notorious sign of a semantic trial is that the users who fail to pass it—those who reject the sentence indicated by a directive are never treated seriously—will be seen as misunderstanding the sentence they reject or simply as using a different vocabulary. Semantic trials can thus be seen as a very specific context that competent language users must recognize.

The second aspect of the nDTM that reminds us of the notion of *context* is the idea of the circumstances C that the user must recognize as correct ones for a given directive. To provide an example of this, imagine a perception of a typical red patch that accompanies an empirical directive "this is red". Both aspects can be described as contextual, as they refer to extra-linguistic aspects of the utterance that must

[4] The reason we are so cautious here is that whether a given directive functions in a given language demands the linguist to do the actual field work and test the users. It cannot be decided *a priori*.

accompany it. It is nonetheless important to remember that, apart from these two aspects, the nDTM does not take *contextual* dependence into account. Most importantly, it does not consider typical contextual parameters that are required to explain the meaning of expressions that are contextually dependent, such as indexicals. The main reason why this omission must be analyzed is that the lack of discussion of contextual dependence in the nDTM prevents us from asking if it is possible for some contextual parameters to influence semantic tests or directives, such as if it is possible for some of them to cancel the markers of a semantic test.

2. Indexicals and Demonstratives in the nDTM: The Problem

The ambition of the nDTM is to provide a theory of linguistic meaning. The realization of this task, however, faces a serious theoretical challenge when it comes to the analysis of indexicals and demonstratives (as well as context-sensitive expressions in general). Roughly speaking, the standard view on the semantics of such expressions assumes a difference between the linguistic meaning (analogous to the principle stating that in the case of 'I', 'I' refers to the speaker of the context; in Kaplanian semantics, such rules are represented as functions from contexts to contents), content expressed on a particular occasion (in Kaplanian semantics represented as functions from circumstances of evaluation—possible worlds in the simplest possible case—to extensions), and the extension of an expression with respect to particular circumstances of evaluation. One might expect, therefore, that the nDTM will provide at least a partial method of arriving at the rules that intuitively are linguistic meanings of indexicals and demonstratives. This is, however, highly problematic due to the following reason: apart from the highly specific case of a semantic trial mentioned earlier, the nDTM uses the notion of accepting a sentence in a situation that makes no reference to the context of use.

As we explained in section 1 among the types of meaning directives considered in the nDTM, the axiomatic and the inferential ones are usually mentioned. The former class contains the directives that specify that certain sentences must be accepted in every situation, while the latter specifies that certain sentences must be accepted if certain other sentences have been accepted. The additional constraint that axiomatic and inferential directives must follow the classic logic (and its standard extensions) does not have to be met by every possible language and set of directives. However, we assume that languages that meet such a constraint are much closer in spirit to natural language, at least if we consider some minimal set of logical truths and rules of inference. If this is the case, the nDTM must say something about the acceptance of sentences that contain indexical and demonstrative constituents and, at the very same time, are either exemplifications of logical truths or exemplifications of components of logically valid inferences. We know, however, that without the appropriate specification regarding the role of context, every such exemplification is prone to counterexamples.

Consider, for instance, the law of identity $A \Rightarrow A$ and its (*prima facie*) indexical exemplification:

(*) If you are a philosopher, then you are a philosopher.

(*) is false in cases in which the addressee of the first occurrence of 'you' in (*) is a philosopher and the addressee of the second occurrence of 'you' in (*) is not. Thus, at least on some occasions, (*) should be rejected. The matrix of the

language that contains (*) and respects the law of identity must therefore tell the difference between the occasions and the corresponding uses of (*).

To see another example, consider an inference that (again, *prima facie*) puts into question the validity of the conjunction elimination rule:

This is a conjunction, and Rome is situated on the Tiber.
Therefore: This is a conjunction.

The proponent of the nDTM should explain why this is not a real exemplification of the conjunction elimination rule at all, just like some uses of (*) are not exemplifications of the law of identity. One may produce examples of this sort concerning every tautology and rule of inference that any system of logic treats as valid.[5]

Note that some intuitive analyses are not available for the proponents of the nDTM. Saying, for instance, that tautologies and rules of inference represent propositions, not sentences, is not helpful at all because, firstly, our task is to construct a matrix of a language by appealing to the notion of an *acceptance of a sentence* and, secondly, define a linguistic meaning in terms of such a matrix. Propositional contents simply cannot do that job. The only way of solving the problem, it seems, is to introduce the context of utterance into the nDTM.

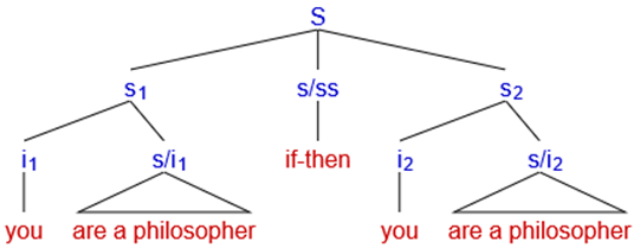
3. Indexicals and Demonstratives in the nDTM: A Partial (non-)Solution

In order to address the problem, we shall introduce the notion of *admissible contextual distribution* of an utterance. We borrow the idea in question from the theory of *contextual perdurantism* (Ciecierski 2019). Contextual perdurantism is one of several interpretations of the *hybrid name* view (other interpretations can be found in: Künne 1992, 2010; Textor 2007, 2015; Kripke 2011; see also Penco 2013 for a comparison between some of the theories) developed originally by Frege and according to which—in the case of indexicals and demonstratives—the circumstance becomes a part of the expression of the thought, that is they might be constituents of linguistic expressions themselves. This results in a special class of expressions being the hybrids of a purely linguistic component and external circumstances.

Contextual perdurantism treats each relevant aspect of context as an independent dimension of contextual space and treats utterances as *aggregates of contextual parts across actual distributions of contextual parameters* or, as one might put it, *contextually perduring objects*. Roughly speaking, admissible contextual distribution of an utterance is the sequence of potential values of contextual parameters. The parameters in question determine the semantic values of all indexicals and demonstratives[6] that are terminal elements of the syntactic analysis of the uttered sentence. Consider (*) again and assume that it is used in the context c, which embraces two potential candidates for the role of addressee: Lauben and Ligens. Given the (categorial) syntax of (*), that is:

[5] The problem in relation to Kaplanian semantics for indexicals and demonstratives has been studied extensively by many authors; some important works include: Kaplan 1989; Braun 1996; Predelli 2005; Radulescu 2015; Georgi 2020; McCullagh 2020.

[6] For the sake of simplicity, we ignore here important differences between the representation of indexicals and demonstratives (we shall sometimes speak about deictic expressions in order to cover both classes). Let us assume that in the former case, the distributions concern demonstrations (considered as aspects of contexts) conceived as complex actions constituted by intentions and an elementary action of indication.



We associate the following set of admissible contextual distributions with it:[7]

{<i_1 – Lauben, i_2 – Lauben>, <i_1 – Ligens, i_2 – Ligens>,
<i_1 – Lauben, i_2 – Ligens>, <i_1 – Ligens, i_2 – Lauben>}.

For terminal elements of the syntactic structure, we shall sometimes represent the fact that a particular terminal element is associated with a value of a particular contextual parameter in terms of an ordered pair: <expression, distribution>. We assume that distribution jointly with the meaning of an expression type determines the reference of an expression in several cases, the reference is just the object that is a value of a contextual parameter, so <expression, distribution> pairs might be indistinguishable from <expression, reference> pairs.[8]

We imagine that the last paragraph could have confused some of the readers. Wasn't nDTM supposed to be a non-referential theory? As explained earlier, the solution we propose should be understood as one of the extensions of the nDTM that go beyond its minimalist nature. We believe that it is quite interesting that the addition of indexicals and demonstratives forces us to add reference to the theory.

We assume that, depending on the speaker's intentions, only one such distribution corresponds to what is communicated in the utterance. We shall call such distributions *relevant*. Relevant contextual distributions represent actual contexts of utterance (note that this concept is defined partially in terms of the speaker's intentions). We might, therefore, treat a pair consisting of a sentence and a contextual distribution as a representation of a particular utterance.

In the original nDTM, the concept of acceptance is defined for sentences and circumstances understood in terms of input data to which the language user is exposed. Introducing contexts into the nDTM requires substantial modifications of that concept. Firstly, the notion of acceptance should now apply not to sentences but to utterances (our sentence-contextual distribution pairs). Secondly, the circumstances of acceptance must be separated from the context of utterance. Although we assume that within the frame of a particular semantic trial both co-occur, they are distinct (but potentially overlapping) aspects of the circumstances in which the utterance occurs and in which it is assessed. The proper subject of acceptance is the sentence with its (fixed) contextual distribution.

[7] Here the aspects of the context are identical with referents of a particular indexical, but in other cases it does not have to be like this: expressions like 'today' or 'yesterday' are linked to moments of utterance while referring to days that are appropriately related to the moments of utterance.

[8] One might also suggest (cf. Ciecierski 2019) that particular aspects of the contexts (contextual parameters) are qua-objects rather than regular objects (cf. Poli 1999; Werner 2020). If this is the case, then Aristotle ≠ Aristotle-addressee, but <*you*, Aristotle-addressee> necessarily refers to Aristotle.

We might observe that among all the distributions, there are the homogeneous ones: they assign the same value of a contextual parameter to each occurrence of a particular indexical (<i_1 – Lauben, i_2 – Lauben> and <i_1 – Ligens, i_2 – Ligens> in our example). We might, therefore, in the first attempt of adding the context to the new directival theory of meaning, consider the modification of the idea behind the notions of axiomatic and inferential meaning directives and define both concepts relative to homogeneous contextual distributions and use the modified notion of acceptance. In the case of axiomatic directives, we might consider:[9]

> If ***cd*** is the relevant and homogeneous contextual distribution of the sentence *S*, the user accepts the utterance <*S*, ***cd***> in every circumstance of acceptance *c*.

The extension of the idea to inferential directives looks as follows:

> If the user accepts utterances <s_1, $\boldsymbol{cd_1}$>…<s_n, $\boldsymbol{cd_n}$>, then she accepts the utterance <s_k, $\boldsymbol{cd_k}$> (for $k = 1…n+1$).

Here, however, we lose any track of the (inferentially relevant) connections between contextual distributions. For instance, the idea given above does not distinguish between (the reader is kindly asked to keep in mind that the first argument of <x, cd> pair is an expression that is mentioned, not used):

> <*This*, cd_1> is a conjunction, and Rome is situated on the Tiber.
> Therefore: <*This*, cd_1> is a conjunction.

And (for cd_1 different from cd_2):

> <*This*, cd_1> is a conjunction, and Rome is situated on the Tiber.
> Therefore: <*This*, cd_2> is a conjunction.

Additionally, it does not distinguish between:

> If <*you*, cd_1> are a philosopher, <*you*, cd_1> are poorer than <*me*, cd_1>.
> <*You*, cd_1> are a philosopher.
> Therefore: <*You*, cd_1> are poorer than <*me*, cd_1>.

And (again: for cd_1 different from cd_2):

> If <*you*, cd_1> are a philosopher, <*you*, cd_1> are poorer than <*me*, cd_1>.
> <*You*, cd_2> are a philosopher.
> Therefore: <*You*, cd_1> are poorer than <*me*, cd_1>.

One might think that to fix this, we assume that among all homogeneous contextual distributions there is a maximal one, which is the one that embraces distributions of all premises and the conclusion of the argument. Since the maximal distribution enables tracking the relevant connections between the values of contextual parameters instead of considering potentially different distributions of premises and conclusion, we *might assign to all of them single maximal distribution*. If $\boldsymbol{cd_m}$ is such a maximal distribution, this results in the following characteristics:

> If the user accepts utterances <s_1, $\boldsymbol{cd_m}$>…<s_n, $\boldsymbol{cd_m}$>, then she accepts the utterance <s_k, $\boldsymbol{cd_m}$> (for $k = 1…n+1$).

[9] In the case of both axiomatic and inferential directives, the complete formulations should include information that the user accepts a sentence or will be treated by the language community as not participating in a semantic trial or not to use a sentence in the meaning it has in a particular language (cf. Grabarczyk 2019: 177). For the sake of simplicity, we use here a shorter and more intuitive formulation.

So far so good. A moment's reflection shows that the solution works for some cases but definitely cannot be generalized into others. Consider an instantiation of the law of conjunction elimination:

(**) If you$_1$ are a philosopher and you$_2$ are a philosopher, then you$_3$ are a philosopher.

The requirement that the homogeneous contextual distribution is required here is way too strong. A particular utterance of (**) is an instantiation of the law of simplification if and only if the third occurrence of 'you' gains reference relative to the same value of the addressee parameter as any of the two other occurrences of 'you'. So, the homogeneity of contextual distribution is not required to arrive at the (intuitively) logically valid instantiation of the law of simplification. We must, therefore, drop the idea of homogeneous distributions and look for the solution elsewhere.

4. Indexicals and Demonstratives in the nDTM: Taking the Hybrid Syntax Seriously

To see what went wrong, we shall use Geoff Georgi's notion of referential promiscuity (Georgi 2020: 129). The characterization of the notion is:

REFERENTIAL PROMISCUITY (RP)
An expression *e* of a language *L* is referentially promiscuous if and only if there are distinct free occurrences O_1 and O_2 of e in a sentence *s*, and some context *c*, such that the content of O_1 relative to *c* is distinct from the content of O_2 relative to *c*.

According to these characteristics, *s* ranges over all sentences, both compound and non-compound. So, it is in principle possible that an expression meets the characteristics only if its content differs across occurrences in *distinct* atomic sentences that are constituents of some compound *s*. But there are actually two ideas of RP that are combined into (RP), namely:

ATOMIC REFERENTIAL PROMISCUITY (ARP)
An expression *e* of a language *L* is atomically referentially promiscuous if and only if there are distinct free occurrences O_1 and O_2 of *e* in an *atomic* sentence *s*, and some context *c*, such that the content of O_1 relative to *c* is distinct from the content of O_2 relative to *c*.

COMPOUND REFERENTIAL PROMISCUITY (CRP)
An expression *e* of a language *L* is compoundly referentially promiscuous if and only if there are distinct free occurrences O_1 and O_2 of *e* in atomic sentences s_1 and s_2, respectively, which are constituents of the compound sentence *s*, and some context *c*, such that the content of O_1 relative to *c* is distinct from the content of O_2 relative to *c*.

(ARP) entails (CRP): if an expression meets (ARP), then combinatorial syntax warrants that there will be cases of (RP) involving this expression in the domain of compound sentences. On the other hand, the reverse entailment depends on the syntactic status of *e* and the vocabulary of the language *L*: if, for instance, *e* is a singular expression and *L* contains relational predicates, (CPR) entails (APR).

The idea of (ARP) points out that the phenomenon that needs to be captured in the logic of indexicals and demonstratives concerns the occurrences of deictic expressions within atomic formulas. This suggests that our concept of contextual

distribution should apply to atomic formulas rather than to compound formulas. Our initial example (*) should, therefore, be associated with the following analyses that provide us with the list of all admissible contextual distributions (the '[' and ']' brackets mark the syntactic structure):

(1) [if-then, [<[*you*, *are a philosopher*], <Ligens, Ligens>>], [<[*you*, *are a philosopher*], <Ligens, Ligens>>]]
(2) [if-then, [<[*you*, *are a philosopher*], <Ligens, Lauben>>], [<[*you*, *are a philosopher*], <Ligens, Lauben>>]]
(3) [if-then, [<[*you*, *are a philosopher*], <Lauben, Ligens>>], [<[*you*, *are a philosopher*], <Lauben, Ligens>>]]
(4) [if-then, [<[*you*, *are a philosopher*], <Lauben, Lauben>>], [<[*you*, *are a philosopher*], <Lauben, Lauben>>]]

from which—if we note that the second argument of the contextual distribution is vacuous when paired with the antecedent of the conditional, while the first argument of the contextual distribution is vacuous when paired with the consequent of the conditional, that is on both occasions there are no indexical expressions with which the respective arguments combine to constitute the hybrid expression—we might easily get:

(1*) [if-then, [<*you*, Ligens>, are a philosopher], [<*you*, Ligens>, are a philosopher]]
(2*) [if-then, [<*you*, Ligens>, are a philosopher], [<*you*, Lauben>, are a philosopher]]
(3*) [if-then, [<*you*, Lauben>, are a philosopher], [<*you*, Ligens>, are a philosopher]]
(4*) [if-then, [<*you*, Lauben>, are a philosopher], [<*you*, Lauben>, are a philosopher]]

It might be useful to coin some terminology here: let us call the forms (1)-(4) (and analogous forms) *hybrid propositional forms* and the forms (1)*-(4)* (and analogous forms) *hybrid nominal forms*. The process of arriving at the list of hybrid sentences is partially bottom-up (atomic sentences are paired with contextual distributions; the occurrences of indexicals and demonstratives select the non-vacuous aspects of contexts) and partially top-down (a contextual distribution as a whole takes into account the entire compound sentence).

Pairing contextual distributions with atomic sentences enables us to track down the occurrences of indexicals and demonstratives as well as to pair them with different values of contextual parameters. The result is the propositional language of *utterances* (LU) with the following *hybrid syntax*.

The class of symbols of LU:
$<\mathbf{P}, \mathbf{CD}, \{\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow\}, \{(,)\}>$

where **P** is the set of propositional variables, and **CD** is the set of potential contextual distributions. We assume now that not all elements of $\mathbf{P} \times \mathbf{CD}$ are possible utterances of LU: in order for some $<p, cd> \in \mathbf{P} \times \mathbf{CD}$ to count as a possible utterance, the contextual parameters of *cd* must provide values for all indexicals and demonstratives in *p*. For the purpose of the current presentation, we might assume that we have at our disposal a relation **R** (we might treat it as primitive or attempt to define it) that selects all and only such pairs from $\mathbf{P} \times \mathbf{CD}$. Having this, we might define a potential non-compound utterance:

Definition (potential non-compound utterance)
u is a potential non-compound utterance of LU iff $u \in \mathbf{P} \times \mathbf{CD} \mid \mathbf{R}$

(that is: u counts as a potential non-compound utterance if and only if it is a pair consisting of a propositional variable and a contextual distribution such that the contextual distribution provides all values for indexicals and demonstratives occurring in the propositional variable).

Next, we define the class of potential utterances of LU in the following manner:

1. All non-compound potential utterances of LU are potential utterances of LU.
2. If $<p, cd>$ is a non-compound potential utterance of LU, then '$\neg<p, cd>$' is a potential utterance of LU.
3. If $<p, cd>$ and $<p, cd'>$ are non-compound potential utterances of LU, then '$<p, cd> \wedge <p, cd'>$', '$<p, cd> \vee <p, cd'>$', '$<p, cd> \Rightarrow <p, cd'>$', and '$<p, cd> \Leftrightarrow <p, cd'>$' are potential utterances of LU.
4. If u is a potential utterance of LU, then '$\neg u'$ is a potential utterance of LU.
5. If u and u^* are potential utterances of LU, then 'u $\wedge$ u*', 'u $\vee$ u*', 'u $\Rightarrow$ u*', and 'u $\Leftrightarrow$ u*' are potential utterances of LU.
6. Nothing else is a potential utterance of LU.

Clauses 2-5 are necessary to allow formulas like '$(<p, cd> \vee <p, cd'>) \Rightarrow <p, cd''>$', as potential utterances of LU while disallowing as potential utterances formulas such us '$<<p, cd>, cd'>$': we assume that contextual distributions apply to propositional variables only.

Thus conceived, the (propositional) language of LU is just a propositional language that replaces standard propositional variables with hybrid expressions of the form $<p, cd>$. So, there cannot exist any logical differences between what counts as a logically valid formula or inference in the standard language of propositional logic and LU. For instance, the tableaux system for LU is just the standard system but allows instantiations of correct elimination rules, like:

[$\neg \Rightarrow$ Elimination]
$\neg(<p, cd> \Rightarrow <p, cd'>)$

$<p, cd>, \neg<p, cd'>$

[$\wedge$ Elimination]
$<p, cd> \wedge <p, cd'>$

$<p, cd>, <p, cd'>$

which enable, for example, the following derivations:

[1] $\neg[<p, cd> \wedge <p, cd'>) \Rightarrow <p, cd>]$	[Assumption]
[2] $<p, cd> \wedge <p, cd'>$	[$\Rightarrow$ *elimination,* [1]]
[3] $\neg<p, cd>$	[$\Rightarrow$ *elimination,* [1]]
[4] $<p, cd>$	[$\wedge$ *elimination,* [2]]
[5] $<p, cd'>$	[$\wedge$ *elimination,* [2]]
Contradiction: [3]-[4].	
[1]* $\neg[<p, cd> \wedge <p, cd'>) \Rightarrow <p, cd''>]$	[Assumption]
[2] $<p, cd> \wedge <p, cd'>$	[$\Rightarrow$ *elimination,* [1]]
[3] $\neg<p, cd''>$	[$\Rightarrow$ *elimination,* [1]]

[4] <*p, cd*> [∧ *elimination,* [2]]
[5] <*p, cd'*> [∧ *elimination,* [2]]
No contradiction.

This shows that [1] *is* and [1]* *is not* an instantiation of the '(A ∧ B) ⇒ A' law. This is exactly what we want to have when it comes to propositional tautologies and inferences. This, for instance, enables also distinguishing valid cases of:

This is a conjunction, and Rome is situated on the Tiber.
Therefore: This is a conjunction.

Namely:

<This is a conjunction, "This is a conjunction, and Rome is situated on the Tiber"> and <Rome is situated on the Tiber, *cd*>[10]

<This is a conjunction, "This is a conjunction, and Rome is situated on the Tiber">

and invalid ones, that is:

<This is a conjunction, "This is a conjunction, and Rome is situated on the Tiber"> and <Rome is situated on the Tiber, *cd*>

<This is a conjunction, "This is a conjunction">

(Note that the contextual distribution must contain as a demonstration or demonstratum the expression without the contextual distribution; otherwise, we arrive at an infinite regress.)

This, of course, leaves the problem we are facing partially open as we also need an account that would enable us to have adequate treatment of cases in which we are dealing with distinct but appropriately related contextual distributions, as potentially in:

This is bigger than this. This is bigger than this.
Therefore: This is bigger than this.

which might be an analytic statement also in cases different than:

<*This is bigger than this*, cd>, <*This is bigger than this*, cd>.
Therefore: <*This is bigger than this*, cd>

which is just a matter of propositional logic.

The solution naturally appeals to our distinction between hybrid propositional and *nominal propositional forms*. The solution that works for propositional entailments made use of the former by identifying the non-compound propositional formulas with formulas having hybrid propositional forms. *Per analogiam*: we might dwell into the structure of non-compound indexical formulas by identifying them with formulas having hybrid nominal forms. This would enable us to trace referentially the same occurrences of indexicals and treat them as occurrences of a single referential expression. In that manner, we might arrive at:

<*This*, a> is bigger than <*this*, b>, <*This*, b> is bigger than <*this*, c>.
Therefore: <*This*, a> is bigger than <*this*, c>

which becomes the case of:

x is bigger than *y*, *y* is bigger than *z*.

[10] Since the second conjunction contains no indexicals, it might be paired with an arbitrary contextual distribution.

Therefore: *x* is bigger than *z*.

which cannot be said of:

<*This*, a> is bigger than <*this*, b>, <*This*, c> is bigger than <*this*, d>.
Therefore: <*This*, a> is bigger than <*this*, d>

The solution we arrived at bears interesting relation to the view on the logical form developed by Iacona (2013; see also Georgi 2020: 138-39), which assumes that distinct symbols refer to distinct objects. On the one hand, nothing in our view prohibits co-referential hybrid expressions like <*I*, a> and <*you*, a>, which directly contradicts Iacona's assumption. At the same time, our view automatically secures the referential difference for any pair of expressions <*x*, a> and <*x*, b> by treating them—if a and b are parameters that point to different objects—as *distinct hybrid symbols*. What is important here is that enabling hybrid syntax enables us to keep the intrinsicality assumption regarding logical form, which is the claim that the logical form of an argument is intrinsic to the syntax of the argument.

Our considerations suggest that the properties such as validity and analyticity bear no special complications for indexical languages. From the viewpoint of the nDTM, we, therefore, have at our disposal the method of telling apart the uses of deictic sentences that are candidates for axiomatic and inferential rules from the uses of the very same sentences that are not such candidates. However, one might question this conclusion, if there are exemplifications of correct inferences or indexical tautologies that cannot be distinguished—using the method outlined above—from incorrect inferences or possibly false uses of sentences of the same form. In other words, the opponent of the view presented above must indicate cases of uses of indexical sentences or indexical reasonings that are sometimes valid and sometimes invalid and show that the difference between them cannot be captured by appealing to the idea of hybrid syntax.

Let us, therefore, consider two candidates for deictic reasonings:

(5) Josh says "It is raining today". The next day, Josh says "It rained yesterday" (Radulescu 2015: 1844).

(6) I utter, "David Kaplan is older than I". Then somebody addresses me with: "Therefore, you are not older than David Kaplan" (cf. Bar-Hillel 1963).

It is probably not obvious that the examples are cases of reasonings but let us suspend the skepticism regarding that issue and stick rather to the observation that there is a sense in which one could correctly claim that—in the appropriate contexts—one cannot consistently hold one of the utterances listed in (5) or (6) while denying the second. This, we think, suffices for treating (5) and (6) as cases of inferences. Our hybrid nominal forms of both arguments are respectively (with enthymematic assumptions made explicit):

It is raining <*today*, d_0>.

It rained <*yesterday*, d_1>[11]

David Kaplan is older than <*I*, a>.
For every pair of objects: if the first is older than the second, then the second is not older than the first.

<*You*, a> are not older than David Kaplan.

[11] d_0 and d_1 are the consecutive days at which today and yesterday are respectively uttered.

Here is the way in which one might attempt to analyze such cases. Since '<*today*, d_0>' and '<*yesterday*, d_1>', <*I*, a> and <*You*, a> are respectively co-referential, in order to secure the inferences, we need nothing more than the assumption that co-referential hybrid terms are interchangeable *salva veritate*. In fact, the idea in question is nothing more than the idea of direct reference, and indeed, we believe that indexicals and demonstratives are directly referential.[12] One may worry, however, that something is missing here because we do not have an explanation of the fact that the appropriate change of indexicals (from to 'today' to 'yesterday' and from 'I' to 'you') is required for arguments to be valid. Moreover, it says nothing about reasonings such as:

(7) Josh says, "Today is Monday". Then he says, "Therefore, tomorrow is Tuesday"

which contain indexical expressions that are not co-referential. This, however, is not a problem, as the theory might claim that inferences of this sort might be simply described by appealing to perfectly general enthymematic assumptions of the arguments. Note that the complete hybrid nominal form of the argument looks as follows:

<*Today*, d_1> is Monday.
<*Today*, d_1> precedes <*Tomorrow*, d_1>.
For any pair of days: if the first precedes the second and the first is Monday, then the second is Tuesday.

<*Tomorrow*, d_1> is Tuesday.

All specifically indexical premises of the argument are empirical, and the validity of the entire argument hangs on the correctness of the perfectly general assumption three, which describes the simple arithmetic of days and which, without any doubt, should be included among the class of axiomatic directives for 'Monday' and 'Tuesday'.

Other authors working on the logic of indexicals prefer to analyze the cases in terms of constraints on relations between contexts for premises and the conclusion (Radulescu 2015) or the coordination of contexts of premises and the conclusion of an argument (Georgi 2020). This approach avoids this: it aims to achieve precisely the same results only by appealing to hybrid syntax, the interchangeability principle (and this is a referential principle!) (or principles, if other types of designators are at stake), and perfectly general meaning postulates.

It should be observed additionally that the existence of specific indexical tautologies[13] (the thesis that such tautologies exist was the subject matter of intensive criticism in recent decades) is not the problem for the nDTM unless the sentences occurring in such tautologies have no corresponding false uses *and* unless the nDTM is unable to describe the difference correctly. We simply deny that there are such cases. The language matrix we shall arrive at will contain all the connections and regularities for indexicals and demonstratives and, by the same token,

[12] It also reminds of the version of Fregean semantics for indexicals that incorporates the idea of the primitive sense (cf. Tichy 1986) according to which (as originally for Frege), for instance, 'yesterday' and 'today' as considered in the appropriately related contexts have the same sense.

[13] Among the candidates, one might mention:
I exist.
I am self-identical (Yagisawa 1993: 480-82).

will enable us to read all the regularities and systematic dependencies between uses of indexicals in various contexts. We see no reason for claiming that some of such regularities will not be reflected in the matrix.

The solution sketched above might be treated as a general account of indexical validity, but since our motivation was to apply it to the nDTM, we now should consider the issue of the general form of axiomatic and inferential directives in the nDTM. The approach sketched above entails that no special modification—in addition to introducing the hybrid syntax into the picture—is needed. In the case of axiomatic directives, we have:

> If ***cd*** is the nominal contextual distribution of the sentence *S*, the user accepts the utterance <*S*, ***cd***> in every circumstance of acceptance *c*.

while in the case of inferential directives, we simply have:

> If the user accepts utterances $<s_1, cd_1>\ldots<s_n, \boldsymbol{cd_n}>$, then she accepts the utterance $<s_k, \boldsymbol{cd_k}>$ (for $k = 1\ldots n+1$) for nominal contextual distributions: $\boldsymbol{cd_1}\ldots\boldsymbol{cd_k}$.

However, this gives us a completely new idea of the language matrix and two distinct concepts of the meaning of an expression. Let us explain. From now on, language matrices apply to utterances (i.e., sentences in their hybrid nominal forms). The context of utterance becomes, therefore, a part of the syntactic structure. The circumstances of acceptance apply now to utterances. So, distribution within the language matrix is defined in terms of utterances in which an expression (hybrid or non-hybrid) occurs. It follows that notions such as meaning or synonymy are also defined relatively to utterances. Therefore, we have at our disposal the idea of the utterance's or use's meaning that replaces the original idea of an expression-type meaning. But one might still, if necessary, look at the distribution of expression types within the language matrix. In cases of indexicals and demonstratives, the distribution will simply involve an abstraction from the contextual parameter that is present in the hybrid nominal forms of a particular expression type. This is the second idea of the meaning of an expression that we might arrive at within the theory. The latter concept of meaning is the closest we might get to the intuitive idea of the linguistic meaning of an indexical.

5. Conclusion

As we saw, the original nDTM does not have an easy way of handling indexicals and demonstratives and remains prone to counterexamples, such as the one presented in (*). In this paper, we have sketched the way in which context can be added to the nDTM. We argue that the main reason why this cannot be done easily is that the theory operates only on the notion of an acceptance of the sentence (and not the notion of acceptance of an utterance). We presented an extension of the nDTM that modifies its key notion of a meaning directive, adding the parameter of contextual distribution to it. This, among other things, enables introducing into the nDTM the representations of particular utterances and distinguishing between the context of an utterance and circumstances of acceptance. The resulting view that wholly incorporates the idea of hybrid syntax avoids special complications regarding indexicals and demonstratives as occurring in the axiomatic and inferential directives.

The framework presented in this paper says nothing about various other functions the expressions might play in particular contexts, nor about the broadly conceived idea of context-dependence. It has been suggested by some (cf.

Ciecierski 2021) that this would require a generalization of the notion of meaning that incorporates the idea that, on the one hand, the expressions might be used in several manners and that, on the other, its meaning is something that enables *all* such different uses and tells us when uses become abuses of an expression. This would require a theory of context sensitivity that presupposes a particular account of the functions of language. Such a theory should also describe how contexts affect the change of particular linguistic functions and how they affect the change of one use of an expression into another. Our modest concept of contextual distribution is not designed for this role. It is also not designed as a tool for dealing with the examples discussed within contemporary contextualism-minimalism debates. We do not consider this feature as a drawback, however, as narrowly conceived indexicality generates problems that are challenging themselves. Considering them in the context of the problem of alleged context-sensitivity of *red*, *know* or *tasty* might only obscure the nature of the problems with *I*, *here* or *now*.

The approach sketched above says also nothing about the functioning of indexical expressions in empirical and promotive directives. This issue—due to its complexity—requires a separate study. The main reasons for this are ambiguities related to the question of whether certain epistemological and motivational aspects that systematically accompany the use of indexical expressions should be represented at the level of their meaning. For example, certain direct speech acts seem to require particular personal pronouns. Requests, thanks, congratulations, and advice require either a second person pronoun, promises require the first-person pronoun, thanks require both types of pronouns, etc. Uses of these pronouns (given the appropriate attitudes of the participants of the situation) are systematically connected with the actions of the speaker and the hearer. Should the matrix of a particular language take these actions and behaviors into account when a list of promotive directives is considered? If so, should the category of the promotive directive include not only behaviors but also a change in the normative situation of the sender and receiver of an utterance, an aspect stressed by speech act theorists?

Considering the empirical directive category also leads to problems. For instance, in a semantic trial, we might show someone an object (e.g., a book) and ask: "Is this a book?" or "Is this a zebra?". We will consider a negative or affirmative answer to these questions as relevant for the meaning of nouns *book* and *zebra*, but rather not to the meaning of the demonstrative *this*. Someone may explain this asymmetry and say that the pronoun *this* does not occur in the indicated sentences in an essential manner. But what determines the essential or nonessential occurrence of this pronoun? Let us imagine, for example, that in a certain situation there are two objects in the environment—a book and a zebra, the first one closer and the second one far from the participants of the semantic trial. If the question is now asked, "Is this (we gesture to point to a book) a zebra, and that (we gesture to point to a zebra) a book?" and the user denies, we seem to be faced with a dilemma: either we recognize that she does not know the meaning of one of the respective nouns, or that she does know their meaning but somehow takes into account that aspect of the meaning of *this* and *that* which makes the former pronouns refer to closer and the latter to farther located objects. A directival theory of meaning that accounts for the phenomenon of indexicality and applies to languages with empirical and promotive directives must somehow resolve such issues.

Regardless of directions the further analysis should take, we would like to emphasize that the approach presented in this paper, which takes into account only axiomatic and deductive directives, can also be applied outside the new directival theory of meaning. This is because it enables us to solve the problem of the existence of specifically indexical tautologies and inferences in an elegant way. The cost we pay here is the adoption of the idea of hybrid syntax, the principle of interchangeability, and the inclusion of enthymematic premises in the description of indexical inferences. The solution presented in this paper is, for this reason, not only an attempt to supplement the directival theory of meaning but also a voice in the general discussion about the logic of indexicals and demonstratives.[14]

References

Ajdukiewicz, K. 1934, "Sprache und Sinn", *Erkenntnis*, 4, 1, 100-38.

Ajdukiewicz, K. 1978, "Language and Meaning", in Giedymin, J. (ed.), *The Scientific World-Perspective and Other Essays*, 1931-1963, Dordrecht: Springer, 35-66.

Bar-Hillel, Y. 1963, "Can Indexical Sentences Stand in Logical Relations?", *Philosophical Studies*, 14, 6, 87-90.

Braun, D. 1996, "Demonstratives and Their Linguistic Meanings", *Nous*, 30, 145-73.

Ciecierski, T. 2019, "Hybrid Expressions", in Bella, G. and Bouquet, P. (eds.), *Modeling and Using Context*, Berlin: Springer, 51-61.

Ciecierski, T. 2021, "Indexicality, Meaning, Use", *Semiotica*, 238, 73-89.

Georgi, G. 2020, "Demonstratives in the First-Order Logic", in Ciecierski, T. and Grabarczyk, P. (eds), *The Architecture of Context and Context-Sensitivity: Studies in Linguistics and Philosophy*, Vol. 103, Cham: Springer, 125-48.

Grabarczyk, P. 2017, "Directival Theory of Meaning Resurrected", *Studia Semiotyczne*, 31, 1, 23-44.

Grabarczyk, P. 2019, *Directival Theory of Meaning*, Cham: Springer.

Iacona, A. 2013, "Logical Form and Truth-Conditions", *Theoria: An International Journal for Theory, History and Foundations of Science*, 28, 3, 78, 439-57.

Kaplan, D. 1989, "Demonstratives", in Almong, J., Wettstein, H., and Perry, J. (eds), *Themes from Kaplan*, Oxford: Oxford University Press, 481-563.

Kripke, S. 2011, "Frege's Theory of Sense and Reference: Some Exegetical Notes", in Kripke, S., *Philosophical Troubles: Collected Papers*, Oxford: Oxford University Press, 254-91.

Künne, W. 1992, "Hybrid Proper Names", *Mind*, 101, 721-31.

Künne, W. 2010, "Sense, Reference and Hybridity", *Dialectica*, 64, 529-51.

[14] The work on this paper was funded by National Science Center, Poland, grant under award number 2018/29/B/HS1/01868. We would like to express our gratitude to Dominik Dziedzic, Maria Matuszkiewicz, Carlo Penco, Jakub Rudnicki as well as to participants of the 2021 Workshop on Context: Semantics, Pragmatics and Cognition. Additionally, we want to express our gratitude to the anonymous reviewer of *Argumenta* for numerous important comments that helped us to revise the original submission.

McCullagh, M. 2020, "Distributed Utterances", in Ciecierski, T. and Grabarczyk, P. (eds.), *The Architecture of Context and Context-Sensitivity, Studies in Linguistics and Philosophy*, Vol. 103, Cham: Springer, 113-24.

Penco, C. 2013, "Indexicals as Demonstratives: On the Debate between Kripke and Künne", *Grazer Philosophische Studien*, 88, 55-73.

Poli, R. 1999, "Qua-theories", in Albertazzi, L. (ed.), *Shapes of Forms*, Dordrecht: Kluwer, 245-56.

Predelli, S. 2005, *Contexts: Meaning, Truth and the Use of Language*, Oxford: Oxford University Press.

Radulescu, A. 2015, "The Logic of Indexicals", *Synthese*, 192, 6, 1839-60.

Textor, M. 2007, "Frege's Theory of Hybrid Proper Names Developed and Defended", *Mind*, 116, 947-81.

Textor, M. 2015, "Frege's Theory of Hybrid Proper Names Extended", *Mind*, 124, 823-47.

Tichy, P. 1986, "Frege and the Case of the Missing Sense", *Grazer Philosophische Studien*, 27, 1, 27-47.

Werner, K. 2020, "The Little Word 'As': On Making Contexts and Aspects Explicit", *Axiomathes*, 30, 1, 69-90.

Yagisawa, T. 1993, "Logic Purified", *Nous*, 27, 4, 470-86.

Proper Names as Demonstratives

Maciej Tarnowski

University of Warsaw

Abstract

The paper considers the hypothesis that proper names are simple demonstratives. In the first part, I provide the general motivation for an indexical treatment of proper names as well as assess the strengths and weaknesses of existing indexical accounts. The second part is devoted to proposing a new account that treats proper names as simple demonstratives, where referents are determined by the speaker's referential intention. In my proposal, I use the hybrid approach toward indexical expressions developed by Wolfgang Künne (1992) and Stefano Predelli (2006). I argue that this approach allows countering many of the problems haunting existing indexical accounts of proper names. I also consider the two possible objections: the Humpty-Dumpty objection to intentionalism regarding demonstrative reference (Gorvett 2005) and García-Carpintero's (2018) argument against indexicalism from misdescription of linguistic competence of proper name users. I show how the proposed approach may counter them. The considerations concerning this problem also demonstrate how the treatment of proper names as hybrid demonstratives may allow solving the problems posed by the "Madagascar argument" to the causal-chain theory of proper name reference (Evans 1973) as well as explain the presence of predicative uses of proper names in natural language (Burge 1973).

Keywords: Proper names, Demonstratives, Hybrid expressions, Intention.

Proper names serve as a long finger of ostension over time and place.
Barcan Marcus (1986: 122)

1. Motivation for Indexical Treatment of Proper Names

In recent years there has been a substantial number of proposals treating proper names as a class of indexicals (see Recanati 1993; Pelczar & Rainsbury 1998; Rami 2014), which aim at accommodating the directly referential nature of proper names (Kripke 1972) and the explanation of "proper name ambiguity" or "nambiguity" (Korta, Perry 2011), which troubles many classic proposals of semantic analysis of proper name reference, such as Kripkean causal chain theory. The phenomenon of nambiguity may be captured by appealing to the intuitive notion of "name-sharing": we are inclined to say, for example, that there are at least two

Argumenta (2022): 55-71
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.tar

people named 'Alfred' who deserve wide recognition for their contribution to mathematical logic, or that in the past year 'Michael' was the most popular male name in the United States. As an example, we may consider the following statement (Korta & Perry 2011: 75):

(1) If John would quiet down, John could hear what John is saying

where each token of 'John' is intended to refer to a different individual at some meeting. Intuitively, all three mentioned individuals share a property: "being called 'John'". According to traditional views on proper name semantics, this claim, however, should be regarded as false or metaphorical.[1] According to a causal chain theorist, every token of the word 'John' in (1) is a token of a different name-type identified by its causal history. Proper names as types however, *contra* causal chain theorist, do not seem to be tied to a particular individual. It is even more visible in cases where the use of a proper name is preceded by a quantifier or pluralised, as in:

(2) Every John I met during this visit was nice to me.

(3) There are several Johns in the room right now.

After Burge (1973), many (e.g., Bach 2002) regard this observation as an argument for a different treatment of proper names as metalinguistic predicates. Such treatment allows for expressing general statements similar to (2) and (3), while singular uses of proper names are treated as preceded by an unarticulated demonstrative expression (as in "[That] John was nice to me"). On the other hand, many regard this picture as lacking a core appealing property of causal chain theories, which is the directly referential character of proper names, since metalinguistic predicates are not rigid designators. Statements like:

(4) It is possible that John could have not been named 'John'

seem similarly intuitively true as (2) and (3); however, under the predicativist view, it is hard to provide an analysis that would yield a similar result,[2] since the

[1] Kaplan (1990) disagrees: using his distinction between common-currency and generic names (108-10) he tries to maintain that although all the three people called 'John' in this example have different *common-currency names,* which name particular individuals in a manner similar to logical constants, they have the same *generic name*, which are "the genera, or species, of our individual common currency names" (108). However, using such distinction to interpret (1) is problematic. According to Kaplan, a generic name "doesn't name any*one*, perhaps it names or is an unnatural kind" (111); therefore one cannot be called or named with a generic name. I thank the anonymous reviewer for pointing this out.

[2] This is true at least of two popular proposals concerning the semantics of statements of the form "That F is G", which take the predicate F to have direct input into the statement's truth conditions (e.g., Neale 1993; Richard 1993). Theories that reject this premise (e.g., Larson & Segal 1995), on the other hand, are difficult to be regarded as consistent with the predicativist account. As pointed out by the anonymous referee, "[That] John" could also receive a quantifier interpretation [The x: x is called 'John'], which, assuming the wide scope of the description, would yield the desired verdict. Such proposals may be however subject to further similar counterexamples. Consider the pair of statements: "John is a father. This could have been true". The interpretation yielded by such a proposal would be that this pair of statements is true if in some worlds the demonstrated individual (even if he's called 'Tony' and is childless) is called 'John' and is a father—which seems counterintuitive assuming the predicativist account. Of course, the point of this argument is more modest than saying that predicativist theories cannot provide a satisfying account of modal rigidity of names; I rather wish to argue that it certainly is an obstacle for these theories, and a theory which avoids it should be preferred if available. Some predicativists argue for that reason that proper names are, in fact, non-rigid and try to explain away this intuition (see Bach 2002: 85-88).

first occurrence of 'John' in (4) refers to only those individuals who are named 'John' in a given possible world.

The indexicalist theory of proper names—a semantic theory that identifies them as a class of indexicals—seemingly allows for an intuitive accommodation of these two properties of names: the possibility of ambiguity and their modal rigidity. Pure indexicals such as 'I' or 'now' or demonstratives such as 'that' clearly exhibit desired properties: neither has a fixed referent if taken as types, but their utterances in context are directly referential and rigid. It is also noted in the literature, that indexicals may have uses similar to the ones exhibited in (2) or (3), such as: "Everyone of them were nice to me" or "There are several *shes* in the room right now" (see Kijania-Placek 2016; I write more about such uses in Section 4.2). Indexicalist accounts defended by Recanati (1993), Pelczar & Rainsbury (1998), which may be regarded as the most influential, liken proper names to 'pure indexicals', expressions that refer to objective features of the context of the utterance (picked by what Kaplan [1989a] calls a 'character' of an expression) rather than the intention or gesture provided by the speaker. Others call for an analysis in terms of complex demonstrative expressions, such as "that *N*" (Rami 2014). In my paper, I want to take a different route and show how proper names may be characterised as simple demonstratives. I will first discuss the problems of the mentioned indexical theories and show how a demonstrative account may address them. Secondly, I will sketch a demonstrative approach based on the notion of a "hybrid proper name" developed, among others, by Künne (1992, 2010), Textor (2015), and Penco (2013). Thirdly and lastly, I will show the benefits of such treatment and counter two possible objections to this view.

2. Problems of Existing Indexical Theories

Although an indexical picture of proper name reference may be found in many texts of early analytic philosophy of language (notably Russell 1950 and Dąmbska 1949), the first two major proposals arguing for it are the ones of Recanati (1993) and Pelczar and Rainsbury (1998). While Recanati states his theory in terms of different modes of presentation, Pelczar and Rainsbury prefer to speak in terms of 'character' and 'content' borrowed from Kaplan (1989a), but their approach may be treated as similar.[3] In this section, I will use Pelczar and Rainsbury's account as representative.

Pelczar and Rainsbury, directly inspired by Kaplan's treatment of indexical expressions, take proper names to be indexicals which reference is established by the contextual presence (or "being in force") of the dubbing, that is the act of naming a particular individual. In standard formalizations of statements containing indexical expressions the context set—paired with the sentence-type to yield appropriate interpretation—consists of the agent, time, location and the possible world of utterance, with different pure indexicals being sensitive to different of these parameters. The utterance of 'I' is sensitive to the change in agent-parameter, while 'here', 'now' and 'actual' to the changes in time, location and possible world parameters (Pelczar & Rainsbury 1998: 294-95). According to Pelczar and Rainsbury, similar treatment and sensitivity may be offered for utterances of proper names, if we assume that dubbings are also parameters of the context:

[3] For a direct comparison of Recanati's and Pelczar and Rainsbury's theories, see Matushansky 2008: 595, 597, 599; Ridley 2016: 108.

> Echoing Kaplan's insight with respect to the indexical, 'I', we say that what a competent English speaker knows, as such, about the word 'John' is not its content with regard to some particular occasion of its use. Rather, a competent English speaker with the word 'John' in his vocabulary knows the rule which assigns to each context of utterance of 'John' a function which assigns to each possible world whatever was dubbed 'John' in the dubbing in force governing 'John' in that context of utterance. Competent speakers know that the proper use of 'John' is—loosely speaking—to refer to something called 'John'. What someone knows by virtue of knowing this rule is the character of 'John', not its content (or, extension) with respect to a given utterance of it. The meaning of a proper name, then, is to be identified with its character (Pelczar & Rainsbury 1998: 297).

Given such characteristics, we may interpret their remarks[4] as suggesting that proper names are pure indexicals like 'I', 'here' or 'now', which are assigned additional contextual parameters, dubbings, when uttered:

(Indexicalism 1) A character of a proper name 'N' assigns different utterances of 'N' in the context c a certain dubbing being in force[5] in c.

According to Pelczar and Rainsbury, we might interpret utterance of (1) as being supplied with a contextually salient (brought to the attention of the speakers and thus being in force) *set of dubbings—d_1, d_2, d_3*—which are assigned to different uses of 'John' in (1). We may therefore speak of the same name (contrary to the causal-chain view) being uttered three times in (1), while still those distinct uses rigidly refer to different bearers, which are assigned to different uses of 'John' in (1).

A commonly raised objection to (Indexicalism 1) is that it relies heavily on the notion of contextually salient dubbing (or naming convention). Although it seems intuitive, the way of bringing certain dubbing into force (and making it a constituent of the context set) is not provided by the authors. A common way to describe it would be to employ a Gricean maxim of quality[6]—for example, the circumstances of the utterance of (1) may make it clear who are the intended referents of uses of the name 'John' given that the speaker wants to convey meaning

[4] This claim comes from the parallel that Pelczar and Rainsbury draw between proper names and pure indexicals rather than demonstratives (although the distinction is, of course, vague and hotly debated). In fact one may interpret this theory in an intentionalist way, where the appropriate dubbing comes into force through the intention of the speaker (and therefore reference of a proper name utterance is intention-dependent). Pelczar (2001) in fact suggests such a reading when he writes that choosing the appropriate dubbing is a matter of "speaker's discretion" (149-50), however he also states that "[a]ny effort to state precisely the conditions under which a given dubbing of a thing with a particular name is in force (in any context) quickly leads to the consideration of multifarious factors, semantic, pragmatic, and even extra-linguistic, that, to say the least, resist tidy encapsulation" (143). If Pelczar and Rainsbury's theory is interpreted as an intentionalist-demonstrative treatment of proper names, their theory is still subject to similar objections as Rami's theory criticized further in the article. I would like to thank an anonymous reviewer for pointing this out.

[5] In line with Recanati's interpretation, one should speak here of "naming conventions" instead of "dubbings-in-force".

[6] Pelczar and Rainsbury suggest, although do not explicitly commit to, such reading of their theory: "one of the competing dubbings must be brought to prominence in order to determine a unique referent for the name (in that use). This might be achieved by a variety of mechanisms. One important factor in this raising to prominence might be relevant features of the conversation (if any) of the context of utterance of the name" (1998: 295).

in accordance with conversational rules. However, this approach would make the semantic value of a proper name dependent on pragmatic conventions. If we grant that the speaker of (1) might refer to Johns in the room in a different order or to some other people named 'John' in this situation and do so without bringing their dubbings into the conversation, the Gricean way is not a way out of this problem.

One might also notice that there is a problem in sentences like (1) with pairing different uses of 'John' with the contextually salient dubbings, given that in Kaplan-style semantics we use a whole sentence as a point of reference and assign the appropriate values of the context to utterances within it. If we look at the statement:

(5) John is here now

according to Pelczar and Rainsbury, we should represent it as a pair of the sentence type and the context set:

(5') < ["John is here now"], <a, t, l, @, d_1>>

and in the standard manner pick t as a referent of 'now', l as a referent of 'here', and d_1 as a dubbing picking out the referent of 'John'. However, if we will represent (1) in the same manner:

(1') < ["If John would quiet down, John could hear what John is saying"], <a, t, l, @, d_1, d_2, d_3>>

then we have no semantic rule that picks, for example, d_1 instead of d_3 as an appropriate dubbing for the first occurrence of 'John' in (1). Even if there was a clear way of bringing dubbings into force, then matching of utterances with them still seems to employ some pragmatic clues. It seems doubtful, therefore, that we may characterise proper names as pure indexicals with a strict character insensitive of the speaker's intentions.

Dolf Rami's (2014) theory was aimed at improving these flaws of Pelczar and Rainsbury's theory, which it does by tying referents identified in context with a particular occurrence of a proper name within an utterance and listing three principles of identification of the referent, which replace the dubbing-in-force or a naming convention. Rami presents his idea of establishing the reference of a contextually sensitive proper name in the following manner:

(Indexicalism 2) "$[[N_x]]_{c, <w, t>}$ is the object that is identified *demonstratively, descriptively or parasitically* in c_w in respect to the occurrence *x* of 'N' by c_a and that is a bearer of 'N' at c_t" (Rami 2014: 139).

According to Rami, demonstrative, descriptive, or parasitic identifications are ways of determining the referent by the speaker in a given context. In his characterisation, it seems clear that proper names are no more conceived as pure indexicals as in Recanati's or Pelczar and Rainsbury's works, but as a class of complex demonstratives. Three types of identification listed are mechanisms available to the speaker to single out his desired reference: demonstrative identification concerns cases of the direct presence of the named object, while descriptive and parasitic are indirect forms of unique identification. The speaker may do so by using a definite description or an intention "to use the name 'N' in the same way as [...] a certain person or a certain group of people" (Rami 2014: 127).

The second condition visible in (Indexicalism 2) is a condition according to which the intended referent must be a bearer of the name at the time of utterance. This comes from the classic analysis of complex demonstratives of the form "that F", which takes the predicate 'F' to be present in the truth conditions of the sentence or reference conditions for the complex demonstrative (such analysis was

provided, for example, by Richard 1993; Borg 2000). However, these theories are unable to explain a class of examples of successful reference without the referent being the bearer of a specific name as witnessed in the following exchange:

(6) A: John has been acting weird lately.
B: That's true, especially since he changed his name to 'Clark'.

As one may see, under Rami's characteristics, A's (and hence B's) statement cannot be true—Clark is not called 'John' anymore at c_t, and therefore A has failed to refer to him. B, however, recognised A's intention and asserted his words, although he considered Clark not to be the bearer of 'John' anymore.[7, 8]

3. Proper Names as Hybrid Demonstratives

Responding to the problems of these theories, I believe that a promising analysis of proper names may be given by treating them as a class of simple demonstratives analogous to 'this' or 'that'. As mentioned in the discussion of conversation (6), it allows us to avoid problems connected with the analysis of the semantics of complex demonstratives mentioned as well as to keep the directly referential character of names.

In my characteristics, I want to use the "hybrid approach" towards indexicals. This notion has been coined by Wolfgang Künne (1992) in his interpretation of a passage from Frege and further remarks by Wittgenstein and Schlick. Unlike the classic Kaplan-style approach, the hybrid approach treats utterances of indexicals as hybrid entities composed of the utterance and its referent (Künne 1992; Predelli 2006) or a pointing gesture (Kripke 2008; Künne 2010; Penco 2013; Textor 2015). The idea is roughly that in the case of utterances of indexicals we should extend our notion of being a linguistic entity to include also extra-linguistic factors such as features of context. Instead of analysing in a Kaplanian fashion sentences containing indexicals by pairing sentence types with respective contexts of utterances, we analyse particular utterances with contextual parameters being present as parts of the utterances of indexicals. The indexicality is, in this view, captured by the phenomenon of introducing extra-linguistic objects or acts as parts of utterances instead of the change of content in differing contexts. Treating proper names as indexicals would mean in the hybrid approach representing their different utterances as pairs of their tokens and referents ("utterance-referent" view) or demonstrations ("utterance-demonstration" view). On such view, the same name-type "John" might be used to refer to $John_1$ and $John_2$, and this fact is captured by

[7] A similar point is raised by Ridley (2016), who considers an example of people calling St. Petersburg by its Soviet name 'Leningrad'. Both Pelczar and Rainsbury and Rami agree that proper uses of names that accord with an outdated convention are impossible, either because of the "synchronic invariability" (Pelczar & Rainsbury 1998: 297) of proper names or the perceived oddity of statements such as "Leningrad has over 2 million inhabitants in 2022" (Rami 2014: 143). If one wants to revise Rami's view (as suggested by an anonymous reviewer) to account for using such outdated conventions, one might further consider examples of similar conversations which would involve cases of identity theft, where the appropriate naming convention was never created.

[8] One may also see that the similar line of argument may be pointed against classic predicativist theories of proper name reference, if one (as, e.g., Tyler Burge [1973]) regards statements such as the one made by A to contain an unarticulated article 'the' or 'that' ("[That] John has been acting weird lately") and maintains that the proper name is taken as a predicate.

representing the two referring utterances of "John" as pairs of its token and either $John_1$ or $John_2$ themselves, or uniquely referring demonstrations of them.

I will opt for the "utterance-referent" view of hybrid proper names in my analysis for two reasons. Firstly, most uses of proper names lack any associated pointing gesture; this would be the case only in situations where (to borrow Rami's phrase) demonstrative identification is possible. The concept of a "demonstration" would have to be highly stretched and artificial to fit the situation in which, for example, I want to refer to an Egyptian pharaoh by uttering the name 'Ramses'. Secondly, as noted by Predelli (2006), this view does not seem to secure the rigidity of proper names, since the same entity composed of an utterance and a pointing gesture may have a different referent in some other possible world due to some other object being pointed.

If we rule out this option, then it is natural to opt for an intention-based view of demonstrative reference. This view is based on the concept of the referential intention of the speaker (an intention to refer to a particular object *n*). Under intentionalism, we may accommodate Rami's remarks about the semantic importance of ways of identifying an object by the speaker in establishing reference. Such accounts originate in Kaplan's (1989b) remarks about the demonstrative reference in "Afterthoughts" and were defended, for example, by Bach (1992) and Åkerman (2009). For simplicity's sake, we may simply assume that the object that is a component of the hybrid proper name is picked out by the referential intention of the speaker, which may be driven by the demonstrative, descriptive, or parasitic act of identification in Rami's sense.

In my view, therefore, the utterance of a proper name N performed by the speaker with an intention to refer to *n* should be formalised as:

$<)N(<l, t>, n>,$

where "$)N(<l, t>$" should be read as "the token of an expression 'N' uttered at place *l*, at time *t*" (which is a use of "token quotation" invented by Reichenbach [1947] as developed by Ciecierski [2020]). The idea behind such formalism is that whenever a speaker refers to a particular object by using a proper name, one uses a specific proper name-type 'N' to reveal her referential intention ("to refer to n") to the hearer in a given communication context.

As noted by Stefano Predelli (2006), a hybrid approach to indexicals also allows for better treatment of elliptic reference. Consider the following exchange:

(7) A: I am a great fan of Truman Capote.
B: I am not [a great fan of Truman Capote].

Suppose that B actually does not know anything about Truman Capote or that he mistakenly believes that the name in question actually refers to Charles Dickens. Under Kaplan-inspired semantics, both of these statements need to be assessed under their respective context of utterance:

(7') A: $<$["I am a great fan of Truman Capote"], $<a_1, t_1, l_1, @, d_1>>$.
B: $<$["I am not [a great fan of Truman Capote]"] $<a_2, t_2, l_2, @, d_2>>$.

And as we may see, the referent of the tokens of "Truman Capote" may differ if, for example, this conversation is carried out by two different agents in different settings and with different dubbings or conventions in force. The hybrid approach allows to analyse this conversation in a uniform way, which secures the reference of the proper name across contexts:

(7") A: [<)I(<l_1, t_1>, a_1>] [am a great fan of] [<)Truman Capote(<l_3, t_3>, n].
B: [<)I(<l_2, t_2>, a_2>] [am not] [[a great fan of] [<)Truman Capote(<l_3, t_3>, n>]].

As we may see in the same manner, we may escape the trouble of pairing tokens and dubbings seen in Pelczar and Rainsbury's analysis of (1). In the hybrid approach, (1) ought to be represented as:

(1") [If] [<)John(<l_1, t_1>, j_1>] [would quiet down], [<)John(<l_2, t_2>, j_2>] [could hear what] [<)John(<l_3, t_3>, j_3>] [is saying],

where the respective referents of tokens of the name 'John'—j_1, j_2, j_3—are picked by the referential intentions of the speaker of (1) formed with the demonstrative act of identification. The appropriate matching of name tokens and their referents is another reason to support the hybrid treatment of demonstratives instead of the traditional Kaplanian account. The formal problem of pairing demonstratives within a sentence with respective intentions/gestures present in the context was raised, for example, by Braun (1996).

Contrary to Rami's theory, under the provided analysis, the referential success in uttering (1) does not require the referent to be the bearer of the name 'John' in the present context. This, of course, may raise some issues (which I will address in the last part of my paper), but let us now focus on the two benefits it brings. Firstly, this allows us to counter the objection sketched at the end of the previous section: the cases of referential success when the referred object is not (currently) the bearer of the name used. What is profound about this is that this objection applies not only to Rami's theory, but also to many contemporary theories of proper name reference, such as causal-chain or predicativist approaches. If we are inclined to say that A successfully referred to Clark in (6), then we should count this as a massive benefit of the simple-demonstrative treatment.

Secondly, the hybrid-demonstrative view escapes the objection raised, for example, by García-Carpintero (2018: 1150), who argues that by using the notion of "being the bearer of 'N'", Rami presupposes the functioning of proper names in their non-indexical form.[9] In the view defended here, there is no theoretical motivation for introducing them. Naming conventions may be seen as chains of uses of demonstratives uttered with a parasitic intention essentially dependent on the reference of some demonstrative (e.g., a speech act of Kripkean baptism) or descriptive use, as I will explain further. Therefore, the hybrid-demonstrative view escapes the threat of theoretical redundancy that may haunt other indexical theories.

4. Possible Objections to the Demonstrative Theory

Both the hypothesis that proper names are simple demonstratives and the use of a hybrid approach may raise some objections that prevented many philosophers of language from looking at proper names this way before. In the last sections of my paper, I would like to deal with the ones I take to be the most widespread: the "Humpty-Dumpty" objection against strong intentionalist accounts of demonstrative reference and Manuel García-Carpintero's objection against indexical theories from the linguistic competence of ordinary language users.

[9] García-Carpintero's universal argument against all indexical theories is discussed in the subsequent section (4.2).

4.1 Humpty-Dumpty Objection

Historically, the Humpty-Dumpty objection dates back to Alfred MacKay's (1968) critique of Donnellan's distinction between referential and attributive uses of description and has since been adapted against intentionalism regarding the reference of simple demonstratives (e.g., by Corazza et al. [2002] and Gorvett [2005]). This objection is based on the statement that, considering the intention of reference as a sufficient condition for reference by a given expression (description, expression indicating whether, in our case, a proper name), we are forced to consider that the speaker can quite freely refer to any object by using any linguistic expression, even if it would break the linguistic convention. A similar position is most often summarised by the words of the hero of *Through the Looking-Glass* by Lewis Carroll, Humpty Dumpty: "If I use a word [...] it means just what I choose it to mean—neither more nor less".

This requires some explanation. Surely, the intentions of the kind represented by the character of a Lewis Carroll novel seem somewhat unnatural. Can we, in the normal sense of the word, intend to refer to a certain object (say, John) by any proper name that is not related to it in any way (e.g., the name "World War II")? A classic analysis of the concept of intention would force us to ask if in such a case the subject can be convinced that he will be able to refer to John by employing the name "World War II". A supporter of the Humpty-Dumpty objection might rightly point out here that if the only condition for a valid reference is to have such an intention, then such a belief would not be surprising. However, one should distinguish here, after Bach (1992) and Åkerman (2009), two kinds of intentions present in communication. The intention to refer is one thing, and the intention to communicate certain content to your audience is something else. The agent could therefore be convinced that he would be able to refer to John by employing the name "World War II", but the conviction that by using this name he could convey the content of his referential intention to other participants in the act of communication would be completely irrational. If a subject considers that conveying this content by using this name is impossible due to the applicable conventions, then they cannot create a communicative intention of the given type, as the subject cannot have the intention of doing something that he considers impossible to do. Just as I cannot, correctly speaking, have the intention of flowing out of the window of the Institute of Philosophy on a broom (since I consider it physically impossible), nor can I have the intention of communicating my wish to refer to Saul Kripke by using the name "World War II" to my audience[10]. Since the speaker usually wants his referential intention to be correctly interpreted, he will employ means understood by others.

According to the proposed view, the relationship between employing a certain naming convention and referential success is not necessary, but rather it is pragmatically necessary for the speaker to use the appropriate convention to communicate their referential intention to the hearer. It is useful to think about the question of what constitutes the propriety of a certain naming convention. Such conventions

[10] Unless I introduce others to my bizarre way of speaking, e.g. by saying "Recognizing historical importance and impact of the author of "Naming and Necessity", I will refer to him during today's lecture as 'World War II'". Then it would not be odd to say that I actually referred to Saul Kripke with the name "World War II", since my intention was easily communicated to the audience.

can, but need not be introduced in any formal way, such as administrative decision or baptism. Consider the following example introduced by Paul Ziff:[11]

> A hungry long white whiskered kitten wanders into my garden. I say to one in the family 'Let's feed that thing'. We do so. The next day I ask 'Where's the cat?' and my son replies 'Whiskers is in the kitchen': no baptismal act no act of ostension and yet the cat has acquired the name 'Whiskers' (1977: 321).

The convention of calling the cat 'Whiskers' may spring into existence even without any particular dubbing. What matters is that, from now on, this name would be applied to the cat and the intention to refer to it would be understood by any other member of the community. What matters in constituting the convention of naming *n* with the proper name 'N' is therefore a constant practice of successfully conveying the intention to refer to *n* to the hearers with the use of 'N'.

However, there are certainly cases where the referential intention is properly formed and the speaker misuses the appropriate naming convention. A similar problem is known in the literature as "the mismatch cases". It is useful to distinguish between two possible variants of such cases. In one of these, the use of a name occurs when the subject is directly familiar with the object that he wishes to refer by a proper name: (a) is mistaken in saying the appropriate name, or (b) has misconceptions about the adopted naming convention. Case (a) is well captured by the following passage from David Kaplan:

> I have witnessed and been the subject of such experiences on many occasions concerning proper names. 'Wait a minute,' says the person, 'Did I just say "Eleanor?" I meant "Harriett"'. Some dark force reached the speaker's psyche and misdirected the hand of intention (Kaplan 1990: 105, note 11).

However, while Kaplan, as a supporter of the causal-historical theory, must accept that in a similar case we are dealing with a very peculiar way of saying the name 'Harriett', a supporter of a demonstrative analysis of proper names does not have to resort to it. In the above case, the intuition that the speaker nevertheless said something about Harriett is explained by his referential intention, while the misunderstanding by his failure to communicate this intention to the audience. We can assume that Kaplan's utterance of the proper name had the form <)*Eleanor*(<1, t>, *h*> (since he referred to Harriett by using the token of the name "Eleanor"), although his intended form was, of course, <)*Harriett*(<1, t>, *h*>.

Cases of type (b) are more subtle. Imagine a group of unrepentant supporters of the Communist Party who did not accept that in 1991 their hometown's name was changed from 'Leningrad' to 'St. Petersburg'. Such people, when communicating with others, keep using a name that has ceased to function in the official or colloquial nomenclature. Should we say that when using the name 'Leningrad' they do not refer to St. Petersburg or merely that they have misconceptions about the functioning of certain names in the language? In my opinion, intuitively, one should admit the second option is right, in line with the intuition represented in the discussion of (6). Regardless of the naming convention, these people talk

[11] Note that such an example is also problematic for the account of Pelczar and Rainsbury, for if we agree that in saying "Whiskers is in the kitchen" Ziff's son actually referred to the cat, he did so without any particular act of dubbing the cat 'Whiskers' even existing.

about a certain city, although perhaps in an outdated and ideological way. This is a pragmatic error at best, not a semantic one.

The same is true when considering cases similar to the famous "Madagascar argument" by Gareth Evans (Evans 1973). Evans points to cases in which the practice of referring to a certain object (Madagascar Island) differs significantly from the reference intended in the original act of naming (according to historical sources, the fragment of the African land on the territory of present-day Somalia was originally "christened" with the name 'Madagascar'), which is supposed to indicate the erroneous predictions of Kripke's causal-historical theory.

To explain this phenomenon, Evans appeals to the notion of a community-based naming practice (1973: 202). But what constitutes the "practice" adopted by the community? From what point do the single uses of the name 'Madagascar' used to refer to an island constitute a practice that allows language users to use it in a "new sense"? The answer suggested by the demonstrative name analysis is: from the moment when some directly demonstrative uses have been mediated by other users who are not directly familiar with the object in question. We can imagine such first uses in line with this practice: sailors who come to Europe use the name 'Madagascar' to denote where they came from or sea traders pointing to an island off the coast of Africa on a map stating, "The next frigate will go to Madagascar to buy pepper and ivory". Such sailors were *referring to the island*, despite the fact that the widespread convention at the time took the name 'Madagascar' to refer to the mainland. If we disagree with this fact, we do not have any easy explanation of why certain naming convention or practice came to exist, and why suddenly the name shifted referents. It is certainly a piece of philosophical fiction to think that we may identify one original mistake in reference which should be designated as a *dubbing* of the island with the name 'Madagascar'.

The situation may be quite different if the subject does not have direct knowledge of the reference of the name. Such uses are widespread: we often use names whose reference we are briefly acquainted with, unable to specify a bundle of features distinguishing a given object. We can define similar reference cases as parasitic, in line with the terminology proposed by Rami. We are then inclined to use the name to refer to the same object that someone else referred to by using that name. To compare this to other demonstratives, you can refer to the class of intentions specified by Kent Bach (1992) in a discussion with Marga Reimer—intentions to refer to the object you point to. Then the subject does not have a specific, unambiguously identifying method of indicating the reference other than referring to a certain naming practice, borrowed from another person.

For example, I may be wrong in thinking that the world record holder in the sprint is Olusoji Fasuba, although, in fact, the record holder is Usain Bolt. Even so, when I say "Olusoji Fasuba is the fastest sprinter in the world", I say something false, because my referential intention in this case is to refer to the referent of the name ("<)*Olusoji Fasuba*(<l, t>, *f*')) used by someone from whom I borrowed it. If I intend to state that Usain Bolt is the fastest sprinter and accidentally use the name "Olusoji Fasuba", then we are dealing with a similar case as Kaplan's slips-of-tongue, and this name should be represented, due to the intention accompanying the statement, as "<)*Olusoji Fasuba*(<l, t>, *u*>".

The given theory, at least at first glance, seems to respond well to cases of mismatch by distinguishing between referential and communicative intentions. It also does not appear to be susceptible to the devastating consequences of the

Humpty-Dumpty argument, and even if they were derivable, one could, by pointing to the cases of 'Madagascar' or 'Leningrad', answer with the slogan popular among computer scientists: *It's not a bug; it's a feature!*

4.2 Linguistic Competence Objection and Predicative Uses

Another interesting remark aimed more generally against indexical theories of proper name reference was recently made by Manuel García-Carpintero (2018). The argument, as presented, aims to show that indexical theories fail to account for our intuitive conditions of linguistic competence for using proper names, which presents no problem for other theories of proper name semantics, especially predicativism. Manuel García-Carpintero, in his article "The Mill-Frege Theory of Proper Names" (2018), quotes the following passage from Schoubye:

> Another reason to think that there is a difference between the context-sensitivity of pronouns and lexical ambiguity is this: it would be entirely unsurprising and unremarkable if some speaker *S* was competent with only one meaning of a lexically ambiguous word. For example, *S* might know that 'bat' is used to talk about a piece of sporting equipment, but not know that it is also used to talk about nocturnal animals. In contrast, it would be surprising and quite remarkable if, say, the word 'she' was part of *S*'s vocabulary (i.e. suppose *S* used the word to refer to his mother), yet *S* was unaware that it can be used to refer to different individuals. Indeed, one might think that *S*, in this case, is just not competent with the meaning of 'she'. In contrast, in the previous case, it does seem that *S* is competent with one of the meanings of 'bat' (Schoubye 2017: 731-32).

Then García-Carpintero himself writes:

> Similarly, someone who assumed that names, like social security numbers, have only one bearer would not thereby be incompetent in their use. For commoners, these differences should just be answerable to contingent beliefs about how widespread is the extension of the predicative conditions with which the expressions are associated (García-Carpintero 2018: 1143).

García-Carpintero's argument therefore states that the indexical answer to the problem of multiple references does not reflect well the nature of the linguistic competence of using a proper name. A child who learns proper names and assigns each person a single, uniquely identifying name is not an incompetent user of it only because he has not yet met another person who would have the same name as, for example, his family members. If we admit, following Kaplan, that to grasp the meaning of a certain expression one must know its character, and such knowledge is a necessary condition for considering a person to be a competent user of the expression, then it turns out that our theory requires too much from average users of names.

I believe that García-Carpintero's argument is incorrect, but in order to capture the error, attention should be paid to a very important issue in the context of our considerations. To fully accept this argument, the concept of *linguistic competence* used by Schoubye and García-Carpintero should be clearly formulated. This, however, seems difficult. A "competent language user" is as real as, in the words of Daniel Dennett, "the Equator or an average Canadian"—it is merely an idealisation, useful only in some cases. The only way to make it useful is to define

what *de facto* competence we are dealing with. Is being a competent user of 'N' in Schoubye's sense necessary for such a user to refer with 'N'? This seems at least not obvious. *S* from Schoubye's story seems to refer to his mother using the word 'she', although he is not a competent user of 'she' in the sense in which he is unable to refer to other persons with the use of this pronoun. As mentioned by García-Carpintero, the mistake of *S* seems to lie in a *metalinguistic* belief he holds—his belief that *the pronoun 'she'* can be applied solely to one person. One could of course disagree with García-Carpintero on this point while still maintaining his argument[12]—the competent use of a pronoun could be understood then as a disposition to refer to certain people with the word 'she', without invoking any beliefs about its extension. But then it seems that "someone who assumed that names, like social security numbers, have only one bearer" is also incompetent. If someone is unable to use the same name-type "John" to refer both to $John_1$ and $John_2$ in their presence and is disposed to call only one of them by that name, might be considered similarly incompetent as *S*. To run García-Carpintero's argument one necessarily needs to claim that *assuming* that indexicals do not refer to more than one person is odd, while assuming the same about proper names is not. Therefore one needs to invoke mistaken metalinguistic beliefs of the speaker and claim, that such understood competence (that is, accurately representing the character of an expression) is crucial for reference.

Paul Ziff in *Semantic Analysis* (1960) draws attention to an important element of the question regarding the nature of the competence to use a proper name. He writes: "If I say 'are you familiar with *Hsieh Ho's* views on art?' I am speaking English: I am not speaking a combination of English and Chinese" (Ziff 1960: 86); knowledge of proper names is not an element of the competence of any language, as they do not belong to any of them separately.[13] Proper names are not part of the dictionaries of the correct Chinese or English, nor are they (in the traditional sense) translatable from one language to another. Thus, they are primarily a reference tool. To show how this perception of proper names can be related to their interpretation as hybrid demonstratives, a simple thought experiment can be offered. Imagine a language in which demonstratives are completely replaced by gestures. For example, instead of saying "Bring it to me", I say in this language: "Bring me [the gesture pointing to the item]". Such a language is conceivable; the types of gestures we make, in the same sense as in the example of Ziff, do not belong to any language (although perhaps different cultures could have used, for example, different hand positions or complement the statement with an eyebrow movement) and are not, in the traditional sense, its expressions. Could we express with such language all that we are able to express with ours? Sure, some expressions might look strange, but it seems that for simple demonstrative uses, there would not be much of a difference between the two languages.

[12] I wish to thank an anonymous reviewer for bringing this possibility to my attention.

[13] Some (e.g., McKinsey 1984) reject Ziff's argument by saying that, in the given example, by simply employing the name "Hsieh Ho" in an English sentence, we establish this name as an English name. But this way of looking at this issue is, in my opinion, dubious. We tend to think that a necessary condition for being a competent speaker of some language *L* is knowing the complete vocabulary of *L*, which is impossible if we are to count all (that is, an infinity of) proper names as its potential parts. Similarly, a person who does not know some specific proper name (say, "Bertrand Russell") does not seem to particularly lack knowledge of *English*.

We can imagine that the grammarians of such a language would list particular types or forms of demonstrative gestures, just as we list on the basis of sounds the words 'this', 'that', or 'he'. But can everything that can be expressed in our language be directly translated into that imaginary language? We can notice that some difficulties may be caused by conversations similar to these:

A: *He* is a bit aggressive, but *she* is such a cute little monkey.
B: I think you are confused. Both of my kittens are *shes*.[14]

In order to allow for a meaningful formulation of such sentences at all (without even considering whether B's statement is true), B would have to specify two different gestures (corresponding to 'he' and 'she') and assign them to types of some kind of content (indicative of gender) or at least a syntactic property (gender) that can pragmatically imply certain properties of the referred object. Additionally, B would have to be aware that with a certain type of gesture (equivalent to 'she'), he is able to refer to more than one object. This would require the existence of conventional rules linking the use of certain gestures to the information conveyed and that users have a notion of the type of gesture by which they can refer to a certain class of objects.

Is it so, however, that ignorance of these facts prevents competent reference by means of the pointing gesture? Probably not, because communication by means of these gestures is still possible even when users do not have this metalinguistic knowledge. We can therefore distinguish between two levels of linguistic competence that interest us. One of them will be *the competence to refer* with a name, the other *metalinguistic competence*, which is related to the knowledge of the semantic properties of the name as a linguistic object.

The first of them does not require too much from the average language user. According to the picture outlined earlier, referring with the use of a name requires only an appropriate referential intention, whether in a situation of direct demonstration or in a parasitic way, mediated by the practice of other users. The second one requires the user to have the concept of a proper name-type 'N' so that they can assign certain semantic properties to it, such as referring to individual objects or having a character.

Based on this distinction, we can make an observation on how the link between these two competences is treated by different theories of proper names. Predicativism (in line with García-Carpintero's argument) will maintain that both competences are identical; referring to a single object requires selecting a specific (context-relevant) object that is the salient reference of the description "bearer of the name 'N'". Causal-historical theories and classic descriptivist theories must hold that the two competences are completely distinct; if names behave similarly to individual constants or definite descriptions, then there is *de facto* nothing that would link two names with different references other than shape similarity. Meanwhile, the indexical theory (as presented here) postulates that one of the competences may ground the other. If we can refer to a specific object with a proper name, we learn at the same time that with a proper name we are able to refer to

[14] Example from Schoubye (2017: 29). An interesting analysis of descriptive and plural uses of both indexical expressions and proper names (e.g., in the proverbs: "Never do *tomorrow* what you can do *today*" or "What *Johnny* will not learn, *John* will not be able to do") can be found in Kijania-Placek (2016). I believe that the method of dealing with the cases proposed in this article may also be applied if proper names are taken to be demonstratives; therefore, readers interested in explaining this phenomenon are referred to the above text.

certain objects with communicational success. The metalinguistic competence is acquired by observing that the proper name 'N' can be effectively used by us to refer to many different objects, just as observing that with the use of 'she' *S* may refer to other people than his mother. Therefore, on the basis of the demonstrative theory, we can indicate a probable mechanism linking these two competences: if I am able as a user to refer to a specific object with the proper name 'N', then I can probably also describe this process in an abstract way ("referring with 'N' to an object *o*") and thus formulate general statements similar to Burge's examples.

Note that although the demonstrative theory liberally treats the question of reference (it is possible to refer with the name 'N' to any object that is the subject of the speaker's adequate reference intention), in the case of predicative uses, we deal with the use of the potential of the proper name-type that is delimited by existing conventions derived from uniquely referential uses.

For example, the sentence:

(3) There are several Johns in the room right now.

could be read as:

(3') There are several people in the room right now who, by some convention, can be referred to with the name 'John'.

A similar analysis has a certain advantage over the classic predicative positions when considering family or hereditary names. The statement:

(8) Every Kennedy has sat in Congress for the Democratic Party

would be true, according to the classic predicative reading, if every congressional holder of the surname 'Kennedy' was a Democratic candidate. A similar statement is false, but Louisiana-elected Republican Senator John Kennedy, unrelated to the famous political clan, would certainly not be considered as a counterexample by the speaker of (4), who intended to refer only to the members of the "Kennedy clan". On the basis of our analysis, however, there is nothing to prevent the intended interpretation from being limited from any convention to a specific naming convention *C*, which includes only members of a certain family among potential referents. Then (4) would be read as:

(8') Every person who can be referred to with the name 'Kennedy' under convention *C* has sat in Congress for the Democratic Party

in line with our expectations. Thus, it can be seen that a thorough analysis of the linguistic competences of users allows us to clearly define the competence that constitutes the basis for reference and predicative uses and also allows us to effectively analyse the cases problematic for classic predicativist theories.

5. Conclusion

This article's focus was to present and defend against possible objections to the treatment of proper names as hybrid-demonstrative expressions—expressions that contain extra-linguistic objects as their parts. I modify the standard treatment of hybrid expressions, presented by Künne (1992), Predelli (2006), and others, to adjust it to desirable features of proper name semantics and argue for regarding specific uses of proper names as ordered pairs of the name token and its referent picked by the speaker's referential intention. I argued that alternative indexical theories of proper name reference as put forward by Recanati (1993), Pelczar and

Rainsbury (1998), and Rami (2014) have either technical or pragmatic problems that make them less preferable to the presented view.

My view presupposes two theses: (1) intentionalism regarding the reference of simple demonstratives and (2) indexicalism about proper names. Since both are regarded as controversial among many philosophers of language, I decided to defend my view against two widely held objections to these two theses: the Humpty-Dumpty objection against intentionalism and García-Carpintero's argument against indexicalism from misdescription of linguistic competence. By rebutting these objections, I also demonstrate how the hybrid-demonstrative theory of proper names may account for 'Madagascar'-style (Evans 1973) cases of reference shift and predicative uses of names (Burge 1973).

Indexicalist approaches to proper name semantics had often been disregarded by philosophers as an implausible *ad hoc* solution to the problem of proper name ambiguity. As David Kaplan writes: "[t]hose who suggest that proper names are merely one species of indexical depreciate the power and the mystery of the causal chain theory" (1989a: 563). In a broader philosophical perspective, the aim of this paper may be regarded as a way of finding out whether disregarding this Kaplanian injunction may lead to interesting philosophical consequences. I hope that this paper provides sufficient evidence for a positive say to this hypothesis.

References

Åkerman, J. 2009, "A Plea for Pragmatics", *Synthese*, 170, 155-67.

Borg, E. 2000, "Complex Demonstratives", *Philosophical Studies*, 97, 229-49.

Bach, K. 1992, "Intentions and Demonstrations", *Analysis*, 52, 140-46.

Bach, K. 2002, "Giorgione was So-Called Because of His Name", *Philosophical Perspectives*, 16, 73-103.

Barcan Marcus, R. 1986, "Possibilia and Possible Worlds", *Grazer Philosophische Studien,* 25, 107-33.

Braun, D. 1996, "Demonstratives and Their Linguistic Meanings", *Noûs*, 30, 145-73.

Burge, T. 1973, "Reference and Proper Names", *The Journal of Philosophy*, 70, 425-39.

Ciecierski, T. 2020, "Utterances, Sub-utterances and Token-Reflexivity", *Theoria*, 86, 439-62.

Corazza, E., Fish, W., & Gorvett, J. 2002, "Who is I?", *Philosophical Studies*, 107, 1-21.

Dąmbska, I. 1949, "Z Filozofii Imion Własnych", in Pelc, J. (ed.), *Semiotyka Polska 1894-1969*, Warszawa: PWN, 1971.

Evans, G. 1973, "The Causal Theory of Names", *Proceedings of the Aristotelian Society, Supplementary Volume*, 47, 187-208.

García-Carpintero, M. 2018, "The Mill-Frege Theory of Proper Names", *Mind*, 127, 1107-68.

Gorvett, J. 2005, "Back Through the Looking Glass: On the Relationship between Intentions and Indexicals", *Philosophical Studies*, 124, 295-312.

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

Kaplan, D. 1989b, "Afterthoughts", in Almog J., Perry, J. and Wettstein, H. (eds.), *Themes from Kaplan*, New York: Oxford University Press, 565-614.

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

Kijania-Placek, K. 2016, "Indexicals and Names in Proverbs", *Studies in Logic, Grammar and Rhetoric*, 46, 59-78.

Korta, K. & Perry, J. 2011, *Critical Pragmatics: An Inquiry into Reference and Communication*, Cambridge: Cambridge University Press.

Kripke, S.A. 1972, "Naming and Necessity", in Davidson, D. and Harman, G. (eds.), *Semantics of Natural Language*, Dordrecht: Springer, 253-355.

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

Künne, W. 1992, "Hybrid Proper Names", *Mind*, 101, 721-31.

Künne, W. 2010, "Sense, Reference and Hybridity: Reflections on Kripke's Recent Reading of Frege", *Dialectica*, 64, 529-51.

Larson, R.K., Segal, G.M.A. 1995, *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press.

MacKay, A.F. 1968, "Mr. Donnellan and Humpty Dumpty on Referring", *The Philosophical Review*, 77, 197-202.

Matushansky, O. 2008, "On the Linguistic Complexity of Proper Names", *Linguistics and Philosophy*, 31, 573-627.

McKinsey, M. 1984, "Causality and the Paradox of Names", *Midwest Studies in Philosophy*, 9, 491-515.

Neale, S. 1993, "Term Limits", *Philosophical Perspectives*, 7, 89-123.

Pelczar, M.W., Rainsbury, J.M. 1998, "The Indexical Character of Names", *Synthese*, 114, 293-317.

Pelczar, M.W. 2001, "Names as Tokens and Names as Tools", *Synthese*, 128, 133-55.

Penco, C. 2013, "Indexicals as Demonstratives: On the Debate between Kripke and Künne", *Grazer Philosophische Studien*, 88, 55-71.

Predelli S., 2006, "Hybrid Indexicals and Ellipsis", *Erkenntnis*, 65, 385-403.

Rami, D. 2014, "The Use-Conditional Indexical Conception of Proper Names", *Philosophical Studies*, 168, 119-50.

Recanati, F. 1993, *Direct Reference: From Language to Thought*. Oxford: Basil Blackwell.

Reichenbach, H. 1947, *Elements of Symbolic Logic*, New York: The Macmillan Company.

Richard, M. 1993, "Articulated Terms", *Philosophical Perspectives*, 7, 207-30.

Ridley, P. 2016, *A Pragmatic View of Proper Name Reference*, PhD thesis, King's College London, https://philpapers.org/archive/RIDAPV.pdf (accessed 1.05.2021).

Russell, B. 1950/2005, *Inquiries into Meaning and Truth*, London: George Allen & Unwin.

Schoubye, A.J. 2017, "Type-Ambiguous Names", *Mind*, 126, 715-67.

Textor, M. 2015, "Frege's Theory of Hybrid Proper Names Extended", *Mind*, 124, 823-47.

Ziff, P. 1960, *Semantic Analysis*, New York: Ithaca.

Ziff, P. 1977, "About Proper Names", *Mind*, 86, 319-32.

Opening Up New (and Old) Vistas on the Contextualist-Minimalist Debate

Ernesto Perini-Santos

Universidade Federal de Minas Gerais

Abstract

The border war between semantics and pragmatics has an early version in the dispute between Mates and Cavell. While Mates argues for a strict separation between semantic inferences and mere pragmatic regularities, Cavell argues for a "logic of ordinary language", identifying the commitments following the act of saying something. This answer gives a clue to the contemporary debate between minimalists and contextualists: we may either think that pragmatic inferences are only effective after the proposition is grasped, or think that it is part of the determination of what is said. However, Cavell has also another answer, that seems much less amenable to logical regimentation: we understand what an agent is committed to as we evaluate what is appropriate to do in a given situation. In this latter case, there is no logic to backtrack to the very determination of what is said: the agent is attuned to the complexity of her contexts of action, but it does not lead to a local adjustment of concepts. This is precisely what a semanticist does, according to Predelli: she is attuned to the complexity of human purposes, and represents this contextual sensitivity at points of evaluation of sentences, but does not build this complexity into his semantic system. Predelli is at odds both with minimalists and contextualists in this respect, and we can see the specificity of his position as we construe different maps of the theoretical landscape. A bit of historical perspective opens up new ways to see some current discussions.

Keywords: Contextualism, Semantics, Pragmatics, Truth conditions.

1. Introduction

It is a border war, or maybe a series of border wars, and it was so even before the territory was charted. There is a distinction between what I imply and what I merely indicate by my words, says Mates, that corresponds to the frontier between semantics and pragmatics (Mates 1958: 72). Mates still lacks the vocabulary to talk about that peculiar sort of inference that is authorized by pragmatics, but not by semantics. The lack of a proper vocabulary is also the lack of a systematic theory, that will soon begin to come to light with the introduction of the concept of implicature in Grice's Lectures *Logic and Conversation* (see Soames

Argumenta (2022): 73-84
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.per

2003: Ch. 9), and, in more complicated ways and with a longer pre-history, around the concept of presupposition (see Horn 1989).

With Grice, we have the beginnings of a theory about different sorts of inferences that do not follow from logical principles and can be identified by a small battery of tests. My aim is not to tell this story, but to notice that, from the start, the frontier between semantics and pragmatics points towards the distinction between what is part of the meaning of words and what follows from their use. *Nota bene*: I do not mean that this is the start of the dispute between ordinary language philosophers and those working within a formal language tradition. Mates is an early way of seeing it *as a distinction* between pragmatics and semantics in the wake of the aforementioned dispute. This is not part, for instance, of the important Quine's review of Strawson (Quine 1953).

Cavell seems to take a further step in the direction of Grice in his answer to Mates. He claims that there is a "*logic* of ordinary language" (Cavell 1958: 181; italics in the original) that explains what a speaker "MUST MEAN" (Cavell 1958: 178; capital letters in the original): what follows from what she says, but not from the meanings of her words. Grice's theory is precisely a theory of such commitments. It is probably far-fechted to read a call to a Gricean-like solution in Cavell, especially in the light of more contemporary theories that share the same Austinian and Wittgensteinian influences, and reject Grice's theory. However, we still have to account for what 'logic' means in his argument.

Be that as it may, I am mainly interested in the other, un-Gricean response, that we can also read in Cavell's paper and that cannot be cashed out in terms of a "logic of ordinary language", or so I will argue. In effect, as he identifies what "we *must* do (or say)" (Cavell 1958: 193; again, italics in the original) with what action is appropriate in a given context, Cavell aims at a much more open evaluation of what one does in "infinitely complex contexts" (Cavell 1958: 185; Mates 1958: 168).

There is a theoretical interest in taking a look at this point of history and devise different ways to see a debate that seems locked in a discussion about where to place pragmatic inferences. And, as we open up different vistas on this subject, I will suggest an unexpected proximity between Cavell's and Predelli's take on the debate.

In the paper, I will go back and forth between Mates and Cavell's exchange in 1958 and the contemporary debate. In section 2, I will describe in broad terms the contemporary opposition between minimalism and contextualism in semantics. Cavell's reaction to Mates is the subject of section 3: I will argue that he has two different responses to Mates, and suggest a first approximation with the present debate. In section 4, we will see different ways to map the contemporary debate. In one to these classifications, Predelli's solution appears as an outlier, as I explain in section 5. The vicinity of Cavell's and Predelli's views is the object of section 6. A brief gesture towards a theory that might cover what is outside the reach of semantics is made in section 7.

2. Minimalism vs. Contextualism

Here is one way to describe the frontier between semantics and pragmatics. Semantics deals with inferences authorized by what is encoded in the language, while pragmatics is a theory about inferences that are not grounded in the code, using rather contextually available information.

> What matters is the cognitive procedure associated with each particular instance of language use, whether it constitutes a conventional code or an ad hoc inference (Ariel 2010: 119).

Ariel is concerned mostly with the distinction between grammar and pragmatics. We are not very far, however, from the way the debate between minimalism and contextualism is usually framed in semantics: whether the proposition expressed by a sentence in a context results exclusively from a compositional mechanism, or non-meaning governed processings are required. If the former theory is correct, then pragmatics enters the story only after the proposition expressed is grasped by the hearer; if the latter is the case, the very determination of the proposition involves already pragmatic mechanisms. Using Ariel's terms, the question is whether what is said is determined exclusively by the conventional code, or *ad hoc* inferences are needed, backtracking to the determination of what is said.

The point can be explained using the contrast between primary and secondary pragmatic processes:

> The contextual processes which, like saturation, are (sub-personally) involved in the determination of what is said I call primary pragmatic processes. In contrast, secondary pragmatic processes are ordinary inferential processes taking us from what is said, or rather from the speaker's saying of what is said, to something that (under standard assumptions of rationality and cooperativeness) follows from the fact that the speaker has said what she has said (Recanati 2004: 17).

The divide between minimalism and contextualism concerns primary pragmatic processes, whether they are exclusively meaning-governed mechanisms, or include also the free enrichment of propositions, as Recanati has it. The latter is, of course, also the Relevance theoretic answer: the content of explicatures is "derived in two distinct ways depending on its source, by linguistic decoding or by pragmatic inference" (Carston 2002: 117).

Both sides are much less uniform than this description might suggest. On the minimalist side, for instance, one may take different positions on whether or not to extend the domain of indexicals, so as to have more or less extended meaning-governed context sensitivity. On the contextualist side, there are different views about the stability of the linguistic input to primary pragmatic mechanisms and how it enters in the determination of what is said. Moreover, neo-Griceans dispute the appropriateness of the exclusive division between what depends on the linguistic code and on *ad hoc* inferences. And, as it is often the case, there are disagreements on who deserves the labels of 'minimalist' and 'contextualist'.

3. Cavell's Contextualism

With this very general picture in mind, it is clear that Mates sides with minimalists, and Cavell with contextualists. The logic of ordinary language he calls for should not be read only in post-propositional processings, but is also part of the fixing of what is said—of what one must say. But this is not the only way to think about the frontiers between semantics and pragmatics, or at least not the only way to react to the position delineated by Mates. In the very same issue of

Inquiry, Cavell suggests two ways to respond to Mates. The first answer is that what we say authorizes inferences beyond what is encoded in the language:

> What needs to be argued now is that something does follow from the fact that a term is used in its usual way: it entitles you to (or, using the term, you entitle others) to make certain inferences, draw certain conclusions. (This is part of what you say when you say that you are talking about the logic of ordinary language.) (Cavell 1958: 180-81).

One way to understand what is a "*logic* of ordinary language" points to a set of inferences that are authorized by one's use of words, and not by the meanings of the words used, inferences that may be systematically described—otherwise, we could not speak of a "logic". So understood, we are not very far from distinguishing, by inferential tests, what is encoded in the language and what belongs to a theory of human action. And, as I have already noticed, seen in the light of more contemporary worries, this logic of ordinary language will have both pre- and post-propositional effects.

If this is so, Cavell's remark seems to demand precisely the sort of theory that will soon emerge with Grice. Let me say, again, that I am not suggesting that this is the way to develop Cavell's argument. On the contrary, I do not think that the sort of normativity that he has in mind (what one *must* say) is what regimented in Grice's theory. We may consider, for instance, Travis' critical reaction to Grice from a point of view that has much in common with Cavell's.[1] But if there is a logic to be unearthed here, it will be something in the vicinity of Gricean theories.

However, in the very same paper, there is another way to react to Mates' view that cannot be cashed out in inferential terms:

> A statement of what we *must* do (or say) has a point only in the context (against the background) of knowledge that we are in fact doing (or saying) a thing, but doing (or saying) it […] badly, inappropriately, thoughtlessly, tactlessly, self-defeatingly, etc.; or against the background of knowledge that we are in a certain position or occupy a certain office of station, and are *behaving* or *conducting ourselves* inappropriately, thoughtlessly, self-defeatingly (Cavell 1958: 193).

In order to identify what we must do, we have to look at situations in which one's actions, both linguistic and non-linguistic, receive a negative evaluation. In the words of Austin, "the abnormal will throw light on the normal, will help us to penetrate the blinding veil of ease and obviousness that hides the mechanisms of the natural successful act" (Austin 1979: 180). We are no longer looking for inferential patterns, but for ways to act in "infinitely complex contexts". And this opens up new perspectives on the contextualist/minimalist debate.

4. Different Solutions to Travis's Challenge

Of course, this very early reaction is not in itself a stance on the debate as it has unfolded, say, since the last decade of the last century. However, I think that if

[1] For Travis' assessment of Grice, see Travis 2008. For the proximity of Cavell's and Travis' views, see, e.g., Cavell 1976: 52 and the purposes in the various Travis' examples.

we use different criteria to group together at least some current positions, we make room to something that is, in a certain respect, similar to Cavell's second response.

Minimalists claim that a contextual effect on the proposition expressed by a sentence can only be triggered by the meanings of its terms. This account keeps a strictly compositional determination of the truth-conditions of sentences. For contextualists, on the other hand, the contextual effects on the proposition expressed can be linguistically controlled, as it is the case of sentences with indexical terms, but they may also result from non-meaning governed mechanisms. In the latter case, the proposition is said to be freely enriched. These very general theoretical constraints lead to different treatments of cases in which different occurrences of a sentence, with no (apparent) indexical morpheme, seem to have different truth-conditions.

Let us consider the famous story of the green leaves, due to Travis:

> Pia's Japanese maple is full of russet leaves. Believing that green is the colour of leaves, she paints them. Returning, she reports, 'That's better. The leaves are green now'. She speaks truth. A botanist friend then phones, seeking green leaves for a study of green-leaf chemistry. 'The leaves (on my tree) are green', Pia says. 'You can have those'. But now Pia speaks falsehood (Travis 1997: 89).

Let Pia's first utterance be

(1) The leaves are green,

and the second utterance be

(2) The leaves are green.

1 seems to be true, and 2 false, but both talk about the same state of the world. There are a couple of options open for the minimalist. She may consider that 1 and 2 have different truth-conditions, either in virtue of a hidden indexical component — 'green' is an indexical term (e. g. Szabó 2001) —, or because 'green' has different meanings in each utterance, so that they are not utterances of the same sentence after all. (Kennedy and McNally 2010) The minimalist may also claim that 1 and 2 have the same truth-conditions but convey different implicatures (Sainsbury 2001). Only in the first approach is there a contextual effect triggered by the meaning of a word, but they all deny any sort of free enrichment. Let us call these solutions, respectively, INDEXICAL, AMBIGUITY, and IMPLICATURE.

Contextualists claim that the difference in truth-conditions between 1 and 2 is not explained by a compositional process, but by an adjustment of the concept 'green' to the context of use, in relevance theoretic terms, by the creation of an *ad hoc* concept (see Carston 2002). If this is the case, then the truth-conditions of sentences are not determined exclusively by compositional processes but involve also a conceptual adjustment that is not meaning-governed. This solution may be called FREE ENRICHMENT. There is another way of seeing 1 and 2 as having different truth-values, without any *ad hoc* conceptual adjustment, nor with any of the minimalist strategies: 1 and 2 are evaluated at different circumstances, and that is all it takes to take care of Travis' example (Predelli 2005: 119-69). Let us call this solution CIRCUMSTANCE.

There are different ways to organize the responses to this case. We may put together the solutions that keep the intuitive difference of truth-conditions be-

tween 1 and 2 (INDEXICAL, AMBIGUITY, CIRCUMSTANCE, and FREE ENRICHMENT), in opposition to IMPLICATURE, that claims that they have the same truth-conditions but convey different implicatures. We may also use the principle of compositionality as the criterion: INDEXICAL, AMBIGUITY, and IMPLICATURE keep a strict compositional determination of truth-conditions, as opposed to FREE ENRICHMENT. I am not sure where to place CIRCUMSTANCE in this respect, but I will put it together with minimalists, since there is certainly no conceptual adjustment. But there are other ways to classify these solutions. For INDEXICAL, AMBIGUITY, and IMPLICATURE, every difference in truth-conditions, if any, is explained by a difference in the stable meanings of the terms, and this is rejected by CIRCUMSTANCE and FREE ENRICHMENT—indeed, for Predelli, INDEXICAL and AMBIGUITY are "hopelessly *ad hoc*" (Predelli 2005: 135). But there is still another way to classify these solutions: INDEXICAL, AMBIGUITY, IMPLICATURE, and FREE ENRICHMENT claim that if there is any distinction in truth-conditions, it is due to the different contributions of the terms of the sentence, while for CIRCUMSTANCE the difference lies in the circumstances of evaluation and cannot be traced back to sub-sentential components. Here is a representation of these different ways to classify theories:

I

Different truth-conditions	**Same truth-conditions**
INDEXICAL AMBIGUITY CIRCUMSTANCE FREE ENRICHMENT	IMPLICATURE

II

Difference in truth-conditions (if any) only compositional	**Difference in truth-conditions not only compositional**
INDEXICAL AMBIGUITY IMPLICATURE CIRCUMSTANCE?	FREE ENRICHMENT

III

Difference in truth-conditions (if any) only in stable meanings	**Difference in truth-conditions not only in stable meanings**
INDEXICAL AMBIGUITY IMPLICATURE	FREE ENRICHMENT CIRCUMSTANCE

IV

Difference in truth-conditions (if any) only in the meanings	**Difference in truth-conditions not only in the meanings**
INDEXICAL AMBIGUITY IMPLICATURE FREE ENRICHMENT	CIRCUMSTANCE

There are some lessons to be gleaned from these tables. The first one is that INDEXICAL and AMBIGUITY are always together—and, with the exception of table I, IMPLICATURE also patterns with them. They constitute the minimalist

cluster. FREE ENRICHMENT, unsurprisingly, is often in a different case. CIRCUMSTANCE is less clear. To begin with, I am not sure whether it should be placed with the minimalist cluster in table II, echoing a Predelli's remark concerning the likely dissatisfaction of traditionalists with his approach (Predelli 2005: 150). More importantly, like FREE ENRICHMENT, it does not situate the difference between 1 and 2 (given that they have distinct evaluations) in the stable meanings of the terms. Still more relevant, it is the only solution that does not take the difference in the truth-conditions to be in the meanings of terms, whether in the stable meanings or in some *ad hoc* concept. We might say that the difference is in the global utterance, but this is not very informative: there is an important theoretical reason why CIRCUMSTANCE is an outlier in last table.

5. The Point of Linguistic Productivity

In order to understand both the proximity and the distance between FREE ENRICHMENT and CIRCUMSTANCE, consider the explanation of linguistic productivity. According to minimalism, linguistic creativity is only explained by a compositional mechanism. Without a compositional account, there is no explanation of how we produce and understand sentences we never encountered before (see, e.g., Stanley 2007: 8-9). With a compositional account, on the other hand, there is "no mystery why our understanding of complex linguistic items has an indefinite range" (Borg 2004: 56). It is clear that the productivity of human languages cannot be explained without the principle of compositionality, but it does not follow that the understanding of sentences is determined *exclusively* by a compositional process. Part of the creativity lies, precisely, in knowing how a concept may have different contributions to propositions in different situations. The reason is that

> [we] cannot predict or give a qualitative exact specification of all past and future human interactions with nature […] Hence this kind of multiplicity of senses or uses remains an uneliminable aspect of our use of [words] (Moravcsik 1998: 43).

This reaction seems to lead to FREE ENRICHMENT—after all, it seems that we face a contextual adjustment of the sense of 'green'. But we may also think that what we really need is the contextual adequacy of the utterance of sentences, or even of sub-sentential utterances (Predelli 2011) taken as whole, and not a local adjustment of its components.

How can we agree with Moravcsik's claim without postulating *ad hoc* concepts? Here is Predelli's answer. In a semantic system, utterances are represented as sentence-context pairs, or, in Predelli 2005's terminology, clause-index pairs. An interpretive system should assign to a clause-index pair an intuitive assignment of truth-vales across circumstances or points of evaluation, that is, an intuitive t-distribution:

> The discussion of the relationship between clause–index pairs and utterances is important, because systems—namely, procedures that operate on the former—aim at empirical adequacy; i.e., at consistency with pre-theoretic intuitions pertaining to the latter. What is desired, among other things, is that the interpretive system, when supplied a clause–index pair appropriate to a certain utterance u, gives results suitably related to (at least some among) our intuitive verdicts about u (Predelli 2005: 6).

A formal system is empirically adequate if it has an intuitive t-distribution, in this case, the same clause-index pair being true at the point of evaluation 1, and false at 2, and that is all.[2] And that is why, according to Predelli, we should be satisfied with CIRCUMSTANCE: a semantic theory does not see beyond t-distributions. The theorist should be attuned to variations in purposes leading to 1 being true and 2 false, and choose accordingly a representational system with an intuitive t-distribution. This sensitivity is not itself, however, part of a semantic theory. While IMPLICATURE does not deliver an intuitive t-distribution, the mistake of INDEXICAL and AMBIGUITY, but also of FREE ENRICHMENT, is to construe this sensitivity back into the meanings of the terms:

> if the relativization of truth-value to points is to yield any informative account of the relationship between meaning and truth, the type of information provide may not renegotiate the very meaning of the expression in question (Predelli 2005: 13n).

Incidentally, this refusal of unstable concepts explains why CIRCUMSTANCE is grouped with the minimalist cluster in table II. What is crucial is that a semantic theory aims at an intuitive distribution of truth-values for 1 and 2 (*contra* IMPLICATURE — table I) without the *ad hoc* theoretical moves of INDEXICAL and AMBIGUITY (table III), nor, indeed, FREE ENRICHMENT's ad *hoc* concepts (table IV). The result is that there are variations in truth-evaluations due to factors that are not represented within semantics. More precisely, the different evaluations of 1 and 2 have a proper place in the theory—they are distinct points of evaluation. However, why sentences have this t-distribution, that is, the way they are represented by a semantic system, is part of the pre-theoretical task of choosing the t-distribution—in Carnap's terminology, it is an external question. That is why CIRCUMSTANCE is an outlier. Let us start now our way back to Cavell.

6. Cavell and Predelli's Semanticist View

As we have seen, Cavell has two lines of response to Mates: calling for a yet-to-be-created logic of ordinary language, and suggesting that the sort of commitment that goes beyond what is encoded in the language should be found in a more elusive understanding of what kind of action is appropriate in a given circumstance. There are different ways to deploy a logic of ordinary language,

[2] Predelli presents his theory as a mere development of Kaplan's semantics, which is, of course, the origin of the term 'circumstance'. It is not clear, however, that aiming at an intuitive t-distribution, and only at that, is faithful to everything Kaplan requires of circumstances. For him, circumstances "must include all elements with respect to which there are content operators" (Kaplan 1989: 511n). If we add that circumstances include only such shiftable elements, that is, elements that "can be sufficiently well defined and isolated" (Kaplan 1989: 504) so as to be the object of an intensional operator, Predelli makes a step beyond Kaplan: there are no operators over the purposes of a speech act, which is what explains the distinction between 1 and 2. In other words, two utterances of the same sentence might receive different truth-evaluations in virtue of a distinction in circumstances that is not accounted by a parameter shifted by an intensional operator. Maybe this difference is due to the distinct concerns of Kaplan (indexicals and demonstratives) and Predelli (wider contextual variations), but it is nonetheless a difference that should be acknowledged.

specifying inferences that will explain what is conveyed beyond what is said. These pragmatic inferences will find its way back into the determination of what is said, in different contextualist frameworks.

It is less clear how this latter move could be built from Cavell's second response. For a start, we can act inappropriately in many ways, and it would be useless to try to give a precise description of how this may happen—actions take place in "infinitely complex contexts". We can be more optimistic about a "grammar" of ordinary actions, probably at the cost of a less precise notion of grammar. But another way to react is simply to say that there is no systematic theory to be found here. Cavell's second answer and Predelli's stance converge in claiming that what explains the variation of t-distribution across points of evaluation cannot be systematically represented—it does not belong in a semantic theory, as Predelli has it, nor in some sort of logic of ordinary language, as Cavell sways from his first to his second answer, or, at least, to my reading of his second answer.

For Cavell, what is appropriate or not in a given context comes naturally to us. This also seems to be a good description of Predelli's view. The semanticist is one who understands the "infinitely complex contexts" in which 'green' is used and represents this complexity in an intuitive t-distribution, a sheer intuitive t-distribution, as it were. However, why she has made this choice is not part of the theory.

CIRCUMSTANCE is an outlier in table IV because what motivates the construction of an intuitive t-distribution is not in the meanings of terms, neither encoded in the stable meanings, nor in locally adjusted concepts.

This solution avoids a problem pointed out by Picazo, following Travis' lead, for FREE ENRICHMENT: if a speaker can build *ad hoc* concepts, we would expect them to be stable across at least a subset of contexts of use, and "it should be possible to coin occasion-insensitive words" (Picazo 2020: 235). However, we can reiterate the argument so as to devise pairs of occasions of use in which a putative *ad hoc* concept created, say, for 'green' in 1, would still lead to different verdicts. In Predelli's picture, since we explain the variability only at the level of points of evaluation, and without having to construe it back in mental representations, nor in the meanings of words, there is no pressure to find, at some level, an invariant item. It is also a way to explain the complexity of our use of words without multiplying (*ad hoc*) senses, breaking apart Moravcsik's disjunction ("Hence this kind of multiplicity of senses *or* uses remains an uneliminable aspect of our use of [words]"): multiplicity of uses *without* multiplicity of senses.

Much depends on what one expects of a semantic theory. If we refrain from building metaphysical or cognitive theories on the back of the meanings of words (as Borg would probably agree), we may still have an intuitive t-distribution, by understanding the distribution of truth-evaluations across points of evaluation as responding to our intuitive judgements, instead of mending stable meanings, or refusing intuitive t-distributions (while Borg seems to be caught between IMPLICATURE, AMBIGUITY and INDEXICAL see Borg 2004: 265).

7. Theories of Language and Theories of Action

The fact that context sensitivity is not accounted for in semantics does not mean that it is outside the reach of any theory. The contextual variation between 1 and 2 is inscribed in the "action landscape" (Rietveld and Kiverstein 2014): we

understand the distinction between these utterances as we understand what a botanist does, and what is involved in redecorating a room (supposing that this is Pia's motivation to paint the leaves green). We know what an agent does, as she engages in these activities, and we also know what she is expected to do in each case. I take it that this is what Cavell means, as he talks about the "complementarity of rule and statement":

> When we say how an action is done (how to act), what we say may report or describe the way we in fact do it […] but it may also lay out a way of doing or saying something which is to be followed (Cavell 1958: 184).

The normativity to which agents are sensitive is partially inscribed in their understanding of the action landscape—not only what is expected from the role each agent takes on, but also how they interact with objects, how to start and to end an interaction etc. And it is also the sort of capacity that we, as interpreters, deploy in understanding that 1 is true, and 2 false.

A semanticist building her system is attuned to these different kinds of activity, just as she is attuned to what counts as an appropriate outfit, or as a vehicle, in different situations, and so on, for the different contextualist-flavoured examples. This belongs to a "pre-semantic" phase of the construction of semantics systems (see Predelli and Stojanovic 2008: 69). We may have many things to say about what is appropriate or not in each, it is just not part of semantics. My suggestion is that it is in itself an interesting result, to see the proximity of theories of language and theories of action.

8. Conclusions

I have argued that there are two responses to Mates in Cavell's well known 1958 paper: one answer calls for a logic of ordinary language, and, if there is a logic to be found, it will be something close to Grice's theory. Another answer does not seem to be so optimistic about a systematic theory of what we say and do, as we talk. If we think that there is a logic accounting for the commitments we are bound to have as we talk, we can ask whether this logic is only effective after the fixing of what is said, as Mates and contemporary minimalists want, or whether it is already operative in determining the proposition expressed, as contextualists have it. However, if there is no logic to be found here, as, I think, is suggested by Cavell's second answer, then this exclusive question is no longer available. Seen from Predelli's viewpoint, it means that semantics will not explain context sensitivity, it will only represent it in an intuitive t-distribution.

One might have the impression that, at the end of the day, the proximity of Cavell's second answer and Predelli's theory is rather thin: we know what they *do not want to be* part of a semantic theory. Worst, this proximity may even by a mirage, since it depends on reading two different reactions to Mates in Cavell's paper, an interpretation that might well be rejected. However, I think that it allows us to see that there are different ways to frame the debate between minimalists and contextualists. More precisely, thinking that it is all a matter of where to place pragmatic inferences is unduly restrictive.

While this is certainly not new, given, in particular, Travis' position on these issues, finding a proper theoretical spot to this position, as I have suggested in table IV, may be of some interest. And finding Predelli as an outlier in this

table may also shed a new light on these issues, since he may be a bridge between more formally oriented approaches and "ordinary language" theories. Finally, the connection with Mates and Cavell very early exchanges shows how the theoretical landscape may have been narrowed, in particular with Grice's intervention. A little bit of historical perspective may open new ways to see some current discussions.[3]

References

Ariel, M. 2010, *Defining Pragmatics*, Cambridge: Cambridge University Press.

Austin, J.L. 1979, *Philosophical Papers*, Oxford: Clarendon.

Borg, E. 2004, *Minimal Semantics*, Oxford: Oxford University Press.

Carston, R. 2002, *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.

Cavell, S. 1958, "Must We Mean What We sSay?", *Inquiry*, 1, 172-212.

Clark, H. 1996, *Using Language*, Cambridge: Cambridge University Press.

Horn, L. 1989, *A Natural History of Negation*, Chicago: University of Chicago Press.

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

Kennedy, C. and McNally, L. 2010, "Color, Context, and Compositionality", *Synthese*, 174, 79-98.

Mates, B. 1958, "On the Verification of Statements about Ordinary Language", *Inquiry,* 1, 161-71.

Moravcsik, J. 1998, *Meaning, Creativity, and the Partial Inscrutability of the Human Mind*, Palo Alto: CSLI Publications.

Picazo, C. 2020, "Occasion-Sensitivity and What is Said", in Ciecierski, T. and Grabarczyk, P. (eds.) *The Architecture of Context and Context-Sensitivity*, Dordrecht: Springer, 227-43.

Predelli, S. 2005, *Contexts: Meaning, Truth and the Use of Language*, Oxford: Oxford University Press.

Predelli, S. 2011, "Sub-Sentential Speech and the Traditional View", *Linguistics and Philosophy*, 346, 571-88.

Predelli, S. and Stojanovic, I. 2008, "Semantic Relativism and the Logic of Indexicals", in García-Carpintero, M. and Kölbel, M. (eds.), *Relative Truth*, Oxford: Oxford University Press, 63-80.

Quine, W.V. 1953, "Mr. Strawson on Logical Theory", *Mind*, 62, 433-51.

Recanati, F. 2004, *Literal Meaning*, Cambridge: Cambridge University Press.

Rietveld, E. and Kiverstein, J. 2014, "A Rich Landscape of Affordances", *Ecological Psychology*, 26, 325-52.

Soames, S. 2003, *Philosophical Analysis in the Twentieth Century*, Vol. 2, Princeton: Princeton University Press.

[3] I am very grateful to the *CNPq* for supporting this research, and to the audience at the *2021 Workshop on Context*.

Sainsbury, R.M. 2001, "Two Ways to Smoke a Cigarette", *Ratio*, 14, 386-406.

Stanely, J. 2007, *Language in Context*, Oxford: Oxford University Press.

Szabó, Z. 2001, "Adjective in Context", in Kenesei, I. and Harnish, R.M. (eds.), *Perspectives on Semantics, Pragmatics, and Discourse*, Amsterdam: John Benjamins, 119-46.

Travis, C. 1997, "Pragmatics", in Hale, B. and Wright, C. (eds.), *A Companion to the Philosophy of Language*, Oxford: Blackwell, 87-107.

Believing the Formless?

Giuseppe Varnier and Salvatore Pistoia-Reda

University of Siena

Abstract

In this note, we discuss the analyticity puzzle affecting the logicality of language hypothesis. The analyticity puzzle is the fact that only some analyticities result in ungrammaticality, which seems to conflict with the idea that an inferential device plays a role in determining the set of the possible sentences of the language. The literature includes two solutions to account for this puzzling evidence. According to one of the solutions, the deductive system can access both ungrammatical and grammatical trivialities, though only the latter can be rescued, i.e. made informative, via application of a pragmatic repair strategy, which modulates the meaning of the nonlogical material. It is then argued that syntax only excludes logically trivial (i.e. unsalvageable) structures, and that nonlogically trivial structures may even be used under their trivial readings. Our focus in this note is on a possible implication of this discussion for the analysis of belief ascriptions. In particular, we discuss that occurrences of the formula 'Bel p' are acceptable when p is nonlogically trivial but unacceptable when p is logically trivial. Since the ascribed propositions differ just on a logical dimension, we suggest, against classical discussion, that belief ascriptions are sensitive to logical considerations.

Keywords: Logicality, Logical form in natural languages, Formal pragmatics, Contradiction.

The logicality of language hypothesis is the idea that the language system, i.e. the combinatorial device building structures out of a lexicon, is not merely interfaced with—but actually contains—a deductive inferential device, sometimes referred to as a "natural" logic (cf. Chierchia 2013; Fox and Hackl 2006; Gajewski 2002, 2009). Assuming this perspective, the set of the possible sentences of a language is restricted to structures that, beyond being syntactically acceptable in a standard sense, are logically fruitful, i.e. are not analytic ("say something" in a Tractarian sense; cf. e.g. Frascolla 2017). This idea breaks with traditional generative approaches to the syntax/logic interface, but also with philosophical well-established doctrines on logic and language, including the Husserlian distinction between nonsense and countersense (cf. Husserl 1901) and the Carnapian separation between formation and transformation rules (cf. Carnap 1934; cf. also discussion in Pistoia-Reda 2021).

Argumenta (2022): 85-91
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.var

Standard evidence in favor of the logicality of language hypothesis comes from the unacceptability (i.e. ungrammaticality) of certain analytic structures. An example is given in the familiar contradiction reported below in (1). The standard account for this case (cf. Fintel 1993; cf. also Gajewski 2008) is that it generates a content such that the structure is true if we subtract the complement of the exceptive from the set of students, and false otherwise; however, in light of the left upward monotonicity of the existential quantifier, the content generated is bound to result in a contradiction. Note that the property of left upward monotonicity allows inferences from sets to supersets (e.g. from 'Some philosophy students passed the exam' to 'Some students passed the exam'), but this clearly conflicts with the contribution of the exceptive, since ⟦students-John⟧ ⊂ ⟦students⟧—in other words, if the structure is true on ⟦students-John⟧, it must be true also on ⟦students⟧. Clearly, however, this behavior cannot be generalized, since not all analyticities result in ungrammaticality and incomprehensibility of this kind. Indeed, it is an established fact that in normal contexts contradictions can be used naturally, in conversation and in internal thought, since they are potentially informative and meaningful ("espressive" Gramsci would say in this connection, though on some quite different underlying assumptions). Thus, the acceptability of the example in (2) reveals what seems to be an analyticity puzzle for the logicality of language hypothesis.[1]

(1) *Some students but John attended the meeting

(2) It is raining and it is not raining.

In this note we begin by focusing on this puzzle. In the literature one finds different solutions to the observed acceptability asymmetry. According to the standard solution, the logicality of language hypothesis should be combined with a modularity vision according to which the inferential device does not access word meanings (cf. Chierchia 2013; Gajewski 2002, 2009; cf. also Abrusán 2019 for useful discussion). Since, in difference from ungrammatical analyticities, acceptable ones owe their analytic status specifically to word meanings, one can simply assume that the mechanism assessing grammaticality is blind to the contradiction in (2), while being able to access the contradiction in (1). This solution thus requires assuming representations similar to what Gajewski calls *logical skeletons*, instead of standard logical forms. Logical skeletons are of course connected to logical forms, in that they are derived from logical forms through substituting the nonlogical material with distinct variables belonging to the suitable semantic type. Assuming this perspective, ungrammaticality is predicted only when the logical skeleton, and not just the logical form, is analytic. We report below the logical skeletons, respectively, of the ungrammatical contradiction and of the acceptable case above. Since the contradiction in the first case emerges from a conflict between the quantifier and the exceptive, it is argued that this representation is already sufficient to establish the analytic status of the structure. Things are different, of course, in relation to the second logical skeleton.

(1) Some $P_{1<e,t>}$ but $P_{2<e,t>}$ $P_{3<e,t>}$

(4) It is $P_{1<e,t>}$ and it is not $P_{2<e,t>}$

[1] It is important to keep in mind that, whereas (2) is perceived *prima facie* as a contradiction, though eventually not interpreted as such, (1) is simply not understood, and proof is necessary to even understand that it is contradictory.

An alternative solution has been recently proposed in the literature (cf. Del Pinal 2019, 2021; Sauerland 2014); this solution is pragmatic in spirit and builds on standard discussion in the previous literature concerning modulation processes and contextual enrichment (cf. e.g. Martí 2006; Recanati 2010; cf. also Stanley 2007). Advocates of this solution assume that the logicality of language hypothesis should be combined with a more conservative approach based on modulated logical forms, to avoid the theoretical cost of having to assume an intrinsically natural logic. In this connection, it is important to note that, on the logical skeletons solution, the deductive device included in the language is assumed to be blind to most classical formulas and logical laws, since it does not even access the co-occurrence of content words. To avoid exactly this, the proponents of the alternative solution submit that some analyticities are grammatical because a contextual repair strategy, i.e. *meaning modulation*, modifies the literal meaning of the nonlogical material in the structures, thus apparently preventing an analyticity to be derived in the relevant cases.[2]

To be more precise, the repair strategy is described as the application of a constrained pragmatic *rescale* operator (here 'R^{+}' in symbols) that specializes, i.e. strengthens, the meaning of the nonlogical words, consequently excluding non standard interpretations compatible with the literal meaning of the terms (cf. the definition in (5); cf. e.g. Del Pinal 2019 for more formal details). For instance, assuming this account the acceptable contradiction observed above can be assumed to be associated with the nontrivial representation (i.e. a *modulated logical form*) reported below in (6), which can result in an interpretation such as that reported in (7). In particular, by applying on at least one of the two conjoined predicates, the rescale operator induces a strengthening in the meaning of the relevant terms, consequently making the overall content conceivable and perfectly informative (it should be noted that Del Pinal's version also allows multiple applications of the operator).

(5) $\{x: R^{+}(P)(x)\} \subseteq \{x: (P)(x)\}$

(6) It is raining and it is not R^{+}(raining)

(7) It is raining and it is not e.g. raining heavily.

In conclusion, even if we assume that certain analytic structures are excluded from the language, and that the inferential device interfaced with syntax does not distinguish between the different kinds of analyticity, by adopting the pragmatic solution we can still account for the acceptability of cases such as (2): in these particular cases, the analyticity appears to be avoided, as the rescale operator applies to restore informativity. Crucially, the account is also capable of predicting unacceptable cases; indeed, in his discussion Del Pinal assumes that the repair strategy is crucially limited in its extension, in that it can only apply to nonlogical words. This seems a condition for the intended comprehensibility of such cases. As a consequence, the strategy cannot produce effects on ungrammatical analyticities such as (1), whose analytic status is crucially due to the logical material contained in them, in crucial accordance with our intuitions (for instance,

[2] One may argue, however, that this idea cannot explain the fact that the second structure is perhaps perceived as being, at least *prima facie*, contradictory, so that the modulation cannot be claimed to take place, so to speak, *a fortiori*. On the contrary, the first structure cannot be understood right from the start

strengthening *students* to e. g. *philosophy students* does not restore informativity in that particular case).

The pragmatic solution, we take it, is motivated by important concerns from a philosophical point of view (cf. Pistoia-Reda and San Mauro 2021 for relevant discussion). In addition, it can be shown that this solution can account for the same cases as the logical skeletons solution, while also extending to some other cases (cf. Del Pinal 2021 for more recent discussion). However, recently various authors started to focus on the details of the original version of the repair strategy, and arguments have been submitted to the effect that the repair strategy should be extended so as to include weakening modulations (i.e. '$R^{\square}_{c}$') and applications to variables (cf. respectively Pistoia-Reda and Sauerland 2021 and Chierchia 2021 for relevant proposals). Another interesting line of research is related to an overgeneration problem afflicting the pragmatic solution (cf. e.g. Abrusán 2019 for relevant discussion). In particular, we would like to submit that acceptable contradictions are not merely acceptable, as we discussed above, but they can even be interpreted under their contradictory reading, as shown in (8). In order to appreciate the significance of this case, recall that, assuming the pragmatic solution, there is no structural difference between functional and nonfunctional contradictions. Thus, the prediction can be made that cases such as this, which contain a nonfunctional contradiction, should sound ungrammatical, unless of course it is rescued, against intuitions, via application of the repair strategy. The unacceptability prediction is clearly incorrect.

(8) Mary is so confused, she thinks that it is raining and it is not raining.

An influential and important response one finds in the current literature involves appealing to the notion of *logical triviality* (i.e. a structure is logically trivial if there is no modulation of the nonlogical material that makes the structure informative), and assuming that structures are excluded only if they are logically trivial, not just trivial (cf. Del Pinal 2021; cf. also Chierchia 2013, 2021 for related discussion on logical vs. grammatical triviality). The decisive and essential point is that, in order for a given structure to sound acceptable, it is sufficient that the structure can be rescued (i.e. made informative) *in principle*; it is sufficient, in other terms, that the structure be associated to at least one interpretation which is not trivial. As a consequence of adopting this notion, we can allow for the possibility that sometimes structures are interpreted under their trivial reading, provided that there is an informative interpretation which is however ignored in that particular occasion.

We believe this notion of logical triviality to be extremely rich philosophically; in this note, we would like to conclude by focusing on a possible implication for the analysis of belief ascriptions. The observed acceptability of the discourse in (8) reveals, we take it, that occurrences of the formula 'Bel *p*' can be meaningful (in this particular case, arguably, it reveals that they can even be true) when the embedded structure *p* is meaningless (*qua* contradictory). To a certain extent, this seems to confirm Mellor 1954's traditional argument to the effect that occurrences of the formula 'Bel *p*' are not necessarily meaningless when the embedded structure *p* is meaningless (cf. also Stroll 1955 for related discussion). Mellor's point, quite famously, was intended to emphasize the fact that, as follows from his general views on meaning, belief ascriptions really pertain to psychology, and that,

consequently, considerations concerning the properties of the content being ascribed (e.g. its logical coherence, its epistemological verifiability) are irrelevant for judging whether our use of the device was adequate.

However, based on the foregoing discussion on logicality, in this note we would like to submit that, if we substitute the reported contradictory embedded structure *p* equivalent to (2) with a distinct contradictory embedded structure *q*, we obtain a surprising result, at least in light of Mellor's essentially psychologistic proposal. In particular, we observe that the discourse in (9) is meaningless (*qua* unacceptable) precisely for the fact that it contains the meaningless (*qua* contradictory) embedded structure *q* equivalent to (1).[3] More precisely, the unacceptability of the discourse in (9) demonstrates, we take it, that some occurrences of the formula 'Bel *p*' cannot be meaningful if the embedded structure *p* is meaningless (*qua* contradictory). It is important to recall that the two reported embedded structures, while being both contradictory, still differ along a logically relevant dimension in that, as we discussed above, while (2) is nonlogically trivial, (1) is logically trivial.

(9) *John is totally irrational. He believes that some students but John smoke.

It should be emphasized, as indeed follows from our description above, that Mellor's argument was not intended to exclude that occurrences of 'Bel *p*' can sometimes be meaningless when the embedded structure *p* is meaningless. But the crucial observation for us is that, given his essentially psychologistic stance, in doing so Mellor merely considers meaningless cases in which *p* is just an uninterpretable "form of words" (Mellor 1954, p. 42). We intend our submitted asymmetry to show, instead, that some purely logical features (i.e. the distinction between nonlogically triviality and logically triviality structures) can make the formula 'Bel *p*' meaningless, thus revealing that "logical or epistemological" (Mellor 1954, p. 43) features are crucial components of belief ascriptions. Granted, if one adopts logicality, one is then forced to assume that the embedded structure with the exceptive is but a mere form of words; but the point is precisely that this structure being just a form of words ("strictly unacceptable", as we said above) is a consequence of its logical features.

In conclusion, in this note we considered the *analyticity puzzle* for the logicality of language hypothesis. This is the fact that only some analyticities result in ungrammaticality. The literature includes discussions on two possible solutions to the puzzle. In particular we focused on one of the two solutions, i.e. the pragmatic one, which has been discussed to radically improve the logicality hypothesis from a philosophical point of view. In particular, this solution maintains a conservative stance concerning logical forms, and assumes that acceptable analyticities are due to the application of a pragmatic repair strategy, i.e. *meaning modulation*, to the nonlogical material. Advocates of this solution further assume that acceptable trivialities may even be used under their trivial readings since, in order for analyticities to be grammatical, it is sufficient that the structures be rescuable in principle. Our focus in this note has been on a possible application of the distinction between logically trivial (i.e. unrescuable) and nonlogically trivial (i.e. rescuable) structures. In particular, we showed that occurrences of the formula 'Bel *p*' are acceptable when *p* is nonlogically trivial but unacceptable when *p* is logically trivial. Since the ascribed propositions differ logically in the two

[3] Granted, we are conscious only of unacceptability, not of its being contradictory *per se*.

cases, we suggested, against classical discussion, that belief ascriptions are sensitive to logical considerations.[4]

References

Abrusán, M. 2019, "Semantic Anomaly, Pragmatic Infelicity, and Ungrammaticality", *Annual Review of Linguistics*, 5, 1, 329-51.

Carnap, R. 1934, *Logische Syntax der Sprache*, Wien: Springer.

Chierchia, G. 2013, *Logic in Grammar*, Oxford: Oxford University Press.

Chierchia, G. 2021, "On Being Trivial: Grammar vs. Logic", *The Semantic Conception of Logic: Essays on Consequence, Invariance, and Meaning*, Sagi, G. and Woods, J. (eds.), Cambridge: Cambridge University Press.

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

Del Pinal, G. 2022, "The Logicality of Language: Contextualism vs. Semantic Minimalism", *Mind*, 131, 522, 381-427.

von Fintel, K. 1993, "Exceptive Constructions", *Natural Language Semantics*, 1, 2, 123-48.

Fox, D. and Hackl, M. 2006, "The Universal Density of Measurement", *Linguistics and Philosophy*, 29, 5, 537-86.

Frascolla, P. 2017, "A Note on Saying Nothing and Saying More in the Tractatus", *History and Philosophy of Logic*, 39, 2, 135-39.

Gajewski, J. 2002, "On Analyticity in Natural Language", Manuscript, University of Connecticut.

Gajewski, J. 2008, "NPI *Any* and Connected Exceptive Phrases", *Natural Language Semantics*, 16, 1, 69-110.

Gajewski, J. 2009, "L-Triviality and Grammar", Handout, University of Connecticut.

Gramsci, A. 1975, *Quaderni del carcere*, Gerratana, V. (ed.), Torino: Einaudi.

Husserl, E. 1901, *Logische Untersuchungen*, Halle: Max Niemeyer.

Martí, L. 2006, "Unarticulated Constituents Revisited", *Linguistics and Philosophy*, 29, 2, 135-66.

Mellor, W.W. 1954, "Believing the Meaningless", *Analysis*, 15, 2, 41-43.

Mill, J.S. 1878, *An Examination of Sir William Hamilton's Philosophy and of the Principal Philosophical Questions Discussed in his Writings*, London: Longmans, Green, Reader, and Dyer.

Pistoia-Reda, S. 2021, "Divertissement quasigramsciano su logica e linguaggio", *Rivista Italiana di Filosofia del Linguaggio*, 15, 1, 148-53.

Pistoia-Reda, S. and San Mauro, L. 2021, "On Logicality and Natural Logic", *Natural Language Semantics*, 29, 2, 501-506.

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

[4] We would like to express our gratitude to the anonymous reviewer for insightful comments on a first version of this note. This work received partial support from Marie Sklodowska-Curie Actions (Pistoia-Reda; 2018 Beatriu de Pinós – grant n. 00070).

Recanati, F. 2010, *Truth-Conditional Pragmatics*, Oxford: Oxford University Press.

Sauerland, U. 2014, "Making Fuzzy Logic Work for Language", *Semantics and Philosophy in Europe*, 7.

Stanley, J. 2007, *Language in Context: Selected Essays*, Oxford: Clarendon Press.

Stroll, A. 1955, "Believing the Meaningless: A reply to Mr. Mellor", *Analysis*, 16, 2, 45-48.

Scalar Implicatures and Presupposition of Existence: Strawson-entailment and the Grammatical Theory

Aldair Díaz-Gómez

Universidad Industrial de Santander

Abstract

Two strong contenders for scalar implicature (SI) computation are the pragmatic and the grammatical theories. While the former sustains that context plays a major role, the latter suggests context is required but is lexically and monotonically constrained (Chierchia 2012). In particular, this paper discusses a processing account for SIs that is dependent on the satisfaction of the Strawsonian presupposition of existence, necessary for the realization of the asymmetric entailment pattern among relevant alternatives. This observation complies with the principles of the grammatical view, for it predicts SIs in the presence of contextually empty domains, unlike the pragmatic account whose necessary access to contextual information will cause propositions with empty domains to always return a truth value false, hence blocking any inference. I present online experimental evidence of an acceptability judgment task and recorded response times of both existential and universal sentences containing entities of three kinds: existent (type-a), non-existent but conceivable (type-b), and non-existent and inconceivable (type-c). The data of 25 Colombian Spanish speaking participants were collected using PsychoPy, powered by Pavlovia. The results suggest SIs are computed in types a and b but relatively unsuccessful in type-c. I conclude that the relevant entailment pattern for implicature computation is the Strawson-entailment relation which, combined with the grammatical account, correctly predicts SIs with non-existent but conceivable entities but avoids inference with inconceivable terms.

Keywords: Scalar implicatures, Strawson-entailment, Grammatical theory, Conceivability.

1. Introduction

This paper intends to answer the following research questions: (i) Under which condition is the entailment pattern among alternatives possible? (ii) Can inferences be derived in the presence of contextually empty domains? Regarding (i), I will discuss the idea of a processing account for scalar implicature computation that is dependent on the satisfaction of the Strawsonian presupposition of

Argumenta (2022): 93-107
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.dia

existence, necessary for the relevant asymmetric entailment[1] relation among relevant alternatives which is part of the grammatical theory pursued by Chierchia (2004) and Chierchia et al. (2012). As far as (ii) is concerned, my hypothesis is that SIs are computed in existential sentences featuring empty-conceivable terms but not necessarily computed with empty-inconceivable elements due to a conceivability restraint. First, section 2 introduces the theoretical discussion revolving around the case of scalar implicatures. Then, section 3 presents theoretical instances involving context sensitivity in implicature processing in relation with presupposition satisfaction and entailment patterns. Section 4 outlines the hypothesis of a refined theory of entailment based on Strawson-entailment (von Fintel 1999). Finally, section 5 reports experimental evidence that supports the addressed hypotheses. Section 6 closes this paper with some concluding remarks.

2. Scalar Implicatures

Observe the dialogue in (1):

(1) a. Arnold: Did you invite all of your friends?
b. Emily: I invited some of them.
⇝I did not invite all of them.

Sentence (1b) instantiates a case of scalar implicatures (SIs henceforth), a phenomenon observed when a sentence containing a term that is part of an ordered scale triggers the negation of stronger terms that also belongs to such scale, e.g., ⟨some, many, most, all⟩ (Horn 1972).[2] That negation is not precisely uttered by the speaker, but rather implicated by them, that means that it is the hearer who is expected to infer what the speaker has implicated. That sort of inference is known as SI. For more clarity, Emily's reply contains the term 'some' which is part of the scale shown between angle brackets above; in uttering it, Emily implicates that she did not invite all of her friends.

There are two main approaches that account for implicature computations, the pragmatic and the grammatical ones. Neo-Griceans (Horn 1972; Gazdar 1979; Hirschberg 1985; Russell 2006, and others) are responsible for the fortification of the context-driven pragmatics-based enterprise. Similarly, Chierchia (2004, 2006); Fox (2007); Chierchia et al. (2012) have formally submitted evidence for a grammatical theory; it has inspired additional work like that of Magri (2009, 2017), and it has been strongly supported by linguists such as Crnic et al. (2015) and most recently Del Pinal (2021). Both theories agree that implicatures are triggered via exhaustification as justified by van Rooij and Schulz (2004) of a sentence *against* the set of alternatives (or scales) induced by it, so a theory of alternatives is imperative in both approaches. Nevertheless, they differ in that the strengthened meaning of a sentence is a result of different cognitive systems. In

[1] Fox (2007) refines the notation of the set of excludable alternatives as one that excludes alternatives that, if negated, lead to a contradiction, such that only non-weaker (instead of stronger) alternatives are negated. Hence, entailment relations are not purely logical. Magri (2009) presents further examples for the motivation of this move.

[2] Other examples of so-called Horn-scales are:

⟨sometimes, often, usually, always⟩
⟨or, R, L, and⟩ (Sauerland 2004)
⟨possible, likely, certain⟩
⟨can/may, should/ought to, must⟩

the pragmatic system it is "derived from principles of rational cooperation" (Chemla and Singh 2014) while in the grammatical system a sentence is strengthened thanks to compositionality principles of a given linguistic system.

2.1. Pragmatic Theory

The pragmatic program relies on speech acts analysis; the hearer reasons upon the speaker's intentions to utter a sentence. This process develops freely when participants of a conversation are assumed to be cooperative.[3] When a speaker S utters some x is p (sentence), a hearer H's reasoning infers that S believes that not all x is p (SI) because, as mandated by the maxim of quality, S lacks evidence that all x is p—an alternative statement—and, since S observes the maxim of quantity too, S does not say all x is p because S is entitled to the belief that what they are saying is as informative as required. Saying all x is p would be much of an over-informative statement. As an example, imagine S utters sentence (2a) which is under-informative given the contextual piece of information that all elephants are mammals, hence the strengthened meaning of the sentence in (2d).

(2) a. Sentence: Some elephants are mammals (Bott and Noveck 2004).
b. Alternative: All elephants are mammals.
c. SI: Not all elephants are mammals.
d. Strengthened meaning: Some but not all elephants are mammals.

In cases like (3), for instance, because it functions at the level of speaker's intentions, the pragmatic view dictates that (3b) is not a relevant alternative given contextual information. Plus, since it applies to global contexts, by definition, a SI is not predicted in embedded contexts as shown in the subordinate clause of (3a).

(3) a. If some of Judy's students passed the test, she will be pleased.
b. If all of Judy's students passed the test, she will be pleased.

Therefore, this theory fails to account for certain anomalies present in natural language. One example involves Hurford's constraint[4] (HC) (Hurford 1974); compare sentence (4a), which is infelicitous by virtue of HC, with sentence (4b), where HC does not apply; however, this is a question that cannot be answered in pragmatic terms. Similarly, there are sentences with possible strengthened meanings that emerge from two logically independent alternatives such as sentences in (5). Once again, SI calculation, in this case, remains unaccounted for by the pragmatic enterprise due to the fact that while both (5a) and (5b) can be true independently of each other's truth value in a given situation, (5a) still triggers a SI that involves the negation of (5b).

(4) a. # John ate an apple or a fruit
b. Some of Mary's students got an A or all of them did.

(5) a. Exactly one kid ate some of his cookies.
b. Exactly one kid ate all of his cookies.

The above observation escapes neo-Gricean reasoning granted that the Gricean maxim of quantity does not require that one utter (5b) rather than (5a) "even when both are believed to be true and relevant" (Chierchia et al. 2012: 2325). As

[3] A speaker is said to be cooperative if they observe Grice's four maxims of conversation: quality, quantity, manner, and relation (Grice 1989).
[4] HC: A disjunctive sentence is infelicitous if their disjuncts entail one another.

seemingly expected, these cracks of the pragmatic view can be accounted for by the grammatical view by virtue of an invisible operator that can be parsed through syntactic processes.

2.2. Grammatical Theory

While the pragmatic program appeals to a theory of speech acts that require the hearer to pull off reasoning strategies to cast light on SI derivation based on contextually relevant information, the grammatical theory is motivated by the insertion of a silent operator akin to only. This operator is often written as Exh which denotes Exhaustification.[5] To exhaustify a sentence is to factor in activated scalar alternatives as part of its strengthened meaning. Exh happens at a compositional level of sentence meaning which grants it the power to occur freely at any embedded level as well as globally. Furthermore, it is part of the sentence logical form given that SIs are logical entailment patterns after all. For those reasons, Exh is a grammatical operator that, unlike neo-Gricean reasoning, is able to justify both local and global implicatures. It should be noted, however, that computation of local implicatures is possible in Levinson's lexical approach (2000); in a sentence like "if you ate some of the cookies and no one else ate any, then there must still be some left", 'some' is understood as 'some but not all'; nevertheless, Levinson's account faces problems in deriving indirect implicatures found in sentences like "Mika doesn't like all of Beethoven's symphonies" where Mika clearly likes only some of Beethoven's symphonies (see Sauerland 2012 for this discussion).

For example, the oddness present in (4a) is indeed explained via HC. However, the same predicted oddness in cases like (4b) disappear thanks to Exh because, when inserted in the first disjunct, the entailment relation that leads to infelicity disappear, so the strengthened meaning of (4b) is shown in (6):

(6) Exh(Exh(Some of Mary's students got an A) or all of them did).

The resulting reading is comparable to an exclusive disjunction. The first embedding of Exh yields some (but not all) of Mary's students got an A, and the second allows for a reading like some (but not all) of Mary's students got an A or all of them did, leaving HC out of question since the entailment relation between the two disjuncts will not hold anymore; this move is unknown to the pragmatic program.

Concerning the second crack of the neo-Gricean opponent, applying Exh to (5a) does trigger a reading that implicates the negation of (5b) as shown in (7), again, unavailable to the pragmatic reasoning.

[5] There are good reasons to believe that *Exh* is different from the overt use of *only*. Overt *only* is seen as part of the assertive content of the used sentence under the relevant conditions of the occurrence of the utterance. Compare, for example, the two sentences in (8), (8a) yields an oddness effect due to the inconsistency with the context whereas (8b) feels closer to the strengthened meaning, which includes *some but not all of her students*, and it is not deemed odd presumably because the content that overt *only* presupposes is indirectly asserted instead; it is instilling an immediate revision of the contextual information (Del Pinal 2021).

(8) *Context*: Every year, Sue assigns the same grade to all of her students.
a. # This year, Sue assigned an A to some of her students.
b. This year, Sue assigned an A to only some of her students. It was a peculiar year.

(7) Exactly one kid x ate some of x's cookies, x did not eat all of x's cookies, and for all other kids y, y did not eat any of y's cookies.

In sum, not only does the grammatical theory justify implicatures where the pragmatic theory does too, but it also, certainly, predicts implicatures where the pragmatic program fails to do so.

3. Existential Presupposition

Information that is presupposed is information that is taken for granted. During a conversation, a lot of information is assumed by the participants for the purpose of efficient communication. Presuppositions entered linguistics and philosophy realms as felicity conditions for utterances; they were initially regarded as part of the semantic component of sentences that, when satisfied, a definedness condition is said to be met. One of such presuppositions is the assumption that the domain of entities to which an expression refers must be non-empty in order for the expression to be defined, that presupposition is called existential presupposition (term originally used by P.F. Strawson in 1952). In Fregean tradition, empty names, however, have sense but lack reference; an empty term like Pegasus fails to refer, but it does express a way in which the object is presented, so it has sense, which is to be held accountable for its meaning. Although the term 'presupposition' was not explicitly used by Frege, he acknowledged that an assertion carries along a presupposition that the thing being talked about designates something, at least something capable of having a cognitive representation, so the name Pegasus would be awarded existence by virtue of the mere thought of it.

Nevertheless, in modern analysis of Aristotelian logic, categorical propositions constituted the problem of existential import. Modern logicians assumed universal propositions are not existentially loaded while the particular ones are. This is motivated by the assumption that when asserted, a universal proposition does not imply the existence of members of the subject term as it is understood as a conditional of the type $\forall x\ (Sx \longrightarrow Px)$, whose truth value will always be vacuously true given the falsity of Sx in the presence of empty terms. However, this leads to the unbearable conclusion that the particulars—whose existential import is indeed implied—will be false and will not stand in any relation of entailment w.r.t. the universals.[6] Hence, to say that (9a) is true while (9b) is false is to agree that there is no relation of entailment between these two.

(9) a. All unicorns have a spiraling horn.
b. Some unicorns have a spiraling horn.

In *On Referring* (Strawson 1950), P.F. Strawson defended a theory of truth-valuelessness that alludes to cases where a proposition fails to be defined. In other words, (9a) would lack truth value in case of failure to meet the existential presupposition for the set of unicorns. However, in later work (Strawson 1952), he salvaged this situation by appealing to what he termed "uniquely referring use" of an expression; that is, uttering a sentence is using it significantly and for communicative purposes given the conventions of regular conversation, along with it, a speaker should succeed in conveying meaning to a hearer no matter what the existential status of an expression is in the actual world so long as they use it to

[6] Recall the relation of subalternation that states that the truth of the universals entails the truth of the particulars, but when the particulars are true, the truth of the universal is undetermined, so technically speaking, the truth of (9a) should entail the truth of (9b).

refer to something. Therefore, ultimately, the question of the truth of a proposition becomes available once it has been successfully used with a referring purpose. This referring use is tantamount to the existence presupposition in that, once satisfied, an expression can be said to be true or false. With this argument, Strawson rescues the entailment patterns between universal and particular categorical proposition.

3.1. The Blindness Hypothesis

A formal representation of Exh is shown in (10), where φ is the uttered sentence and ψ an alternative of the set of excludable alternatives of φ (Excl(φ)) that are negated.

(10) $Exh(\varphi) = \varphi \wedge \neg\psi$

It should be noted that φ and ψ must stand in a relation of asymmetric entailment, i.e., ψ entails φ and not the other way around ($\psi \rightarrow \varphi$; $\varphi \nrightarrow \psi$). Suppose, (11) is uttered, then (12) entails (11) but not the other way around.

(11) # Some Italians come from a warm country = φ
$\rightsquigarrow$ $Exh(\varphi) = \varphi \wedge \neg\psi$ = Some but not all Italians come from a warm country.

(12) All Italians come from a warm country = ψ

Magri (2009) argues (11) is odd because, when exhaustified, it generates a SI that conflicts with the piece of information that all Italians have the same origin. He goes on to say that this mechanism of generating implicatures operates in an automatic fashion and it works regardless of common knowledge which he calls Blindness. If it was not blind to common knowledge, then no implicature would arise, and oddness would not be felt. How does Magri back up his Blindness hypothesis? He argues against a notion of entailment given common knowledge and in favor of a logical notion of entailment as the relevant notion for the definition of Exh since "the strengthened meaning can never be a logical contradiction. This, of course, does not exclude the possibility of the strengthened meaning being a contradiction given common knowledge" if that is the case, then such mismatch results in oddness (Magri 2009: 258). This explains the preference for the SI algorithm to prefer a logical notion of entailment over entailment given common knowledge. In effect, if the latter were the preferred one, then the SI computation device would prevent (12) from being a scalar alternative of (11), hence avoiding any contextually contradictory interpretation.

As predictable, one could argue that sentences are not always strengthened; plus, it is not mandatory that alternatives be negated since it is known across the literature that implicatures do not always happen. Under those circumstances, Magri assumes a relevance assignment procedure, which the Blindness scheme also overlooks, that renders the application of Exh mandatory in matrix clauses such relevance procedure encloses the uttered sentence—also known as the prejacent—and any other contextually equivalent sentences. In so doing, all the relevant alternatives will be obligatorily negated.

To summarize, Margri's scheme dictates that SIs emerge blindly to contextual information, although constraint by the lexicon and monotonicity as pointed out by Chierchia (2012), and are the result of mandatory application of Exh that negates excluded alternatives—necessarily assigned relevance—and confronts the prejacent against its alternatives by means of logical characteristics (logical

entailment) rather than contextual force. This process may or may not yield oddness effects depending on whether the result is a contextual contradiction or not.

3.2. Contextually Empty Domains

In his paper, Pistoia-Reda (2017a) argues that establishing the asymmetric entailment pattern in the interpretation of existentially quantified sentences with a domain restrictor contextually known to be empty is indeed plausible. He argues that if such pattern is the relevant relation for SI computation, then it is realized with no need to access contextual information to check whether the relevant domain is empty, contrary to what Schlenker (2012) defends, that however the computation device works, it cannot be blind to the piece of information that domains are non-empty. His main empirical evidence relies on the interpretation and appropriateness of this pair of sentences:

(13) # Some Swedes come from a cold country.

(14) # Some Swedish matadors come from a cold country.

His main intuition is that if an existential sentence with contextually empty sets is odd in virtue of a SI generated, then no need to access is necessary for the realization of the entailment pattern, and this would suggest that universal quantifiers are existentially loaded. As a matter of fact, in a pilot experiment run with native speakers of English, he confronted the results of three types of existential propositions; basic non-empty subject terms with predicates producing contextual contradictions (13), contextually empty subject terms also with predicates leading to contextual contradictions (14), and contextually empty subject terms but this time not yielding contextual contradictions (15).

(15) Some Swedish matadors know Latin.

In the first two cases he noted that sentences like (13) and (14) received similar percentage of inappropriateness judgments that, when compared to (15) this latter did not show a significantly low acceptability. So, if inappropriateness is to be accounted for via SI computation conflicting with contextual information, then the implicature computation arises irrespective of the non-emptiness status of the relevant domain, i.e., no access to context is mandatory; case (15) provides evidence of high appropriateness potential that could be explained via SI generation not conflicting with contextual information. In other words, if the asymmetric entailment is maintained in order to account for mismatching inferences with empty domains, then universal sentences must carry an existence presupposition. If this is so, then this existence presupposition is satisfied without recourse to contextual knowledge when the universal sentence is counted as relevant alternative. Pistoia-Reda (2012) elaborates on a modified version of the relevance assignment procedure for the universal variable featuring non-existing entities, through a careful Meinongian analysis that takes into account the principle of unrestricted freedom of assumption, where predicated properties of entities are said to be possessed by them independently of their existential status. With this modification, entailments derived from predicated properties can explain the oddness in (14), namely that Swedish matadors, while not existing, come from a cold country.

Having this panorama into account, I find it only logical to contribute to this debate with the aim of elucidating the conditions for implicature computations, particularly in the interpretations of contextually empty domains. My predictions

are in line with Pistoia-Reda (2017a); according to the results I present, the SI mechanism overlooks contextual contradictions that result from the negation of relevant alternatives. However, there is certain level of context involved (though constrained) in the realization of the relevant entailment pattern, such involvement of context is the assumption that a presupposition of existence in Strawsonian terms is satisfied. However, failure of this results in a rejected sentence with no defined truth value.

4. Strawson-Entailment in Implicature Computation

Though slightly controversial, Magri's Blindness hypothesis make appropriate pre-dictions on SI behavior, hence the preference of the view that SIs are indeed computed at a compositional level via entailment patterns that allow for the exclusion of alternatives which are negated so that a SI is derived without any needed access to contextual information. Despite that, I propose a revised version that refines the entailment pattern required for the SI generation. I argue that the entailment relation in the Blindness formulation is restricted to the satisfaction of the definedness condition which, in turn, must be met by means of the Strawsonian presupposition of existence; otherwise, no entailment pattern will be realized resulting in SI failure. The main evidence for this is an extended analysis of observations concerning contextually empty terms (Pistoia-Reda 2017a, b; Pistoia-Reda and Sauerland 2021).

Let me take stock, assume scalar implicatures are the result of mandatory *Exh*, 1) alternative members are activated through a relevance assignment procedure, 2) non-weaker alternatives standing in an asymmetric entailment relation with the base form of the uttered sentence (the prejacent) enter the set of excludable alternatives (Fox 2007), 3) excludable alternatives, part of the entailment relation, are negated (SI), 4) if there is a mismatch between the SI just generated and the information contained in the common ground, the outcome will be an odd sentence.

Now, let us focus on the entailment relation mentioned in step 2. Since entailment relations are contingent to presupposition satisfaction, there is a restriction that should be present for the entailment pattern to be computed successfully and prevent the sentences from lacking truth value; it is that the expressions for which the entailment pattern had ensued must be defined following the definedness condition. If the expressions are undefined, they provoke truth-valuelessness; hence, it is expected that no entailment relation arises. The particular case of seemingly empty terms is salvaged by virtue of Strawson-entailment (von Fintel, 1999) for it presupposes existence of entities under the assumption that speakers take for granted that utterances carry truth values and are logical.

However, this treatment is *restrained by conceivability* because committing to the presupposition of existence of an entity used in common conversation is dependent on our epistemic status as well as our ability to conceive of objects, including those that are not contextually known to exist on account of our linguistic knowledge, feature which equates to context retrieval. This move is pivotal for the reconciliation between context and grammar, our SI derivation device is intrinsically endowed with the inherent definedness of the domain of entities, this, of course, does not contradict the Blindness filter, in fact, this reconciliation ratifies that scalar implicatures are not derived because of context but in spite of

context; and it is only under the Blindness hypothesis that oddness is explained via SI, if it were not, access to context would prevent any inferences given that contradictions ought to be avoided.

With that modification to the relevant entailment relation in implicature computation, we are now in a position to explicate possible scalar inferences triggered in quantified propositions containing contextually empty terms, adding to the discussion the plausible explanation founded in the inclusion of a tacit premise that the relevant domain of discourse is assumed to be non-empty. In spite of that, Pistoia-Reda (2017b) predicts presupposition failure in propositions containing inconceivable entities such as *round squares*; even though they also intend to denote empty terms, they fail to be defined. However, in his current analysis, through the same Meinongian modification introduced before (Pistoia-Reda, 2022), he predicts an oddness effect due to relevance assigned to propositions containing empty domains. As I see it, propositions containing *round squares* are not necessarily always undefined, for the likelihood to assign a truth value to it by virtue of interpretation strategies applied by the speaker is not null. A round square could be understood to be a square with round corners, or even a circle inside a square. Though not understood logically, there are different ways speakers can make sense of it at the cost of significantly high cognitive processing time. This latter observation is approached more in detailed in Del Pinal (2021) and Pistoia-Reda and Sauerland (2021) amidst discussions on logicality of language. However, I sustain that definedness is restrained by conceivability which may either block interpretations or take significantly long times for acceptance. On a par with this theoretical postulation, I will present experimental evidence that supports it. Next section presents a sentence reading experiment with acceptability judgments that offers empirical support in the case of Colombian Spanish.

5. Experiment

5.1. Methods

The experiment consisted of a self-pace reading task that elicited acceptability judgements and recorded reaction times carried out with 25 Colombian Spanish-speaking undergraduate students of academic backgrounds different from linguistics, whose ages ranged between 18 and 30 years old. The participants were presented a number of Spanish sentences[7] split into 4 chunks, and they had to rate them following a Likert scale of 1 to 7 (Likert 1932) basing their decisions on their interpretative intuitions. This is a 3 X 2 factorial design. The first factor includes the degree of compatibility of the NPs with the actual world, it involves three levels: existing entities (*type-a*), non-existent but conceivable entities (*type-b*), and non-existent and inconceivable entities (*type-c*). In addition, the second factor holds two inference levels: universal sentences with no inference triggers, and existential sentences with inference triggers (see Table 1 for a clearer picture of each sentence kind. Their English equivalent is shown in (16) respectively).

[7] 16 sentences for each compatibility level (8 for each inference level) for a total of 48 sentences of the critical condition.

Table 1. Experimental conditions

Compatibility	Inference (+/–)
(1) Existent entities	(–) Todas las rosas rojas son flores. (+) Algunas rosas rojas son flores.
(2) Non-existent but conceivable entities	(–) Todos los príncipes colombianos son latinoamericanos. (+) Algunos príncipes colombianos son latinoamericanos.
(3) Non-existent and inconceivable entities	(–) Todos los triángulos de cuatro lados son polígonos. (+) Algunos triángulos de cuatro lados son polígonos.

(16) a. All/Some red roses are flowers.
b. All/Some Colombian princes come from Latin America.
c. All/Some four-sided triangles are polygons.

Since this is a sentence reading task with acceptability judgments and reaction times, the dependent factors will be both on-line and off-line measures. The former are the reaction times before pushing the judgment button, and the latter are the acceptability judgements themselves. Every sentence belonging to the crucial conditions have the form '*some S are P*' and '*all S are P*' as shown in (16). The predicate in the existential one induces a mismatch that allows for two readings of the sentence, a logical one that yields a positive truth value and a contextual one that yields a negative one because of the inference inconsistent with common knowledge. This is due to the fact that when an inference is made, the sentence meaning gets strengthened and this is what clashes with the common ground information. Hence, higher response time with low acceptability in the condition with 'some' (*some*-sentence henceforth) is expected, compared to the condition with 'all' (*all*-sentence henceforth) which carries no inference trigger and whose meaning does not conflict with contextually known information.

5.2. Procedure

The task was run using PsychoPy (Peirce et al. 2019), powered by Pavlovia to be run online; participants were made to use a computer since the task was not available in another device. As soon as they started the exercise, they were warned that they would be presented sentences divided in four chunks which they would read progressively, and soon after, rate on an acceptability scale of 1 to 7. The software recorded reaction times before participants response (RTs) and, of course, participants' acceptability judgments (AJs).

5.3. Results

Fig. 1 summarizes AJs of types a, b and c. It shows the contrast between non-inferential *all*-sentences and inferential *some*-sentences. The results for type-a concur with the predictions, high ratings for non-inferential items against low ratings for inferential ones due to oddness effects. Even though in types b, and c the

difference between inferential and non-inferential sentences was not salient, type-b sentences did receive higher levels of appropriateness.

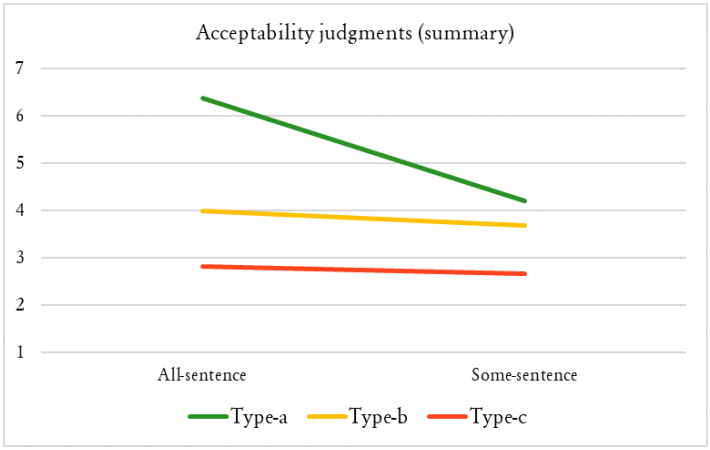


Fig. 1. Acceptability judgments.

Fig. 2, on the other hand, reports higher RTs in the inference conditions for type-a. This is explained by virtue of the cost of SI computation; the moment the sentence was found to be odd, it received low AJs as reported above. For type-b, RTs were higher in the inference condition compared to the non-inference trigger sentences. This points to similar reasons to those of type-a. Finally, for type-c, a closer behavior to type-b is reported, however, the crucial difference is that inferential sentences of this type took less amount to be rated, that is, to be rejected considering the low AJs reported in Fig. 1.

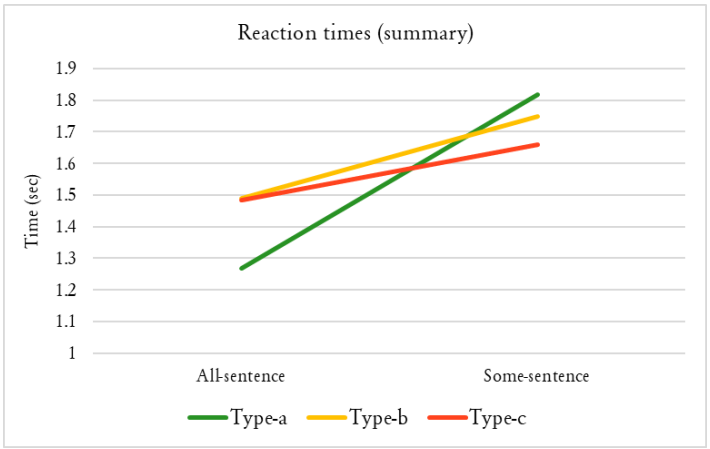


Fig. 2. Reaction times.

5.4. Discussion

Overall, type-a items behave according to previous research conducted by Bott and Noveck (2004); rejections are due to SI computation conflicting with

contextual knowledge which renders the *some*-sentence odd. Besides, a SI is said to occur owing to the higher RTs reported in the inference condition.

Although AJs in type-b sentences were not noticeably different across conditions, the pattern of higher RTs in the presence of inference triggers is indeed observed, which marks SI computation effort that results in a moderately acceptable sentence, certainly less acceptable than non-inferential sentences. However, the fact that type-b sentences are not prominently accepted across conditions does not render them utterly unacceptable. One could argue that there is a possible domain of discourse where 'Colombian princes' exist; this interpretative component might as well be what is preventing the judgment from plummeting in the Likert scale. This AJ difference becomes more relevant when we acknowledge they are rescued from utter unacceptability by their being conceivable, unlike type-c sentences which are lying at the bottom of the scale, so under these circumstances, the *some*-sentence appears odd because of its possible incongruity found at the intersection between the subject and the predicate terms (equivalent to SI), and its appropriateness should not be much higher than the universal sentence; hence, their higher RTs compared to *all*-sentences.

If we look at the RTs of *all*-sentences of types a and b, they are higher in the latter than in the former, and this may be due to a verification process aimed at finding elements of the crucial kind in our epistemic world construal which requires significantly more time than in type-a sentences with no reference failure. Likewise, this process is equivalent to effort for presupposition satisfaction, which is rather restrained in type-c elements but achieved more freely in type-b items given they both have the same RTs for non-inferential conditions, but different ones for the elements with inference triggers where type-b sentences showcase higher AJs. In fact, such low acceptability for type-c *some*-sentences may be due to their being highly hindered by conceivability, namely, the presupposition of existence is hard—if not impossible—to be achieved, leading to definedness failure.

Be that as it may, type-c results raise the following questions: 1) why did participants give, on average, a rating between 2.0-3.0 to these items and not a straight up 1.0? 2) Why can we still perceive a tendency for rejection in the inference sentences?

The most reasonable answer is the availability of interpretation strategies undertaken by the participant who, presumably, attempts to verify the existence of possible elements within the cardinality of the subject term. 1) It may have been the case that a small subset of the participants did figure out a way to make some loose sense of at least some type-c sentences, or every participant found some items of the same condition less infelicitous than others, and based on this, they judged the sentence with the inference trigger a tiny bit less acceptable than the non-inference condition. 2) This move will result in higher cognitive effort evidenced in the reaction times, though not as high as in type-b, most probably due to presuppositions failure which actually leads to rejection.

An interesting recent approximation to this observation appears in Pistoia-Reda and Sauerland (2021) who investigate a pragmatic repair strategy discussed in Del Pinal (2019). In their analysis, Pistoia-Reda and Sauerland envisage the infelicity of sentences belonging to this kind but rescues them via application of a pragmatic operator that modifies and weakens the meaning of the crucial terms. To my view, it could actually be the case that a sentence recovers from infelicity via application of a "silent RESCALE operator" that strengthens interpretation and modulates meaning "via exclusion of logically available interpretations" (Del

Pinal 2019: 4); as a matter of fact, this agrees with the process of linguistic retrieval for meeting the definedness condition, but then again, the results suggest high infelicity effects.

6. Conclusion

In sum, the results of this experiment corroborate the data presented by previous research that dictates higher cognitive effort for conventional sentence interpretation with no referential failure carrying scalar implicatures (type-a). In addition, the critical data illustrate that propositions containing elements with referential failure but which are conceivable (type-b) are judged more felicitous than sentences containing elements that, from a logical perspective, fail to be conceived in the speaker's mind (type-c).

Regarding scalar implicature calculation, in type-b conditions, SIs are triggered and arguably achieved by virtue of successful definedness condition satisfaction in the spirit of Strawson, who assumes commitment to the non-emptiness status of the domain of discourse. A diverse prediction is forecast in the type-c conditions featuring inconceivable entities, for the data offers the interpretation of possible accommodation of the target definedness condition but low likelihood for scalar inference derivation given its infelicity and lower RTs compared to type-b items.

These results are in keeping with the grammatical account in that they suggest strong likelihood for computation of SIs in contextually empty domains, which is at odds with the pragmatic account. To recall, SI computation in contextually empty domains is not predicted by the pragmatic account, for if free access to contextual information is essential, expressions such as "Colombian princess" are said to always return a value false given today's actuality; hence, no inference of any kind is predicted.

Moreover, there is a conceivability restraint, which I alluded to in section 4, that represents a huge obstacle for SI processing in type-c elements since it hinders presupposition satisfaction in logically inconceivable empty terms; therefore, it does not allow for the realization of the Strawson-entailment pattern, that requires the definedness condition for it to be achieved. On the other hand, the possibility of pragmatic repair mechanisms applied by the speaker, via a weakening device applied to non-logical elements in order to make sense of an utterance (Pistoia-Reda and Sauerland 2021), cannot be cancelled at all. At the same time, I suggest a contrast between my results and their theoretical observations. They allow room for the acceptable interpretation of inconceivable entities, while my results suggest otherwise.

In the light of Magri's standard cases, infelicity in type-c is explicated by dint of propositions not meeting the definedness condition, and not through a contextual clash via mandatory SI computation, yet Magri does not deal, to a deeper extent, with inconceivable objects. Nonetheless, my intuition is that this may well be attained but only at the cost of heavy cognitive exercise. After all, natural language understanding is as subjective as it is flexible, but this does not mean that anything can be said that is meaningful.

References

Bott, L. and Noveck, I.A. 2004, "Some Utterances are Underinformative: The Onset and Time Course of Scalar Inferences", *Journal of Memory and Language*, 51, 3, 437-57.

Chemla, E. and Singh, R. 2014, "Remarks on the Experimental Turn in the Study of Scalar Implicature, Part 1", *Language and Linguistics Compass*, 8, 9, 373-86.

Chierchia, G. 2004, "Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface", in Belletti, A. (ed.), *Structures and Beyond*, Oxford: Oxford University Press, 39-103.

Chierchia, G. 2006, "Broaden Your Views: Implicatures of Domain Widening and the 'Logicality' of Language", *Linguistic Inquiry*, 37, 535-90.

Chierchia, G., Fox, D., and Spector, B. 2012, "Scalar Implicature as a Grammatical Phenomenon", in von Heusinger, K., Maienborn, C., and Portner, P. (eds.), *Semantics: An International Handbook of Natural Language Meaning*, vol. 3, Berlin: Mouton de Gruyter, 2297-2331.

Crnic, L., Chemla, E., and Fox, D. 2015, "Scalar Implicatures of Embedded Disjunction", *Natural Language Semantics*, 23, 271-305.

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

Del Pinal, G. 2021, "Oddness, Modularity, and Exhaustification", *Natural Language Semantics*, 29 (1), 115-58.

Fox, D. and Hackl, M. 2006, "The Universal Density of Measurement", *Linguistics and Philosophy*, 29, 537-86.

Frazier, L. 2009, "Computing Scalar Implicatures", in Friedman, T., and Ito, S. (eds.), *SALT XVIII Proceedings*, Cornell University: Linguistic Society of America, 319-39.

Gazdar, G. 1979, *Pragmatics: Implicature, Presupposition and Logical Form*, New York: Academic Press.

Grice, H.P. 1989, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Hirschberg, J.L.B. 1985, *A Theory of Scalar Implicature*, Garland: University of Pennsylvania.

Horn, L.R. 1972, *On the Semantic Properties of Logical Operators in English*, Ann Arbor: University Microfilms.

Hurford, J.R. 1974, "Exclusive or Inclusive Disjunction", *Foundations of Language*, 11, 3, 409-11.

Likert, R. 1932, "A Technique for the Measurement of Attitudes", *Archives of Psychology*, 140, 1-55.

Magri, G. 2009, "A Theory of Individual-Level Predicates Based on Blind Mandatory Scalar Implicatures", *Natural Language Semantics*, 17, 245-97.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J.K. ,2019, "PsychoPy2: Experiments in Behavior Made Easy", *Behavior Research Methods*, 51, 1, 195-203.

Pistoia-Reda, S. 2017a, "Contextual Blindness in Implicature Computation", *Natural Language Semantics*, 25, 2, 109-24.

Pistoia-Reda S. 2017b, "On Conceivability and Existence in Linguistic Interpretation", in Brézillon, P., Turner R., and Penco, C. (eds), *Modeling and Using Context. CONTEXT 2017: Lecture Notes in Computer Science*, Vol. 10257, New York: Springer, 203-12.

Pistoia-Reda, S. 2022, "Relevance Without Existence: Empty Domains in Blind Exhaustification", Università di Siena, unpublished manuscript.

Pistoia-Reda, S. and Sauerland, U. 2021, "Analyticity and Modulation: Broadening the Rescale Perspective on Lan-Guage Logicality", *International Review of Pragmatics*, 13, 1, 1-13.

Russell, B. 2006, "Against Grammatical Computation of Scalar Implicatures", *Journal of Semantics*, 23, 4, 361-82.

Sauerland, U. 2004, "Scalar Implicatures in Complex Sentences", *Linguistics and Philosophy*, 27, 3, 367-91.

Strawson, P.F. 1952, *Introduction to Logical Theory*, London: Routledge.

Van Rooij, R. and Schulz, K. 2004, "Exhaustive Interpretation of Complex Sentences", *Journal of Logic, Language and Information*, 13, 4, 491-519.

von Fintel, K. 1999, "NPI Licensing, Strawson Entailment, and Context Dependency", *Journal of Semantics*, 16, 97-148.

Metaphor Identification beyond Discourse Coherence

Inés Crespo
NYU Paris
Andreas Heise
Institut Jean Nicod (CNRS, EHESS, ENS-PSL)
Claudia Picazo
UNED

Abstract

In this paper, we propose an account of metaphor identification on the basis of contextual coherence. In doing so, we build on previous work by Nicholas Asher and Alex Lascarides that appeals to rhetorical relations in order to explain discourse structure and the constraints on the interpretation of metaphor that follow from it. Applying this general idea to our problem, we will show that rhetorical relations are sometimes insufficient and sometimes inadequate for deciding whether a given utterance is a case of metaphor. They are insufficient, since rhetorical relations fall short at times of providing a basis for disambiguating between literal and metaphorical interpretations. In such cases, contextual information other than previous discourse needs to enter the picture. To this effect, we bring the idea of external consistency into play. Beyond that, though, we will argue that rhetorical relations are sometimes inadequate to account for coherence, if conceived as relations among sentences only. The reason is that extra-linguistic elements of the situation in which the sentence is uttered may be crucial for getting at the preferred interpretation. To account for these cases, we allow rhetorical relations to connect both with previous discourse and with extra-linguistic situations. In our final refinement of the notion of contextual coherence, we forfeit any appeal to rhetorical relations in favour of Questions Under Discussion (QUD). We defend the view that this account does not only explain the same sort of cases. What is more, it solves the issue of metaphor identification in impoverished contexts.

Keywords: Metaphor identification, Context, Coherence, Rhetorical relations, Questions under discussion (QUD).

1. Introduction

What clues can interpreters rely on in deciding whether a sentence uttered in a specific context had better be interpreted in a metaphorical sense, rather than in a literal sense? Our aim here is to address this question about metaphor identification. When

Argumenta (2022): 109-124
ISSN 2465-2334

DOI 10.14275/2465-2334/202200.cre

faced with an utterance of a sentence, one of the things the interpreter might need to determine is whether it is best understood as a metaphor or as a literal claim. This choice is particularly difficult when it comes to sentences that do not involve any category mistake but that are nonetheless reasonably interpreted in a metaphorical sense—so-called twice-true metaphors[1]—as illustrated by Disraeli's utterance of (1):

(1) I have climbed to the top of that greasy pole.

We propose that considerations about coherence play a crucial role. Following Asher and Lascarides, we hold that coherence governs interpretation and that discourse structure constrains metaphorical interpretation.[2] However, we go beyond their approach and focus on how discourse structure often needs to be supplemented with extra-linguistic information available to the interpreter, without being properly construed as part of previous discourse.

We make two claims. First, although discourse structure constrains metaphorical interpretation with regard to previous discourse, as Asher and Lascarides show, we argue that rhetorical relations are ultimately insufficient to identify metaphor. We think that background information, comprising what Asher and Lascarides call world knowledge (like knowing that a political career for someone normally starts at the level of clerk and might culminate in becoming Prime Minister) and perceptual information also play a role in deciding whether a sentence should be interpreted as a metaphor or as a literal claim. Second, we argue that rhetorical relations, understood as relations among sentences, are sometimes not just insufficient but simply inadequate to account for metaphor identification via contextual coherence. We go beyond Asher and Lascarides's view to accommodate examples where an utterance is identified as a metaphor, not by virtue of discourse relations established with previous discourse, but by how it stands in relation to perceptual information or background knowledge. In some cases, extra-linguistic elements of the situation in which the sentence is uttered can play a role analogous to that of the linguistic context. Our view integrates Stone et al.'s (2013) account of the role perceptual information plays in making an utterance coherent. This is to say that contextual coherence is sometimes determined by discourse relations established between an utterance and perceptually accessible features of a situation. We also consider cases where metaphor identification takes place in impoverished contexts, where we know relatively little aside from the topic under discussion. We finish with a suggestion about how Questions Under Discussion (QUD) could be used in metaphor identification, given that QUD are a way of checking whether an utterance addresses the topic under discussion.

We thus unfold and refine a notion of contextual coherence based on Asher and Lascarides's work, but broader than the notion of discourse coherence they develop. This allows us to account for metaphor identification in cases they would not be able to account for.

2. Metaphor Identification and Interpretive Ambiguity

As mentioned, our focus in this paper is on how to tell whether a particular sentence, uttered in a specific context, is an instance of metaphor. We call this the

[1] This expression was coined by Cohen (1976) so as to draw attention to a range of cases that serve as counterexamples to deviance theories of metaphor identification, which characteristically relied on category mistakes as a cue.

[2] See for instance Asher and Lascarides 1995, 2001 and 2003.

problem of metaphor identification.[3] Stern notes that identifying a sentence as a metaphor is not the same as interpreting a metaphor (Stern 2000: 7). One might have good reasons for believing that an utterance is best interpreted as a metaphor without knowing what precisely it says. For instance, we may suspect that an utterance of (2) is metaphorical:[4]

(2) Sam is a pebble.

At the same time, we may fail to draw the relevant comparison between Sam and a pebble. And, conversely, we may have a pretty good idea of what a certain sentence would say if it were to be taken metaphorically, while being ignorant about whether a given utterance of that sentence is best construed as a case of metaphor. This could happen with example (1).

In order to understand an utterance, interpreters typically have to do a number of things, such as determining the proposition expressed, which includes resolving syntactic and lexical ambiguities, and determining the referents of indexicals and context-sensitive expressions; identifying the speech act performed; or inferring implicatures. In some cases, the interpreter must also decide whether the utterance is to be understood as literal or not. Take again example (2), "Sam is a pebble". This sentence can be used as a metaphor, but also as a metonymy (if Sam is a child dressed as a pebble, in the context of a school play) and perhaps even as a literal claim (in a fictional work as a cartoon, or a surrealist poem, for instance).

In this sense, the situation is similar to that of lexical ambiguity. In a case of *lexical* ambiguity (homonymy), the interpreter has to decide, using contextual information, which of a list of potential meanings is the intended or otherwise correct one. For instance, if faced with an occurrence of the word "bank", the interpreter has to find out whether the utterance is about a financial institution or the side of a river. Similarly, when faced with an occurrence of "Sam is a pebble", the interpreter has to find out which of different possible types of interpretation is most appropriate for that sentence, given the context in which it was uttered: Is the utterance to be understood as a metaphor, a metonymy, or a literal claim? In what follows we restrict the discussion to the distinction metaphorical versus literal interpretation, and proceed as if the interpreter only had to choose between these two options of resolving what we might call *interpretive* ambiguity.

It might seem that, in the case of (2), the figurative ambiguity between literal and metaphorical interpretation is very easy to resolve. Assuming that Sam is the name of a person, the sentence would be semantically deviant, and thus either

[3] Stern (2000: 3) officially calls it the question of metaphor "recognition", while describing it as addressing "the conditions, heuristics, clues, cues, trains of reasoning, or steps followed by speaker-hearers by which they identify or recognize particular utterances as metaphors". We find it more intuitive to speak of metaphor identification here, thus adopting for our own theoretical purposes a terminology that is established in cognitive linguistics (cf. e.g. Pragglejaz Group 2007).

[4] This example appears in Asher and Lascarides 2001. It is an example of a so-called nominal metaphor (of the form 'A is B'). Metaphors come, of course, in a host of other syntactic shapes. This is noteworthy insofar as syntactic structure exerts distinctive constraints on intrasentential coherence. There is an interesting literature on the procedures, such as coercion (cf. e.g. Pustejovsky 1995, Asher 2011), by which intrasentential coherence is established when the specifications of semantic types conflict, as in category mistakes or metaphor (cf. Magidor 2020). Our focus, however, is on intersentential or contextual coherence. As Prandi (2021: 64) notes, the latter wins out in case the demands of the two types of coherence conflict.

meaningless or false. The interpreter could use a Gricean-style reasoning to conclude that the speaker must mean her words in a metaphorical sense. However, it is important to note that many metaphors are not semantically deviant. Contrary to what deviance theorists might have thought, no internal feature of the sentence is a reliable guide to metaphoricity, and the interpreter has to make a choice using the available contextual information. Consider a little story surrounding example (1):

> After being appointed Prime Minister, Disraeli said: "I have always despised politics. But I have climbed to the top of that greasy pole".

Example (1) does not involve any category mistake, yet it can be identified as a metaphor. Such an utterance does not have internal features that can be used as a reliable mark of metaphoricity. Our proposal, which we flesh out in sections 4 and 5, is that the key to the identification problem might be in the structure of the discourse in which the target sentence is embedded—as Asher and Lascarides hold—together with extra-linguistic elements of context such as perceptual information and background knowledge available to the interpreter. Before putting it forward, we review Asher and Lascarides's view and argue that it needs to be supplemented with extra-linguistic information.

3. Asher and Lascarides's View on Metaphor

In their paper "Metaphor in Discourse", Asher and Lascarides (2001) argue that it is possible to specify the principles of metaphorical interpretation. In their view, lexical rules and discourse structure—in the form of rhetorical relations—constrain metaphorical interpretation. Interestingly, their account works for metaphors involving category mistakes and for those that do not bear this mark. In this section, we will briefly present their view, by assessing its merits in regards to the question of metaphor identification.

Asher and Lascarides have put forward a theory of interpretation, the Segmented Discourse Representation Theory (SDRT),[5] in which the principle governing interpretation is *Maximize Discourse Coherence*. The underlying idea is that utterances in discourse are connected to one another. The principle establishes that we should prefer interpretations that maximize coherence—roughly, those allowing for as many connections as possible. The connections at stake are called rhetorical relations, and include Elaboration, Narration, Contrast, Question-Answer-Pair, etc. Rhetorical relations describe rhetorical roles utterances play in discourse context (Asher and Lascarides 2003: 3), for instance, constituting an explanation of a previous utterance. What is of interest for us here is that rhetorical relations can be used to explain constraints on interpretation. In SDRT, rhetorical relations are used to model the semantics/pragmatics interface. According to this theory, the logical form of discourse contains rhetorical relations, and these rhetorical relations can have truth-conditional effects on the sentences they link. Consider the following example, in which "bar" is ambiguous and the interpreter must find out which sense is correct (Asher and Lascarides 1995):

(3) a. The judge asked where the defendant was.
b. The barrister said he was in the pub drinking.
c. The bailiff found him slumped beneath the bar.

[5] See Asher and Lascarides 1995 and 2003.

How does discourse structure constrain interpretation—by resolving a lexical ambiguity in this case—and what role does coherence play? Asher and Lascarides's account goes roughly as follows. The discourse structure in (3) is narration. Utterance (3b) is connected to (3a) by the relation Narration, and so is (3c) to (3b). This relation imposes spatio-temporal constraints in the events described. The narrative links are tighter when the interpretation of the sentence—here, utterance (3c)—complies with expectations created. In this example, the expectation would be that the situation described in (3c) is set at the place introduced in (3b)—namely, the pub. "Bar" is disambiguated accordingly. Hence, Asher and Lascarides hold that discourse structure influences ambiguity resolution.

In their 2001 paper, they apply this framework to metaphor and identify two types of constraints on interpretation. First, lexical rules predict that some word occurrences take metaphorical meaning. For instance, there is a rule that establishes that "rock" can take an argument of the type *human* and that, when this happens, the original physical object meaning applies in a metaphorical sense to the relevant human, as happens in "John is a rock".[6] Second, rhetorical relations can trigger metaphorical interpretations. One of the examples discussed involves (1), now slightly expanded by being preceded by the sentence (1*a), which couches the original example in a little story:[7]

(1*) a. I have always despised politics.
b. But I have climbed to the top of that greasy pole.

"But" in (1*b) signals the rhetorical relation of Contrast. This relation connects (1*b) with (1*a), and ensures that they match at the structural and semantic level. Since the two sentences are thus connected, the referent of the anaphoric expression "that greasy pole" is provided by the first sentence: "that greasy pole" refers to politics. Moreover, given that one cannot climb politics in a literal sense, the resolution of the anaphora forces a metaphorical interpretation. As Asher and Lascarides write, "the relation, together with the proposition expressed by the first sentence, triggers the metaphor in the first place" (2001: 285).[8]

Discourse coherence rendered via rhetorical relations with previous discourse, however, cannot be the whole story. Asher and Lascarides note that "computing rhetorical relations involves nonmonotonic reasoning on the reader's background knowledge" (2001: 283). Deciding which rhetorical relation(s) hold(s) between the sentence and its preceding discourse might not be as automatic as the presence of "but" suggests.[9] Furthermore, even if the rhetorical relation Contrast is clearly identified, background knowledge plays a role in the identification of metaphor. Asher and Lascarides (2001) invoke knowledge about "fundamental values in our culture" which allow the interpreter to construct a scale, from clerk to prime minister, at stake in the spatial projection forced by the change of location denoted by the verb

[6] "Predicts" is probably too strong, since it is possible to use "John is a rock" in its literal sense, or as a metonymy.

[7] (1*) includes (1) in the sense that (1*) expresses a conjunction of two propositions, one of which is the proposition semantically expressed by (1) (ignoring the complications created by the indexical).

[8] This is Asher and Lascarides' explanation of the example. However, as a reviewer has noted, the presence of 'but' could also suggest a different interpretation. It could establish a contrast between politics and another domain that is as challenging as politics. This is precisely what we show in what follows.

[9] In Asher and Lascarides' view (2003), rhetorical relations are computed on the basis of composition and lexical semantics together with domain knowledge.

"to climb". Yet they do not offer a systematic integration of this sort of background knowledge when it gets to establishing discourse relations between the proposition at hand and the utterances preceding it.

In the remainder of this section, we offer three examples showing that discourse structure established upon preceding utterances is insufficient to solve the metaphor identification problem. First, imagine two friends chatting about their holidays.

(4) A: a. Did you enjoy your trip to Moscow?
B: b. Not really.
B: c. Moscow is a cold city.

B's utterance is linked to A's utterance via the relation Question-Answer-Pair, as per the Segmented Discourse Representation Theory (SDRT) framework. A coherent interpretation of B's utterance would take it to address A's question. This relation connects (4b) to (4a). What about (4c)? A plausible interpretation here is that c explains why the speaker did not enjoy Moscow. The link here is given by the rhetorical relation Explanation. Now, either a literal or a metaphorical interpretation of "Moscow is a cold city" would count as an explanation. Taken as a literal claim, B can be interpreted as saying that the temperatures in Moscow are very cold. This explains why she did not enjoy the trip. But a metaphorical interpretation would also work. Taken as a metaphor, "Moscow is a cold city" would be saying, roughly, that the inhabitants are unfriendly. This interpretation would also explain why she did not enjoy the trip.

What clues can the interpreter rely on that would help her decide between the literal and the metaphorical interpretation? In this first example, she could have recourse to what we call background information. Imagine that A knows that B lives in a place where temperatures are lower than in Moscow, and that B does not mind cold weather. With this information, it makes more sense for A to interpret B as speaking metaphorically.

As a second example, imagine a visibly restless teenager who tells her friend:

(5) a. I need to get out.
b. I can't breathe.

Again, (5b) can be interpreted literally, as referring to a physical symptom, or metaphorically, as a description of feelings. In this example, the second sentence is again connected to the first via Explanation, which is again insufficient to decide between the metaphor-literal interpretations. Both the literal and the metaphorical interpretation could constitute an explanation of why the speaker needs to get out. However, the perceptual information available to the interpreter might supplement discourse structure and make the metaphorical interpretation preferable. Information about how the teenager behaves, gestures, or about her physical condition could tip the scale towards one interpretation or the other.

Finally, consider again (1*):

(1*) a. I have always despised politics.
b. But I have climbed to the top of that greasy pole.

Although discourse structure points towards a metaphorical interpretation, it is not difficult to come up with more complex examples in which extra-linguistic information suggests otherwise. Imagine, as historically implausible as it may sound, that Disraeli utters (1*) before becoming Prime Minister and after climbing an actual greasy pole. In this case, it would be reasonable to interpret (1*b) as a

literal claim. What about the whole fragment, (1*)? What role is (1*a) playing, and how does it constrain the interpretation of (1*b)? (1*a) has introduced the domain of politics. Given that the sentences are connected by "but", which cues towards the rhetorical relation Contrast, and the matching congruity this relation imposes, it makes sense to interpret (1*b) as implicitly referring to the domain of politics introduced by (1*a). However, "that" could refer to the actual pole and the utterance would be literal. In this context, in order to make sense of the contrast relation, we could interpret (1*b) as a suggestion that Disraeli is capable of great achievements. The contrast here would be between the claim that the speaker despises politics and the suggestion that he is capable of great achievements in the domain of politics—between (1*a) and an implicature communicated by (1*b).[10]

4. Contextual Coherence

The examples above suggest that coherence should be understood as a notion that comprises preceding discourse as well as extra-linguistic information. Instead of understanding coherence as being based only on rhetorical relations that can be established between the target sentence and previous linguistic utterances, we think that coherence is established by also taking into account the surrounding, perceptually accessible context of utterance, and even background knowledge accessible to the interpreter. We develop a broader notion of contextual coherence that expands on Asher and Lascarides's theory of discourse coherence. In our view, interpretation in general, and metaphor identification in particular, is governed by the following principle:

> **Principle of Contextual Coherence:** The preferred interpretation of some utterance is (the) one that is contextually coherent. As a corollary of this principle, contextually incoherent interpretations should be avoided.[11]

What do we mean by "contextually coherent"? The remainder of the paper elaborates on this. Here is a first sketch.

> **Contextual Coherence (basic definition):** For an interpretation of some utterance to be contextually coherent, two conditions must be met:
> (1) It must cohere with the discourse in which the target sentence is embedded.
> (2) It must be externally consistent, by which we mean consistent with the context, that is, with the preceding discourse, perceptual information available to the interpreter, and background knowledge within her reach.

The first condition can be rendered more precise if we rely, as Asher and Lascarides and others do, on rhetorical relations, and understand 'cohere with the discourse' in (1) as meaning that the target sentence is connected to other parts of the discourse by rhetorical relations. (We shall see in the coming section,

[10] Suppose that, in the context described, it is clear that (1*b) is literal, because Disraeli has just climbed to the top of an actual greasy pole, a pole that is still perceptually salient. This literal interpretation is in principle unrelated to (1*a), for, what relation is there between despising politics and climbing greasy poles? However, speaker and audience can exploit Grice's maxim of relevance and derive the implicature that the speaker is capable of great achievements even in unpleasant domains, such as politics.

[11] This leaves room for there being more than one contextually coherent interpretation, something we explore in Crespo, Heise and Picazo (ms.).

however, that this strategy is too restrictive.) Thus viewed, condition (1) on contextual coherence would establish that the preferred interpretation is one in which the target sentence is connected to at least one previous sentence in the discourse via a rhetorical relation.[12] This already precludes certain interpretations and favours others. Asher and Lascarides's account of discourse (1*) shows that, in absence of further information, the Contrast relation precludes a literal interpretation of "I have climbed to the top of that greasy pole".

Nonetheless, we have argued in section 3 that, in some cases, rhetorical relations do not suffice to account for the preferred interpretation. We may encounter at least two kinds of scenarios in which the decision as to whether an utterance is best understood as a metaphor or as a literal claim is affected by other sources of information. On the one hand, some rhetorical relations do not resolve the interpretive ambiguity (examples 4 and 5 above). On the other, extra-linguistic information can override an otherwise coherent interpretation (example (1*)). This is why we need condition (2) on contextual coherence.

Condition (2) on contextual coherence holds that the interpretation must be consistent with the context. As suggested by our analyses of examples in section 3, we believe it is best to conceive of context as a rich and heterogeneous body of information. Here we adopt the perspective of the interpreter and include in this body of information all the information available to her. This encompasses information from three different sources. First, a context includes the information given by previous discourse. This includes what has been explicitly said, and also implicatures and presuppositions. It is difficult to decide what counts as previous discourse,[13] but let us simply note that sometimes previous conversations might count as previous discourse. Second, the context includes perceptually accessible information given to the interpreter. This is especially relevant for oral exchanges, where the interpreter has information about the conversational setting, the context of utterance. Third, interpreters have what we will call background information, that is, general knowledge that can be used for the purpose of interpretation akin to what Asher and Lascarides call world knowledge, but also other sorts of background knowledge, for instance, knowledge about punctuation, register, tone, and other conventional elements of oral and written discourse.[14] Here we include as well information about the speaker or writer (who she is, etc.) and more generic information, including, for instance, sociological, historical or literary knowledge that may allow an interpreter to classify a text or exchange according to general criteria.

All this presupposes an ideal interpreter, someone able to deal with large bodies of information. Actual interpreters have limitations and are often unable to reason on the basis of the complex body of information that, in our view, constitutes the context. Moreover, actual speakers might be inattentive, fail to reason adequately, commit mistakes, etc. Therefore, in our proposal, we work with an

[12] As in Asher and Lascarides's framework, condition (1) would thus not apply to the first sentence of some discourse.

[13] To give a flavour of the difficulty in question, consider a chat between two friends: It is likely that their exchange draws in part on things they talked about in the past. This is not the case with two strangers who strike a conversation. In view of such problems of delimitation, various proposals have been elaborated on how to restrict the domain of previous discourse. Relevance Theory (Sperber and Wilson 1995) would be one way to go forward, but there are others too. We do not wish to commit here to one particular proposal.

[14] Syntactic constraints would figure among the relevant linguistic conventions (cf. fn. 4).

ideal interpreter. Furthermore, presumably some pieces of previous discourse and extralinguistic elements, but not others, will be salient. In our proposal we assume that the context includes information about what is salient, without deriving salience from more basic contextual information.

Condition (2) on contextual coherence urges the interpreter to avoid interpretations that are inconsistent with contextual information. Here we mean "inconsistent" in a loose sense. Besides being logically inconsistent, an interpretation can be inconsistent because of material incompatibility or presupposition failure. In spite of this rather loose notion of inconsistency, our proposal is quite conservative in that we think of inconsistency as a relation between propositions. Thus, the three types of information that make up the context have propositional format, including perceptual information. The question of whether perceptual experience has propositional content is anything but settled, of course (see Crane and Craig 2021). Thus, we are making a substantive assumption here.

The assumption that inconsistency is a relation between propositions requires that we are able to consider metaphorical meaning in a propositional manner. But is there something equivalent to a proposition when it comes to metaphorical meaning? We propose to conceive metaphorical meaning in terms of minimal paraphrases.[15] A minimal paraphrase is a propositional rendering of what the metaphor expresses. It is minimal in the sense of not being exhaustive. Some of the effects of a metaphor might involve mental imagery and thus be impossible to capture by means of a proposition. Moreover, the full significance of the metaphor might be so rich that it is very difficult to fully articulate its meaning. However, the minimal paraphrase captures enough so as to enable the interpreter to reason on the basis of the metaphor (to derive implicatures, for instance), and to assess consistency with contextual information.

The notion of contextual coherence we propose works well for cases as those considered in section 3 in which rhetorical relations are insufficient to account for the preferred interpretation of an utterance. Let us go back to example (4). In section 3 we noted that the rhetorical relation Question-Answer-Pair is insufficient to decide whether the best interpretation of "Moscow is a cold city" is metaphorical or not. Both the metaphorical and the literal interpretation provide an explanation as to why B did not enjoy her trip. However, we also noted that information about the speaker—including what we have called background information—might lead us to prefer one interpretation over the other. In particular, the information that the speaker does not mind low temperatures is inconsistent—in our loose sense—with interpreting her as meaning "Moscow is a cold city" in a literal sense. The two propositions that are inconsistent can be specified as follows: the proposition that the speaker didn't enjoy her stay in Moscow because temperatures in Moscow are low (derived from the literal meaning of "Not really. Moscow is a cold city" and the rhetorical relation of Explanation), and the proposition that the speaker does not mind low temperatures. By contrast, the metaphorical interpretation of "Moscow is a cold city" makes sense of the rhetorical relation and is consistent with the context.

[15] Non-cognitivists such as Davidson (1978) would disagree with the claim that metaphors have minimal paraphrases, and interaction theorists such as Black (1954) would hold that paraphrases come with a loss of cognitive force. However, there are linguistic reasons for positing such (minimal) cognitive contents for metaphorical utterances. For instance, metaphors can be used as vehicles for implicatures, see Bezuidenhout 2001 and Camp 2006.

Similarly, in example (5) perceptual information might preclude the literal interpretation. Suppose that the hearer looks around and sees nothing that could prevent the speaker from breathing in the literal sense—no smoke, no asthma symptoms, etc. In this case, the literal interpretation would be inconsistent with the perceptual information that there are no signs of a physical inability to breathe.[16] The metaphorical interpretation is preferred. It is connected to previous discourse by a rhetorical relation and is consistent with perceptual information (the speaker is visibly restless).

5. Rhetorical Relations, Situations and Topic under Discussion

In this section, we discuss cases that lead us to modify the principle of contextual coherence sketched in section 4. Here, rhetorical relations, understood as relations between a target sentence and previous discourse, are not insufficient but inadequate to account for metaphor identification. One reason is that some sentences are connected to extra-linguistic context. Another reason is that we are sometimes able to identify metaphor even if all we have at our disposal is an impoverished context and no previous discourse at all. These cases will lead us to modify our conception of contextual coherence.

The first reason why we think it is necessary to go beyond the notion of rhetorical relation that we have relied on so far—relations established between an utterance and propositions available from previous discourse—is that rhetorical relations, or at least something akin to them, can also be established between an utterance and its extra-linguistic context. Let us go back to example (5):

(5) a. I need to get out.
b. I can't breathe.

Imagine again that the utterer is a restless teenager with tears in her eyes. In the previous section, we noted that (5b) is an Explanation of (5a). But we think that this is not the whole story. In the scenario described, it seems that part of the reason why we take (5b) to be metaphorical has to do with how it is connected to the real-world situation, not merely by being consistent with it—as required by our condition (2) on contextual coherence—but as being an explanation thereof. It seems that here the proposition enters into a discourse relation (a rhetorical relation) with extra-linguistic elements of the context in which it is uttered, to the extent that they are perceptually accessible to the interpreter. Note that this relation of explanation would also hold between the situation and (5b) when not preceded by (5a).

Stone et al. (2013) let perceptually grounded elements intervene in the discourse relations that model discourse coherence. They consider the role of perceptual information in the interpretation of situated utterances, and bootstrap this input by means of a simple dynamic semantics which introduces perceptually grounded discourse referents. We do not delve into the formal details of their proposal, but simply point out that such a move allows them to integrate perceptual input into the establishment of discourse relations. In their view, rhetorical relations (what they call discourse relations) can be established between an utterance

[16] Again, it would be inconsistent in our loose sense. In general, these loose inconsistencies can be translated into logical inconsistencies by adding the implicit assumptions that are used in reasoning. For instance, here we could add the premise that the physical inability to breathe always comes with certain visible signs.

and what happens in the real-world situation the utterance pertains to. Following this idea, we are in a position to suggest that (5b) yields an Explanation of what is happening in the corresponding real-world situation, which is perceptually available to the interpreter.

Something similar could happen with example (1). Imagine a recently appointed Prime Minister who receives an old friend in Downing Street. Looking around, she utters: "I have climbed to the top of the greasy pole". We think that this utterance, in the context described, is rhetorically connected to the situation in which the exchange takes place. For instance, it might serve as an explanation of why the speaker is in Downing Street. If this line of reasoning holds good, then the relevant relation of Explanation is stronger than the notion of loose consistency ensured by our condition (2) on contextual coherence.

This raises a concern for our previous characterisation of contextual coherence. The restrictive view on condition (1) advanced in section 4 had it that the target sentence should stand in a rhetorical relation with other propositions made available by previous discourse. Our discussion above suggests that a more accurate condition would be:

> **Contextual Coherence (with situated utterances):** For an interpretation of some utterance to be contextually coherent, two conditions must be met:
>
> (1*) It must cohere with previous discourse or the real-world situation in which it is embedded.
>
> (2) It must be externally consistent, by which we mean consistent with the context, that is, with the preceding discourse, perceptual information available to the interpreter, and background knowledge within her reach.

As in the basic definition, condition (1*) can be spelled out using rhetorical relations. These relations can now connect the target sentence with sentences in the previous discourse or with perceptually grounded discourse referents.

Another reason to revise our conception of rhetorical relations, in our attempt to account for metaphor identification via contextual coherence, is provided by a different kind of example. Let us go back to (1):

> (1) I have climbed to the top of that greasy pole.

In section 1 we provided some contextual information for (1). Crucially, we provided some (very limited) previous discourse. Now, as it turns out, many interpreters of (1) only know that this sentence is a quote from Disraeli, something he said after being appointed Prime Minister, together with some minimal knowledge about who Disraeli was. For many interpreters, (1) comes in the form of a direct quote, as in the following example:

> After being appointed Prime Minister, Disraeli said: "I have climbed to the top of that greasy pole".

Previous discourse is not available to these interpreters, and yet they might be able to identify the sentence as a metaphor—as being about politics, and not about an actual greasy pole. Note that what they identify as a metaphor is Disraeli's utterance of the sentence, that is, they take Disraeli's words to be metaphorical. Their interpretation is not about someone else's use of the same words, say a reporter.

We can speak here of metaphor identification in impoverished contexts.[17] The situation is analogous to a well-known problem of inter-contextual communication. Inter-contextual communication includes all those instances of communication in which speaker and audience are not co-situated and lack a common ground. For example, reading a report of what someone said, either in the form of a direct quote or as a homophonic indirect report, would be such a case. The problem of inter-contextual communication is that the interpreter lacks some important contextual information. In particular, and going back to the notion of coherence introduced above, the interpreter lacks access to the previous discourse or real-world situation with which to establish rhetorical relations. Despite this, non-co-situated interpreters are sometimes able to identify metaphors, including twice-true metaphors.

Impoverished contexts motivate an even more minimal notion of contextual coherence. In our view, the non-co-situated interpreter has what we have called 'background information'. For example, she might have information about the speaker. A promising option of filtering that information is to consider the topic under discussion. When interpreting example (1), we might lack knowledge of previous discourse and have no perceptual access to the situation that the utterance is related to. However, by drawing on background information, it seems possible to reconstruct at least the topic. Since we know that Disraeli was a politician, we can venture the guess that the utterance might be about politics. This piece of information plausibly plays a role in identifying the utterance as metaphorical.

As the foregoing shows, we think that the topic under discussion can sometimes be reconstructed from background information. Remember that this includes information about the speaker and known historical facts, for instance. Thus, background information can play two roles. First, it may be used to infer a topic under discussion. Second, it enters into the assessment of contextual consistency.

This motivates a third version of the notion of contextual coherence:

Contextual Coherence (even in impoverished contexts): For an interpretation of some utterance to be contextually coherent, two conditions must be met:

(1**) It must *address the topic under discussion.*

(2) It must be externally consistent, by which we mean consistent with the context, that is, with the preceding discourse, perceptual information available to the interpreter, and background knowledge within her reach.[18]

To finish, we would like to suggest that this notion of contextual coherence can be applied beyond impoverished contexts. Even in cases in which there is a rich previous discourse, it might make sense to use the topic under discussion, instead of relying only on rhetorical relations, to decide whether the target sentence is literal or metaphorical. We will illustrate this possibility with example (5).

[17] See Picazo 2022 for a discussion of inter-contextual communication and the kind of information that is shared across contexts.

[18] Note that an interpreter with little access to elements of the context of utterance has more chances of misinterpreting what the speaker. The less contextual information, the more interpretive failures can arise. This is so because the interpreter can lack contextual information that would defeat her interpretation of the utterance, or information that would support a different interpretation.

Consider another context for example (5), this time Tamara Bach's (2003) novel *Marsmädchen*. The relevant paragraph reads as follows:

> I have to get out. I have to do something, now, immediately. The room is too small, it's afternoon again, and it's always the same: too similar, too small. Not just my room, heck, the whole house, this city. I have to get out, I can't breathe in here (Bach 2003: sect. 5).

As we mentioned above, (5b) is related to (5a) via Explanation. Other relations can be established in the text, and rhetorical relations can be used to decide that the target sentence is metaphorical. The explanation might go as follows. The two sentences before (5) seem to be related to (5) via Result or Cause-Effect—the room, the house and the city are too small, so the speaker feels confined and has to get out. We think that this relation, together with some background knowledge, constrains interpretation in a way that makes the metaphorical interpretation coherent, and hence preferred. In this fragment, the relations among the sentences impose some constraints on interpretation. First, 'here' in (5) must refer to a location introduced in the previous sentence, if the fragment is to be coherent. We conclude that it refers to the room, the house, the city. Using only the latter, the target sentence says that the speaker needs to get out of the city. We have split (5) into two sub-sentences connected by Explanation. This relation imposes a connection between the two sentences such that not being able to breathe must be interpreted as a reason for needing to get out of the city. At this point, we need to use background information. We know that the size of the city does not cause a physical inability to breathe. It can, however, constitute an oppressive environment. This metaphorical interpretation is more coherent.

Rhetorical relations may be appealed to here. However, we think that the topic under discussion can play the same role. The novel is about a teenager, Miriam, who falls in love with her friend Laura. One topic under discussion concerns Miriam's feelings—how she feels, how she reacts to an oppressive environment. The whole paragraph can be read as addressing this topic. If we think of the paragraph as addressing Miriam's feelings, then the metaphorical interpretation would be coherent and the literal not.

How to account for a topic under discussion? In related work, we have proposed to view the topic under discussion as being given by the Questions Under Discussion (QUD) in the context, as identified from the interpreter's point of view (Crespo, Heise, Picazo, ms.). As rhetorical relations, QUD is a framework to model discourse structure.[19] The basic idea is that the utterances in a conversation are interpreted by the interlocutors relative to the question being addressed, that is, relative to the topic under discussion. We think of QUD as a partially ordered set of questions, i.e., a set of questions ordered by a binary transitive and antisymmetric relation. Peculiar to our view is that we consider that QUDs may issue from the three sources that make up context: previous discourse, perceptual information and background information. That the topic under discussion may be inferred from previous discourse, and also from the target sentence itself via accommodation, is standardly accepted. But we think that it is also important to pay attention to the role that perceptual information and background knowledge play.

[19] Originally proposed in von Stutterheim and Klein 1989 and van Kuppevelt 1995. Later elaborations can be found among others in Ginzburg 1996, Roberts 2012, and van Rooij 2013.

As for perceptual information, whatever is perceptually salient in the situation in which a conversation takes place may become a topic under discussion. Gestures, for instance, may introduce a topic under discussion, as when pointing makes an object or individual salient. These moves can introduce brand new topics but also sub-topics that relate to the main topic under discussion. As for background information, we have already argued that it plays a role in the determination of the QUD, notably in impoverished contexts.

With this in mind, how to conceive of condition (1**) on contextual coherence? We believe that the target sentence should address an issue raised by the heterogeneous, threefold context in which the utterance is given to the speaker, conceived as a live question in the QUD, that is, a question that has not yet been downdated, that has not been addressed already. Rhetorical relations or discourse relations, featured in the previous formulations of condition (1), are perhaps more variegated than the kind of constraint given by (1**) in our third attempt. However, they seem to be too stringent in how rich the context needs to be for them to be established in the first place. Hence we suggest that, at least for many examples, QUD can replace rhetorical relations in condition (1). If successful, this option would allow us to treat the different kinds of examples we have considered so far in a uniform manner.

Contextual Coherence (with QUD): For an interpretation of some utterance to be contextually coherent, two conditions must be met:

(1***) It must *answer a live question in the QUD.*

(2) It must be externally consistent, by which we mean consistent with the context, that is, with the preceding discourse, perceptual information available to the interpreter, and background knowledge within her reach.

It lies beyond the scope of this paper to assess whether QUD are a better tool overall to solve the problem of metaphor identification. Our aim with this fourth definition is more modest: We suggest that this unified definition can be used to deal with many examples, including metaphors whose identification proceeds via previous discourse (cases where previous discourse introduces a question to the QUD), metaphors that involve real-world situations (cases where the question is introduced via perceptual information), and impoverished contexts (cases where background information suggests a QUD).

6. Conclusions

We have proposed a solution to the problem of metaphor identification based on the notion of contextual coherence. Following Asher and Lascarides, we hold that coherence guides the identification of figures of speech. We agree with Asher and Lascarides that coherence is primarily a notion that concerns discourse structure. However, we have shown that in order to decide whether a given utterance is a case of metaphor, rhetorical relations are sometimes insufficient and sometimes inadequate. The reason why they are insufficient is that sometimes rhetorical relations fall short of providing a basis for deciding between literal and metaphorical interpretations. Contextual information other than previous discourse, such as background information about the speaker, may then play a crucial role in identifying metaphors. We have captured this in a first notion of contextual coherence (the basic notion). According to this first notion, coherent interpretations must

cohere with the discourse they are embedded in and must be externally consistent, that is, consistent with the contextual information available to the speaker. Coherence with previous discourse could in principle be explained using rhetorical relations. However, we have argued that rhetorical relations, conceived as relations among sentences, are sometimes an inadequate tool to account for coherence. The reason is that extra-linguistic elements of the situation in which the sentence is uttered may be crucial in getting at the preferred interpretation. In order to account for these cases, we proposed that rhetorical relations may be established both with previous discourse and extra-linguistic situations (second notion, contextual coherence with situated utterances). Finally, we have suggested another possible implementation of the notion of contextual coherence. This alternative notion makes use of QUD instead of rhetorical relations (fourth notion, contextual coherence with QUD). What speaks in its favour is the fact that this alternative notion promises not only to explain the same kind of cases as the version based on rhetorical relations, but it also accounts for metaphor identification in impoverished contexts.[20]

References

Asher, N. and Lascarides, A. 1995, "Lexical Disambiguation in a Discourse Context", *Journal of Semantics*, 12, 1, 69-108.

Asher, N. and Lascarides, A. 2001, "Metaphor in Discourse", in Bouillon, P. and Busa, F. (eds.), *The Language of Word Meaning*, Cambridge: Cambridge University Press, 262-89.

Asher, N. and Lascarides, A. 2003, *Logics of Conversation*, Cambridge: Cambridge University Press.

Asher, N. 2011, *Lexical Meaning in Context: A Web of Words*, Cambridge: Cambridge University Press.

Bezuidenhout, A. 2001, "Metaphor and What is Said: A Defense of a Direct Expression View of Metaphor", *Midwest Studies in Philosophy*, 25, 1, 156-86.

Bach, T. 2003, *Marsmädchen*, Hamburg: Verlag Friedrich Oetinger.

Camp, E. 2006, "Contextualism, Metaphor, and What is Said", *Mind & Language*, 21, 3, 280-309.

Cohen, T. 1976, "Notes on Metaphor", *Journal of Aesthetics and Art Criticism*, 34, 3, 249-59.

Crane, T. and Craig, F. 2021, "The Problem of Perception", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/fall2021/entries/perception-problem/.

Crespo, I., Heise, A., and Picazo, C. (ms.), *Metaphor and Contextual Coherence: It's a Match!*

Davidson, D. 1978, "What Metaphors Mean", *Critical Inquiry*, 5, 1, 31-47.

[20] Andreas Heise wishes to acknowledge support by the grant ANR-17-EURE-0017. Claudia Picazo wishes to acknowledge support by the Ministerio de Ciencia e Innovación (Government of Spain) and European Union; grant PID2019-105728GB-I00 (MINECO/FEDER, EU).

Ginzburg, J. 1996, "Dynamics and the Semantics of Dialogue", in Seligman, J. and Westerståhl, D. (eds.), *Logic, Language and Computation*, Vol. 1, Stanford: Center for the Study of Language & Information (CSLI), 221-37.

Magidor, O. 2020, "Category Mistakes", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/win2020/entries/category-mistakes/.

Picazo, C. 2022, "Homophonic Reports and Gradual Communication", *Pacific Philosophical Quarterly,* 103, 2, 259-79.

Pragglejaz Group 2007, "MIP: A Method for Identifying Metaphorically Used Words in Discourse", *Metaphor and Symbol*, 22, 1, 1-39.

Prandi, M. 2021, "Syntaxe formelle et cohérence textuelle: Deux sources pour le conflit conceptuel", in Paissa, P., Conoscenti, M., Druetta, R., and Solly, M. (eds.), *Metaphor and Conflict. Métaphore et conflit*, Bern: Peter Lang, 55-75.

Pustejovsky, J. 1995, *The Generative Lexicon*, Cambridge, MA: MIT Press.

Roberts, C. 2012, "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics", *Semantics & Pragmatics*, 5, 1-69.

Sperber, D. and Wilson, D. 1995, *Relevance Theory*, Oxford: Blackwell.

Stern, J. 2000, *Metaphor in Context*, Cambridge, MA: MIT Press.

Stone, M., Stojnic, U., and Lepore, E. 2013, "Situated Utterances and Discourse Relations", *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, *Short papers*, Potsdam: Association for Computational Linguistics, 390-96.

van Kuppevelt, J. 1995, "Discourse Structure, Topicality and Questioning", *Journal of Linguistics*, 31, 109-47.

van Rooij, R. 2013, "Questioning to Resolve Decision Problems", *Linguistics and Philosophy*, 26, 727-63.

von Stutterheim, C. and Klein, W. 1989, "Referential Movement in Descriptive and Narrative Discourse", in Dietrich, R. and Graumann, C.F. (eds.), *Language Processing in Social Context*, Amsterdam: Elsevier, 39-76.

Lying and Misleading in Context

Palle Leth

Dalarna University

Abstract

In this paper I question the lying/misleading distinction from three different angles. I argue, first, that if speakers are responsible for what they explicitly say only and hearers for what they infer that speakers implicitly convey, it is practically impossible to enforce speaker responsibility. An implication of this view is that the lying/misleading distinction is untenable. Other attempts at questioning the distinction have been countered by empirical evidence of the robustness of the distinction. However, there is also contrasting empirical evidence that people do think that it is possible to lie by implicit means. I argue, second, that empirical evidence is irrelevant to the question which ought to be at issue, namely whether there are good reasons to make the distinction. Third, I argue that to the extent that the notion of misleading is in the service of inducing false beliefs by the statement of truths, the distinction does not seem to be morally well-founded. In short, I sketch an argument to the effect that there are no conceptual, empirical or moral reasons for making the lying/misleading distinction.

Keywords: Lying, Misleading, Context, Semantics, Pragmatics, What is said, Implicature, Intention, Responsibility.

1. Introduction

Many speakers, hearers and theorists take the distinction between lying and misleading to be a matter of course. It is commonly held that speakers lie when they say something which they believe to be false, whereas they are merely misleading when what they convey in addition to what they say is false. Speakers are responsible for what they explicitly say only; it is hearers who are responsible for what they take speakers to convey by implicit means. Therefore, speakers can be charged with lying, but are not responsible for their misleading utterances.

In this paper I will argue, first, that this division of responsibility is unstable. It is practically impossible to hold speakers responsible for their utterances unless they can be held responsible for hearers' assumptions concerning what they implicitly convey in addition to what they explicitly state. Next, I will examine the consequences of this view for the lying/misleading distinction. In particular, I will look at recent arguments from empirical evidence for the reality or unreality of the lying/misleading distinction. I observe that the question of whether the distinction

Argumenta 8, 1 (2022): 125-136
ISSN 2465-2334

DOI 10.14275/2465-2334/202215.let

should be preserved or abandoned cannot be answered by people's intuitions and habits. Last, I will question the moral respectability of the distinction. I raise the suspicion that the distinction chiefly serves the purpose of avoiding telling lies while inducing false beliefs and I suggest that a vindication of lying would perhaps be more honourable.

2. Responsibility and Hearer Assumptions

It seems natural to distinguish between what the speaker expresses in saying what she says and what the hearer infers from what the speaker says. There seems to be a difference between, on the one hand, what the speaker explicitly commits herself to by dint of the conventional meaning of the sentence she utters as specified by lexicon and syntax and, on the other hand, what the hearer takes to be conveyed by the utterance in the context at hand. The speaker may say "I have eaten" and at most occasions and by most hearers, such an utterance will be taken to convey that the speaker has eaten at some time prior to the making of the utterance at the day on which the utterance is made. But, after all, such a specification of the content of the utterance is just an assumption made by the hearer. This content is not expressed by the linguistic meaning of the sentence uttered, hence it is not explicitly stated. It is inferred by the hearer in the process of making sense of the speaker's utterance for the conversational purposes at hand. It may of course happen that the speaker's communicative intention is different from what the hearer takes it to be. Hearer assumptions may be more or less default and more or less well grounded. They seem to be defeasible in so far as they are concerned with the speaker's communicative intention, over which the speaker presumably has first person authority.

It is also natural to think that there is a division of responsibility going hand in hand with the difference between explicit meaning expressed by the speaker and implicit meaning inferred by the hearer. The speaker is responsible for what she says, the hearer is responsible for what she infers. The speaker's utterance of a certain sentence is an action performed by the speaker, whereas the hearer's assumptions in order to make sense of what the speaker says are performed by the hearer. Each one is responsible for her doings; the speaker is in control of what she utters, not of the hearer's assumptions (cf. e.g. Camp 2008: 8).[1]

Natural as this picture may seem, it is not without complications. I will confine myself to two such complications. Let us first consider the case of indexicals, like *she* and *here*. An indexical is a lexical item the linguistic meaning of which is a character consisting in constraints on, or instructions for recovering, its content or semantic value on a specific occasion of use (cf. Kaplan 1989: 500). When the speaker uses an indexical she certainly commits herself to its having some content or other, but she does not explicitly tell the hearer which the content is. It is the task of the hearer to work out what the content is on a given occasion of use. For some indexicals as used on some occasions, it is possible to contend that their contents are determined by specific parameters of the context of utterance, such

[1] It should be noted that recently some theorists have argued that the notion of assertion should be extended to cover not only explicitly stated meaning, but also some implicit content (see e.g. García-Carpintero 2018, Viebahn 2017, 2020, 2021 and Timmermann & Viebahn 2021). In so far, however, as the idea still is that the speaker is responsible for the implicit meaning which she commits to by dint of her intention, the problem which I will address remains basically the same.

as demonstrations or salience. Though the content of an indexical be not explicitly stated, it may be objectively settled. However, for many other cases it is more natural to consider that the instance of determinacy is the speaker's intention, in which case the content is not settled by objective means, but seems to fall under the speaker's authority.

Force is another important aspect of utterance meaning which is implicit and must be recovered by the hearer. It is true that there are many lexical items and constructions the function of which seems precisely to be to signal illocutionary force. Nevertheless, and in most cases presumably, the force of an utterance must be inferred by the hearer. This is certainly the case at the most general level. Speakers do not explicitly tell whether their utterances are to be taken as assertions or mere suppositions, whether as serious, ironical or jesting contributions to the conversation. Frege's (1972: 111) vertical and horizontal strokes for visualizing the difference between mere thoughts and judgments are well known, but they do not appear in conversation. And if they would, they would hardly be reliable. As Davidson (1984: 274) says: "nothing is more obvious than that there cannot be a convention that signals sincerity". The earnest assertive force of an utterance cannot be something which the speaker commits herself to by explicit means.

In most cases, then, when the hearer takes the speaker to have referred to a certain object or to have made an assertion, the hearer is relying on her assumptions concerning the speaker's intention. It is possible to hold that at the metaphysical level of meaning constitution, these aspects of utterance meaning are determined by the speaker's communicative intention (cf. Neale 2016: 246 and Harris 2016: 176). It is also possible to hold that at the metaphysical level, the speaker is genuinely responsible for (cf. Wallace 1994: 87) the contents and the force that she intends for her utterance. Now in the case of communicative intentions, we come to know about them chiefly through the speaker's words. The problem from a practical point of view is that if we take these crucial aspects of utterance meaning to be determined by the speaker's intention, it will be possible to for the speaker to escape liability at her earliest convenience. When taken to task for a certain implicit aspect of meaning, it would be possible for the speaker simply to deny that her intention was the one the hearer assumed it was (cf. Fricker 2012: 88 and Camp 2018: 51). In so far as the authority over the content of indexicals and the force of an utterance would be at the speaker's discretion, the speaker would not be authoritative at all (cf. Hobbes 2000: 84). Thus, if the speaker is only responsible for what she explicitly commits herself to and cannot be, or should not be, held responsible for what the hearer assumes her to convey, utterance responsibility will hardly be enforceable (cf. Rawls 1971: 213). Unless it be impossible to hold speakers responsible at all, it must be possible for hearers to hold speakers responsible for their assumptions.

If it is recognized that it should be possible to hold speakers responsible for hearer assumptions, the focus will not be on the implicit/explicit distinction, but on the warrantability of the hearer's assumptions. It is not the speaker's actual intention which matters, but what the intention appeared to be. Likewise the importance of truth conditional meaning for responsibility must also be rethought. Many theorists take for granted that the speaker is responsible for the meaning which is required for the utterance to have truth conditions at all (e.g. Saul 2012: 57). This more or less minimal truth conditional meaning is closely related to the explicit meaning of the sentence uttered. The fact that an utterance must be assigned truth conditions in order to make minimal sense does not tell anything

about which truth conditions it should be assigned. It is not clear why the speaker should be responsible for hearer assumptions concerning truth conditionally relevant meaning only. If it is thought that a speaker could be held responsible for having referred to a certain person by means of her utterance of the indexical *she*, it seems that she could also be held responsible for other kinds of implicit meaning. Assumptions about truth conditional meaning will certainly in many cases be more warranted than assumptions about more indirectly conveyed meaning. Sentences may contain constituents the value of which is mandatory. Implicatures, insinuations and other innuendoes do not at all appear in the surface structure of the sentence. But nothing in principle excludes that the hearer might have good reasons for other implicit meaning components than those required for truth conditions.

I have here only sketched an argument to the effect that the division of responsibility customarily taken to go with the distinction between what is explicitly said and what is implicitly conveyed cannot really be upheld. I will not pursue this argument here. Instead, I will look at an implication of the view that a speaker can in principle be held responsible for all kinds of implicit meaning, which may seem counterintuitive. If it is true that the speaker should be liable for what she implicitly conveys provided that the hearer's assumption in this regard is warranted, then it seems that the distinction between lying and misleading is threatened.

3. Implicatures as Misleading

The distinction between lying and misleading is a persistent feature of our conversational, moral and political culture. Even though the exact definitions of these notions are not of our immediate concern, it might be useful to see how they might look. The speaker lies to the hearer if and only if she

1) asserts that p to the hearer,
2) believes that it is false that p,
3) intends to deceive the hearer into believing that p.

It should immediately be noted that the last condition concerning the speaker's intention to deceive is taken by most theorists not to be required. The notion of misleading could perhaps be given along the following lines. The speaker makes a misleading utterance to the hearer if and only if she

1) asserts that p to the hearer,
2) believes that it is true that p,
3) intends the hearer to infer that q on the basis of that p,
4) believes that it is false that q,
5) intends to deceive the hearer into believing that q.

Speakers lie when they explicitly say what they believe to be false, while they are merely misleading when they only convey, in addition to what they say, what they believe to be false. It is commonly held that in cases of lying, the false content is stated by the speaker and falls under the speaker's responsibility, whereas in cases of misleading, the false content is inferred by the hearer and falls under the hearer's responsibility.

What does it take then to assert that something is the case? In the previous section, I argued that crucial aspects of regular assertions are not made explicit, e.g. that they are assertions. Other meaning components than indexical content may also be implicit. In practice it is not so easy to distinguish between the

speaker's asserting that *p* to the hearer and the speaker's merely conveying that *p*. If so, how should the division of responsibility between the speaker's and the hearer's responsibility be effectuated in particular cases? Here is an extract from a talk exchange between Holmes and Lestrade:

> "Are you armed, Lestrade?"
> The little detective smiled. "As long as I have my trousers, I have a hip-pocket, and as long as I have my hip-pocket I have something in it".
> "Good! My friend and I are also ready for emergencies" (Doyle 2006: 582).

Holmes's question is quite straightforward. But what does Lestrade's utterance amount to? Is it an answer to Holmes's question? And if so, does Lestrade by his utterance assert that he is armed? Or does he merely suggest that he is armed? Lestrade's statement would of course be true if he had nothing but a beetle in his pocket. It would also be possible for him to continue, without contradicting himself, by the addition of "namely a beetle". The implication that he is armed is certainly cancellable. It would likewise be possible for Lestrade to say what he does with the intention to communicate only that he has something or other in his pocket. It would certainly be possible for Lestrade to deny that he meant to say anything beyond his always having something in his pocket. These considerations speak of course against Lestrade's having by his utterance asserted that he is armed. Yet, in the context at hand, considering that Lestrade's utterance occurs immediately after Holmes's question, is not Holmes warranted in taking Lestrade's utterance as an answer to his question and furthermore as amounting to an assertion that he is armed, as is required at this stage of their operation? Is not Lestrade's utterance just a roundabout way of saying "of course"? Holmes himself, in response to Lestrade's utterance, does he state that he and some friend of his are ready for emergencies, whereas he merely conveys that he and Watson are armed as well? Would not Holmes, if it later turned out that Lestrade was fully aware that he had nothing but a beetle in his pocket, be justified in complaining about Lestrade's deceiving him into thinking that he was armed while he was not? Likewise, would not Lestrade, if it turned out that Holmes and Watson were ready for emergencies only because of the contents of Lestrade's pocket also be justified in complaining about being deceived? What kind of deception would it be? Would those cases be cases of lying or of merely misleading? Would the circumstance that the speaker did not explicitly state that he was armed be relevant in this regard? What difference would the warrantability of the interpretation make?

Meibauer has argued that false implicatures may well constitute lies. His argument is not based on the difficulty of drawing the distinction between assertion and implicature. His view is that the category of lying should be extended in so far as there is a difference between, on the one hand, deception in general and, on the other hand, deception directly or indirectly by means of linguistic utterances. Since an implicature is brought about by the act of assertion, it depends on it and forms a unity with it; hence from the viewpoint of lying, implicated content counts not less than asserted content (Meibauer 2011: 286). Meibauer supports his argument by his own and reported intuitions regarding cases where the speaker asserts that a certain person is not drunk, thereby implicating that usually the person is or where the speaker in response to whether she has seen a certain person, answers that the person is indisposed, thereby implicating that she has not seen the person. These cases are quite similar to the Lestrade and Holmes utterances quoted above,

so, I take it, that Meibauer would say that Lestrade and Holmes would be lying if they were not armed.

The contention that speakers may lie by means of false implicatures runs counter to the majoritarian view among philosophers. Philosophers generally tend to agree that judgments of lying depend upon what is said as opposed to what is conveyed by indirect means (see e.g. Borg 2019, Camp 2006, García-Carpintero 2018, Goldberg 2015, Green 2015, Saul 2012, Stokke 2018, Viebahn 2017). What is said is constituted by compositional meaning, lexical meaning and syntax, plus whatever else is required in order for the sentence to have truth conditions. What truth conditions requires varies somewhat from one theorist to the other, but, in any case, it is more or less closely related to the conventional meaning of the utterance. What is not required for truth conditions and what arises merely out of the hearer's attempt to make sense of what the speaker says in the context at hand cannot constitute content for which the speaker should be liable. Consequently, most theorists agree that false implicatures, i.e. implicitly conveyed falsehoods, cannot constitute lies, but are merely misleading. Intuitions about the distinction have even served as a test for the borderline between semantic and pragmatic meaning (Michaelson 2016). Soames, discussing the implicature of Grice's recommendation letter example, speaks of the 'temptation' to consider the implicature to be the real assertion of an utterance and says in a typical fashion that "the whole purpose of using indirect means to convey this information [i.e. the implicature] was to avoid having to state it" (Soames 2008: 443; cf. Kölbel 2011: 70). He thereby implicates that there is an advantage in not stating the information, which presumably is that speakers cannot be held responsible for only indirectly conveyed information. Among theorists on lying such a view is fairly widespread.

In addition to the convictions of theorists, there is the testimony from ordinary language users documented in various empirical investigations. However, this testimony goes in divergent directions. Weissman's and Terkourafi's investigation shows that informants in most cases do not consider false implicatures to be lies. The lie judgments of their informants rely on the explicitly stated meaning of utterances. Weissman and Terkourafi say: "Our results thus indicate that the literal meaning of an utterance is what people *tend* to consider when judging if an utterance is a lie or not" (Weissman and Terkourafi 2019: 239; my italics). Weissman and Terkourfi say 'tend', because the result of their investigation actually shows that informants' judgments are not categorical. Actually, the informants do consider some false implicatures to be lies. In fact, the result of the empirical investigation suggests that the lying/misleading distinction is less stable than what is generally assumed. Weissman and Terkourafi point out that "[t]o say that false implicatures are absolutely not lies, however, is an unfounded generalization" (Weissman and Terkourafi 2019: 236).

There are also a number of recent empirical investigations of the lying/misleading distinction where the informants take the opposite view and do consider false implicatures to be lies. Or, Ariel and Peleg observe that "even when what the speaker said is literally true, participants feel comfortable with classifying it as a lie if the PCI [particularized conversational implicature] is false" (Or *et al.* 2017: 101-102). Reins & Wiegmann report that their

> findings indicate that people's concept of lying is broader than commonly assumed, [in so far as] the participants in our study considered all of the deceptive

> presuppositions, deceptive GCIs [generalized conversational implicatures], and deceptive nonverbal actions, as well as most of the deceptive PCIs [particularized conversational implicatures] we investigated, to be cases of lying (Reins and Wiegmann 2021: 27 and 17).[2]

One problem about these investigations is that the scenarios which the informants are presented with depict a situation from which most of the conversational detail is left out. Such details may be decisive when it comes to telling whether the speaker is lying or misleading. Another problem is that the informants are instructed to rate the degree of lying or truth-telling. The informants are not given the option of classifying untruths directly as instances of misleading. The empirical evidence is thus far from conclusive.

This raises the general question of the relevance of empirical investigations for the lying/misleading distinction. Carson (2010: 33) says that "consistency with ordinary language and people's linguistic intuitions about what does and does not count as a lie is a desideratum of any definition of lying". In the same vein, Arico and Fallis (2013: 796) say that "lying does not seem to be the sort of thing about which people can be systematically mistaken". If we conceive of the lying/misleading distinction as carving the joints of natural kinds and also believe that people's intuitions are a reliable guide to natural kinds, the convictions of theorists and the intuitions of various informants have a considerable weight. But is it really credible that the lying/misleading distinction is a natural kind? Is it not rather the case that it depends for its existence on social conventions and institutions? If so, the question is not what lying is, but what we count as lying. Whenever something counts as having a certain status, it is of course possible to ask for the reasons for counting it thus and also to evaluate those reasons. When it comes to counting it is not a matter of intuitions, as Carson suggests, but of reasons. Our theory of lying should not respect pre-theoretic convictions, intuitions and inclinations, but spell out the reasons for counting the communication of certain untruths as merely misleading.

In response to empirical investigations confirming the lying/misleading distinction, it must be observed that they show at most that people are in the habit of making, and willing to uphold, the lying/misleading distinction. Since we knew in advance that in our culture there is such a thing as the lying/misleading distinction, the result that many people have the inclination not to count false implicatures as lies was only to be expected. Empirical investigations disconfirming the lying/misleading distinction present a less expectable result, in so far as it is counter to theoretical convictions and widespread intuitions. But in any case, these investigations do not tell us anything about the informants' reasons for making or not making the distinction, nor whether their verdicts are well supported.

It seems to me that when theorists and informants rely on what they perceive as literal meaning in their lie judgments and believe that they hold speakers responsible only for the explicit meaning of their utterances, they disregard the fact that, e.g., the semantic value of indexicals and the assertive force of utterances are not part of explicit meaning, but depend on their own assumptions. And if it is possible to hold speakers responsible for hearer assumptions in these regards, why should it not be possible to hold them responsible for hearer assumptions also in

[2] For further studies, see also Wiegmann and Willemsen 2017, Antomo *et al.* 2018, Skoczén 2021 and Wiegmann *et al.* 2021.

other regards? I thus think that there are good conceptual reasons to hold that whether the speaker lies is a matter, not of the degree of explicitness of the speaker's utterance, but of the warrantability of the speaker's interpretation (cf. Or *et al.* 2017: 103). It is not because an implicature is implicit and depends on hearer assumptions, that the speaker might not be responsible for it. It depends on whether the hearer is justified in taking the implicature to have been conveyed by the speaker. While the fact that an aspect of meaning is implicit does not by itself invalidate a hearer's meaning assignment it might often be an argument against the warrantability of her interpretation. In general it will be more difficult for a hearer to argue that her interpretation is warranted when it relies in great part on assumptions, rather than on linguistic facts. A hearer charging a speaker with lying must be in a position to argue that her interpretation, even though not grounded in the explicitly conveyed meaning, is justified by other considerations. A speaker defending herself against a charge of lying may of course point to the fact that she did not explicitly state what the hearer took her to convey, but such an argument should be construed as concerned with the justification of the hearer's interpretation. The condition of warrantability, though not as seemingly objective as the condition of explicitness, guarantees that charges of lying not be unfounded. An assumption must be warranted and even maximally warranted in order to support a charge of lying. But nothing in principle prevents a speaker from lying by means of a false implicature.

Here is an authentic case of a false implicature made by a speaker whose mother never helps her with the children:

> Boss (in a job interview): You have small children. How will you manage the long hours of the job?
> H.D.: I have a mother (Sternau *et al.* 2016: 708).

Of course, it would be true to state what the speaker says in virtually any circumstances. It would also be possible to continue by adding 'though she never helps me with the children'. Likewise, it would be possible for the speaker to have the communicative intention to convey nothing except that she has a mother. Yet the speaker, rightly on my view, takes herself to be lying when making this utterance in response to the question of her employer *in spe*. Likewise, according to my lights, Lestrade would be lying if he had nothing but a beetle in his pocket.

We have seen that the idea that false implicatures may be lies can be refuted as well as confirmed by empirical investigations. But the idea that our theory of lying should match intuitions about lying seems misguided. The question is not what lie judgments people are in the habit of making, but what lie judgments they have good reasons to make. My own view is that there are no conceptual grounds to make the distinction. Any lie could be said to amount to no more than misleading, once we take the explicitness condition seriously. It could be that there are nonetheless moral grounds to uphold the distinction. I will turn to this issue in the next section.

4. Deceptive Misleading

Is there really a difference between lying and misleading from a moral point of view, such that lying is impermissible and misleading permissible or at least less blameworthy than lying? In this section I will just make a few remarks in order to

question the idea that the lying/misleading distinction is natural, respectable and preservable.

There seem to be good reasons for condemning lying. For the sake of trust and reliability speakers should be strongly encouraged to tell the truth. By lying, Kant (1999: 612) says, "I bring it about, as far as I can, that statements (declarations) in general are not believed". There are also good reasons for distinguishing cases of lying from cases where the hearer without sufficient justification takes the speaker's utterance to convey a certain meaning which the speaker had no intention of conveying. The speaker should not then be charged with lying if it turns out that the implication is false. However, the notion of misleading appears in a different moral context.

The injunction to tell the truth, though well taken, is nevertheless in conflict with another deeply rooted inclination, namely that some truths had better be concealed. We may want to conceal the truth for more or less noble reasons, to save our face or the skin of others, but the conviction that some truths had better not be revealed is just as ingrained as the conviction that we should tell the truth. With regard to telling the truth, we have thus conflicting convictions. It seems that the lying/misleading distinction comes in handy as a way out of this dilemma.

Let us look at Augustine's (1952: 86 [ch. 13]) discussion of what to tell the persecutor of an innocent person the hiding place of whom the speaker is aware of. If the speaker is asked whether the refugee is or is not at the place where the speaker knows the refugee is, Augustine recommends the speaker not to state the truth 'I will not answer your question'. The reason is that this answer would raise the suspicion that the refugee was at the place asked about. Instead, Augustine advices the speaker rather to state the truth 'I know where he is but I will never disclose it'. The reason why this answer is to be preferred is, according to Augustine, that it will turn the persecutor's attention from the refugee's presence at the place asked about to the speaker's knowledge about the refugee's whereabouts. The second answer is advantageous in so far as it induces the persecutor falsely to believe that the refugee is not at the place asked about.

We see thus that for Augustine it is a matter of course that the speaker is not interested in frankly telling the hearer the truth. On the contrary, the speaker is anxiously interested in concealing the truth. It is equally a matter of course that the speaker does not want to state anything but the truth. And what is more, the speaker actually wants the persecutor to have false beliefs.

Augustine visibly displays a concern with, as Pascal (1866: 276) puts it, "how to avoid telling lies […] when one is anxious to induce a belief in what is false". In the casuist doctrines of equivocation and mental restriction which Pascal chastizes, the focus is clearly not on truth and trust, but on ingenuous and multifarious forms of deception (cf. Jonsen and Toulmin 1988: pt. IV). The categorial duty not to lie is served by all sorts of deceptions. Deceiving others is apparently fine as long as no attempt is made to deceive God (as if God could be). The focus in many discussions of misleading, also outside casuistry, is often not on deploring the eventuality that utterances may convey falsehoods, but on rejoicing at the possibility that truthful utterances may convey untruths. Misleading is conceived of not as an accident, but as an opportunity.

The question of whether the lying/misleading distinction emerged or persists because of its services to deception or is merely exploited by deception is empirical and cannot be settled except by historical investigations. I have done no more

than suggest that the existence of misleading is rather due to interest than discovery, that it is rather a social ploy than a natural kind. But *pace* Nietzsche (2009: 7 [pref. §6]), suspicious conditions of emergence and circumstances of occurrence do not automatically invalidate a concept (cf. Rawls 1971: 451).

Nevertheless, I believe that the readiness to deceive displayed in discussions of misleading should make us pause. If we are prepared to accept and to recommend deliberately conveying untruths, why not explore the permissibility of lying? The supposed advantage of misleading, in comparison with lying, is that the deception is not accomplished by the speaker, but by the hearer; it is the hearer who effectuates the faulty inferences. So rather than being deceived by the speaker, the hearer is deceived by herself. A disadvantage of misleading is that it permits the speaker to deceive herself into thinking that she does not deceive the hearer. What is morally most blameworthy, to frankly deceive the hearer or to deceive the hearer into deceiving herself?

5. Conclusion

I set out to show that the commonly made distinction between implicit and explicit meaning and the division of responsibility supposed to go with it are untenable. Crucial aspects of utterance meaning are implicit and it will be impossible to hold speakers responsible for their utterances at all, unless they can be held responsible on account of hearer assumptions. An implication of this view is that the distinction between lying and misleading cannot be upheld. Various researchers have designed experiments either to prove the robustness of the distinction or to empirically disconfirm it. Against these experimental findings, I stressed that such confirmation or disconfirmation is irrelevant, in so far as it tells us only that some people are in the habit of making the distinction, while others are prepared to consider false implicatures to be lies. These findings do not tell us anything about the justification for making or not making the lying/misleading distinction. The question which ought to be at issue is whether there are good reasons for making the distinction. In the last section I suggested that the lying/misleading distinction is morally problematic anyhow. In short, I cannot see that there are any conceptually, empirically or morally good reasons for letting speakers get away with falsehoods which they do not explicitly state.[3]

References

Antomo, M., Müller, S., Katharina, P., Paluch, M., and Thalmann, M. 2018, "When Children Aren't More Logical Than Adults: An Empirical Investigation of Lying by Falsely Implicating", *Journal of Pragmatics*, 138, 135-48.

Arico, A.J. and Fallis, D. 2013, "Lies, Damned Lies, and Statistics: An Empirical Investigation of the Concept of Lying", *Philosophical Psychology*, 26, 6, 790-816.

Augustine. 1952, *Lying*, tr. M.S. Muldowney, in *Treatises on Various Subjects*, Washington: Catholic University of America Press, 45-110 (original work published 395).

Borg, E. 2019, "Explanatory Roles for Minimal Content", *Noûs*, 53, 3, 513-39.

[3] Thanks to an anonymous referee for comments which have considerably improved the paper.

Camp, E. 2006, "Contextualism, Metaphor, and What is Said", *Mind & Language*, 21, 3, 280-309.

Camp, E. 2008, "Showing, Telling, and Seeing: Metaphor and 'Poetic' Language", *The Baltic International Yearbook of Cognition, Logic and Communication*, 3, 1-24.

Camp, E. 2018, "Insinuation, Common Ground, and the Conversational Record", in Fogal, D. *et al*. (eds.), *New Work on Speech Acts*, Oxford: Oxford University Press, 40-66.

Carson, T.L. 2010, *Lying and Deception: Theory and Practice*, Oxford: Oxford University Press.

Davidson, D. 1984, "Communication and Convention", in his *Inquiries into Truth and Interpretation*, 2nd ed., Oxford: Clarendon Press, 2001, 265-80.

Doyle, A.C. 2006, *The Hound of the Baskervilles*, in *The New Annotated Sherlock Holmes*, vol. III, ed. L.S. Klinger, New York: Norton, 383-627 (original work published 1902).

Frege, G. 1972, *Conceptual Notation*, tr. T.W. Bynum, Oxford: Clarendon Press (original work published 1879).

Fricker, E. 2012, "Stating and Insinuating", *Proceedings of the Aristotelian Society*, 86, 1, 61-94.

García-Carpintero, M. 2018, "Sneaky Assertions", *Philosophical Perspectives*, 32, 1, 188-218.

Goldberg, S. 2015, *Assertion*, Oxford: Oxford University Press.

Green, M. 2015, "Speech Acts", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/archives/sum2015/entries/speech-acts/.

Harris, D.W. 2016, "Intentionalism versus the New Conventionalism", *Croatian Journal of Philosophy*, 16, 47, 173-201.

Hobbes, T. 2000, *On the Citizen*, tr. R. Tuck and M. Silverthorne, Cambridge: Cambridge University Press (original work published 1641).

Jonsen, A.R. and Stephen, T. 1988, *The Abuse of Casuistry: A History of Moral Reasoning*, Berkeley: University of California Press.

Kant, I. 1999, "On a Supposed Right to Lie from Philanthropy", in *Practical Philosophy*, tr. M.J. Gregor, Cambridge: Cambridge University Press, 1999, 605-15 (original work published 1797).

Kaplan, D. 1989, "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", in Almog, J. *et al.*, (eds.), *Themes from Kaplan*, New York: Oxford University Press, 481-563 (original work published 1977).

Kölbel, M. 2011, "Conversational Score, Assertion, and Testimony", in Brown, J. and Cappelen, H. (eds.), *Assertion: New Philosophical Essays,* Oxford: Oxford University Press, 49-78.

Meibauer, J. 2011, "On Lying: Intentionality, Implicature, and Imprecision", *Intercultural Pragmatics*, 8, 2, 277-92.

Michaelson, E. 2016, "The Lying Test", *Mind & Language*, 31, 4, 470-99.

Neale, S. 2016, "Silent Reference", in Ostertag, G. (ed.), *Meanings and Other Things: Themes from the Work of Stephen Schiffer*, Oxford: Oxford University Press, 229-342.

Nietzsche, F. 2009, *On the Genealogy of Morality*, tr. C. Diethe, Cambridge: Cambridge University Press (original work published 1887).

Or, S., Ariel, M., and Orna, P. 2017, "The Case of Literally True Propositions with False Implicatures", in Chiluwa, I. (ed.), *Deception and Deceptive Communication: Motivations, Recognition Techniques and Behavioral Control*, Hauppauge: Nova Science, 67-107.

Pascal, B. 1866, *The Provincial Letters*, tr. T. M'Crie, New York: Hurd & Houghton (original work published 1656).

Rawls, J. 1971, *A Theory of Justice*, Cambridge, MA: Harvard University Press.

Reins, L.M. and Wiegmann, A. 2021, "Is Lying Bound to Commitment? Empirically Investigating Deceptive Presuppositions, Implicatures, and Actions", *Cognitive Science*, 45, 1-35.

Saul, J.M. 2012, *Lying, Misleading, and What is Said: An Exploration in Philosophy of Language and in Ethics*, Oxford: Oxford University Press.

Skoczén, I. 2021, "Modelling Perjuring: Between Trust and Blame", *International Journal for the Semiotic of Law*, https://doi.org/10.1007/s11196-021-09818-w.

Soames, S. 2008, "Drawing the Line Between Meaning and Implicature—and Relating Both to Assertion", *Noûs*, 42, 3, 440-65.

Sternau, M., Ariel, M., Giora, R., and Fein, O. 2016, "A Graded Strength for Privileged Interactional Interpretations", in Allan, K. et al., (eds.), *Pragmemes and Theories of Language Use*, Cham: Springer, 703-23.

Stokke, A. 2018, *Lying and Insincerity*, Oxford: Oxford University Press.

Timmermann, F., and Viebahn, E. 2021, "To Lie or to Mislead?", *Philosophical Studies*, 178, 1481-1501.

Viebahn, E. 2017, "Non-literal Lies", *Erkenntnis*, 82, 6, 1367-80.

Viebahn, E. 2020, "Lying with Presuppositions", *Noûs*, 54, 3, 731-51.

Viebahn, E. 2021, "The Lying-Misleading Distinction: A Commitment-Based Approach", *The Journal of Philosophy*, 118, 6, 288-319.

Wallace, R.J. 1994, *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.

Weissman, B. and Terkourafi, M. 2019, "Are False Implicatures Lies? An Empirical Investigation", *Mind & Language*, 34, 2, 221-46.

Wiegmann, A. and Willemsen, P. 2017, "How the Truth Can Make a Great Lie: An Empirical Investigation of Lying by Falsely Implicating", in Gunzelmann, G. *et al*. (eds.), *CogSci 2017: Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, London, 3516-21.

Wiegmann, A., Willemsen, P., and Meibauer, J., 2021, "Lying, Deceptive Implicatures, and Commitment", *Ergo*, https://doi.org/10.31234/osf.io/n96eb.

Dangerous Liaisons: The Pragmatics of Sexual Negotiation

Claudia Bianchi

Università Vita-Salute San Raffaele, Milan

Abstract

The debate about speech acts in sexual contexts has been dominated by discussions of *consent* and *refusal*, two illocutions strictly connected to definitions of sexual assault and rape, which constitutes a crucial step in fighting male sexual violence against women. Many authors have recently claimed that this emphasis has a distorting and harmful impact on our understanding of sexual communication—for it highlights only its negative aspects (mostly how to avoid unwanted sex). Moreover, an account in terms of consent and refusal seems to presuppose a default *asymmetrical* scenario, with men actively requesting sexual activities and women passively consenting or refusing. The aim of my paper is to assess the different speech-act accounts modelling communication in sexual contexts. I will first summarize the philosophical discussion on consent and refusal in sexual contexts and underline its connections with the debate on hate speech. I will then explore the model of initiations of sex in terms of requests and requests for permission, and analyse the *asymmetry* and *benefit* objections. I will present the models in terms of invitations, gift offers, and proposals, advocated by Kukla 2018, Gardner 2018 and Caponetto 2021b for their *collaborative* nature: invitations and proposals are illocutions presenting the sexual activity as beneficial for both parties and framing sex as a joint activity. My main goal is to criticize such Collaborative Models: I will show that conceiving of initiations of sex in terms of invitations, offers and proposals does not remove but rather actually *masks* the asymmetry.

Keywords: Consent, Refusal, Speech acts, Sexual contexts, Requests, Invitations, Proposals.

1. Introduction

This paper is devoted to a particular conversational context, the context of sexual negotiation, and to the asymmetries and distortions characterising it.[1] This analysis has not only theoretical significance for the contextual dependence debate,

[1] I am interested in asymmetries and distortions that are structured by gender hierarchy, and so I will talk about sexual communication in heterosexual scenarios. Also, I will talk

Argumenta 8, 1 (2022): 137-151
ISSN 2465-2334

DOI 10.14275/2465-2334/202215.bia

but also legal and social import, and a close connection to philosophical discussions on hate speech. The label "sexual negotiation" has no romantic connotation, signalling that we are dealing with non-ideal—indeed *dangerous*—relationships.

As a matter of fact, the philosophical, legal and, generally speaking, public debate about speech acts in sexual contexts has been dominated by discussions of *consent* and *refusal*, two illocutions strictly connected to definitions of sexual assault and rape (at least as concerns the US legal system)—definitions which constitutes a crucial step in fighting male sexual violence against women. Many authors have recently claimed that the emphasis on consent and refusal may actually have a distorting and harmful impact on our understanding of sexual communication—for it highlights the negative aspects of sexual communication (mainly how to avoid undesired sexual activities), while sexual negotiation aims not only to prevent harm, but also to enhance sexual agency. Moreover, an account in terms of consent and refusal seems to presuppose a default asymmetrical scenario, with men actively *requesting* sexual activities and women passively consenting or refusing. The aim of my paper is to assess the different speech-act accounts modelling communication in sexual contexts.

The paper unfolds as follows: I will first briefly summarize the philosophical debate on consent and refusal in sexual contexts (section 2) and underline its connections with the debate on hate speech (section 3). In section 4 I will explore the model of initiations of sex in terms of requests and requests for permission, and analyse the *asymmetry* and *benefit* objections. In section 5, I will present the models in terms of invitations, gift offers, and proposals advocated by Kukla 2018, Gardner 2018 and Caponetto 2022 in virtue of their *collaborative* nature: invitations and proposals are illocutions presenting the sexual activity as beneficial for both parties and framing sex as a joint activity. My main goal is to criticize such Collaborative Models: in section 6, I will show that conceiving of initiations of sex in terms of invitations, offers and proposals does not remove, but rather actually *masks* the asymmetry. The examination of different speech-act models will help reveal key features of the inquiry on communication in sexual contexts itself, and I will make these connections explicit in the final section.

2. Consent and Refusal

When discussing communication in sexual contexts, legal scholars, philosophers and linguists focus almost exclusively on the notions of *consent* and *refusal* for their close connection to definitions of sexual assault and rape. Consent and refusal are indeed key concepts in the US legal debate, both in the so-called *No Model* and *Yes Model*. In the *No Model,* rape is defined as a man obtaining sexual intercourse with a woman who physically resisted, or verbally expressed her refusal (Estrich

about rape and sexual assault where the perpetrator is a man and the victim is a woman. I recognize that men too may be victims of sexual violence, especially queer men, or kids regardless of their sexual orientation. However, since the great majority of rape victims and survivors are women and the great majority of rapists are men, and since the definition of consent and refusal constitutes an important part of the fight against male violence against women, the debate—even when presented in terms of "sexual communication"—has a strong tendency to focus on straight sex. For simplicity, I will omit the qualifications about other kinds of scenarios. I am grateful to Bianca Cepollaro for pressing me to clarify this point.

1987).[2] An unwelcome consequence of this model is that *silence* (i.e. the absence of verbal or nonverbal refusal) equals consent. As underlined by legal scholar Michelle Anderson, it is deeply misguided to consider silence as evidence of consent: "people sometimes experience peritraumatic dissociation and paralysis when confronted with sexual aggression, which causes silence and stillness but does not suggest agreement".[3] In the *Yes Model,* rape is defined as a man obtaining sexual intercourse with a woman without her affirmative consent.[4] The *Yes Model* maintains that silence by itself does not equal consent but continues to imply that an individual may nonverbally consent: "If she doesn't say 'no', and if her silence is combined with passionate kissing, hugging, and sexual touching, it is usually sensible to infer actual willingness" (Schulhofer 1998: 272-73). The *Yes Model*, then, still relies on men's ability to infer consent from women's body language, and still contends that prior intimacy ("passionate kissing, hugging, and sexual touching") is evidence of consent. Again, this may be highly problematic, for a) men tend to overestimate the extent to which women's nonverbal behaviour is evidence of sexual intent, and b) "people substitute other intimacy for penetration in order to avoid the health risks associated with it, so prior instances of intimacy cannot be interpreted to mean agreement to penetrative acts" (Anderson 2010: 83-4). In Anderson's words "When things heat up, then, the *Yes Model* melts into the *No Model*, in which silence constitutes consent".[5]

A clear definition of consent to sexual activities is sorely needed, as indicated by the alarming statistics on US colleges. Sexual violence on campus is indeed pervasive: according to the US's largest anti-sexual violence organization, RAINN (Rape, Abuse & Incest National Network), among undergraduate students, 26.4% of females (and 6.8% of males) experience rape or sexual assault.[6]

Unsurprisingly, the definition of consent plays a pivotal role in University conduct codes. To give an example, Yale Sexual Misconduct Policies require, for any sexual activity, *affirmative consent*, which is defined as "positive, unambiguous, and voluntary agreement to engage in specific sexual activity throughout a sexual encounter. Consent cannot be inferred merely from the absence of a 'no'. A clear 'yes', verbal or otherwise, is necessary". Moreover, consent to intercourse cannot be inferred from contextual factors such as clothing, alcohol or drug con-

[2] See Estrich 1987: 102: "'Consent' should be defined so that no means no".
[3] Anderson 2010: 83, see also Möller *et al.* 2017.
[4] See Schulhofer 1998: 283: "[Consent means] actual words or conduct indicating affirmative, freely given permission to the act of sexual penetration".
[5] Anderson 2005: 105. On the interpretation of nonverbal behaviour as evidence of sexual intent, see Anderson 2005: 117 (and works cited there): "A well-developed body of social psychology literature documents that men interpret women's body language as indicative of sexual intent when women have no such intent […] Men are more likely to misinterpret a woman's consumption of alcohol as conveying sexual intent. Men misinterpret women's friendly body language as indicative of sexual intent. When assessing interpersonal distance, eye contact, and casual touch, men rate women as more seductive and more promiscuous than women rate other women and themselves. Men are more prone to interpret flirting as indicative of sexual intent, whereas women tend to view flirting as 'relational development'. In short, the literature documents the male tendency to see female sexual consent where there is none".
[6] RAINN uses as its primary data source the National Crime Victimization Survey (NCVS), an annual study conducted by the US Justice Department: https://www.rainn.org/.

sumption, flirting or engaging in some form of intimacy: "Consent can be accurately gauged only through direct communication about the decision to engage in sexual activity".[7]

3. Consent, Refusal and Hate Speech

The debate on consent and refusal in sexual contexts holds a close connection with the debate on hate speech in philosophy of language, ethics and political philosophy. The general idea is that some sexist materials (paradigmatically pornographic materials, but also mainstream films, TV shows, romance novels and advertising) help to propagate false beliefs about men and women in sexual contexts, and to foster harmful attitudes and prejudices. According to feminist philosophers, such materials *silence* women by interfering with their ability to perform a range of speech acts, most notably sexual refusals (Langton 1993, Hornsby and Langton 1998). In particular, pornographic materials create a distorted communicative environment reinforcing dangerous gender stereotypes, such as "women always want sex; they enjoy violent, abusive sex, they fantasize about rape; rape is normal or legitimate. Hence the utterance of 'no' and similar locutions is not taken by a man to be a refusal but instead to be a part of the game".[8]

By representing women as only apparently declining sexual proposals while in fact longing for and intending to accept them, this kind of sexist material actually legitimizes men in persisting in their advances and disregarding women's refusals. Sexist stereotypes dictate, for instance, that the act of refusal fails (that is, it does not take effect as a binding refusal) if it is performed by women disregarding bigoted gender expectations—with violations such as going out alone at night, wearing short skirts or tight jeans, drinking too much, flirting or engaging in some form of sexual activity.[9]

At least four different kinds of failure of the acts of refusal of sexual advances may be identified.

1. *Uptake* failure: a man imbued with sexist stereotypes may fail to recognize the illocutionary force of the woman's refusal.

[7] https://smr.yale.edu/find-policies-information/yale-sexual-misconduct-policies-and-related-definitions. Further conduct rules state that "Consent to some sexual acts does not constitute consent to others, nor does past consent to a given act constitute present or future consent. Consent must be ongoing throughout a sexual encounter and can be revoked by any participant at any time. Consent cannot be obtained by threat, coercion, or force. Agreement under such circumstances does not constitute consent. Consent cannot be obtained from someone who is asleep or otherwise mentally or physically incapacitated due to alcohol, drugs, or some other condition". As Michelle Anderson aptly remarks, "AIDS killed the romance of uncommunicative sex twenty years ago" (2005: 136).

[8] Bird 2002: 6; for a similar characterization, see Maitra 2004: 192: "women always want sex, but also […] they tend to be coy in response to sexual overtures […] they try not to appear promiscuous, or overly sexually forward". It is notoriously difficult to define pornography: the authors working on silencing usually refer to a subset of pornography that "presents, endorses and eroticizes a hierarchical sexual relationship": McGowan 2017: 41. On this point, see MacKinnon 1987.

[9] For beliefs and stereotypes on sexual consent in the UK, see the 2018 Report of the *End Violence against Women* Coalition: https://yougov.co.uk/topics/resources/articles-reports/2018/12/01/publics-attitudes-sexual-consent.

2. *Sincerity* failure: a man may fail to recognize that the woman is sincere in her refusal: her act is taken as a form of teasing, part of a sexual game.
3. *Authority* failure: a man may fail to realize that the woman has the requisite authority to refuse—for he falsely believes that women have no authority over their own body (over which fathers, brothers or husbands do have authority).
4. *True feelings* failure: a man may fail to recognize that the woman is refusing if he falsely believes that the (sincere and authoritative) refusal does not accurately reflect the woman's true feelings.[10]

In this sense, pornographic materials endorsing and eroticizing hierarchical sexual relationships may be considered forms of hate speech, for they ultimately contribute to legitimizing sexual assault and rape.

4. Consent and Refusal in Sexual Contexts

Many authors have recently claimed that good-quality sexual negotiation requires that we do much more with language than request, consent to, and refuse sex. Additionally, the emphasis on consent and refusal has a distorting and harmful impact on our understanding of sexual communication. In their 2018 paper in *Ethics*, Quill Kukla (writing as Rebecca Kukla) maintains that the focus on consent and refusal is too narrow. People use language not only to decide whether or not they will have sex, but also "what kind of sex [they] are going to have, involving which activities, what [they] like and don't like, what [their] limits and constraints are, and when [they] want to stop" (Kukla 2018: 70). Moreover, the focus on consent and refusal tends to emphasize the negative aspects of sexual communication (mostly how to *avoid* undesired sexual activities), while sexual negotiation aims to not only prevent harm, but also enhance sexual agency.

With these provisos in mind, we now turn to an examination of the initial moments of a sexual negotiation. Consent and refusal are replies to some illocution previously performed by the speaker.[11] But to which kind of illocution? Let us examine the main accounts of the speech acts that consent and refusal are a reply to, starting with the Request Model. We will see that the Request Model is deemed problematic for two different reasons: both requests and requests for permission

1. involve one-sided activities, to be pursued by the H alone (in the Request Model) or by the S alone (in the Request for permission Model) (the *asymmetry objection*);
2. present the activity as beneficial for the S and costly for the H (the *benefit objection*).

4.1 Requests: The Asymmetry Objection

Consent and refusal are typically interpreted as a reply to exercitive or directive illocutions or, more precisely, to acts of either *requests* or *requests of permission*. When S performs a request, S asks for an action on the *hearer*'s part, while when

[10] See McGowan 2009, and 2017: 45-50; Caponetto 2021a. For a detailed analysis, Bianchi 2021a, ch. 2.
[11] Cf. Kukla 2018 and Caponetto 2017.

S performs a request of permission, S asks for permission relative to an action on the *speaker*'s part.[12]

Many scholars find it objectionable to frame the initial moments of a sexual negotiation in terms of exercitives or directives, for they seem to presuppose a default *asymmetrical* scenario, where one individual asks for a sexual activity (or asks for permission to perform a sexual activity) and the other individual either consents or refuses to perform the sexual activity (or to let the speaker perform the sexual activity upon them). This *asymmetry objection* has been raised by, among others, Anderson (2005: 108: "Consent […] is permission to be acted upon in some way. By itself, it suggests a passive response to the actions of another"); MacKinnon (2016: 440: "Consent as a concept describes a disparate interaction between two parties: active A initiates, passive B acquiesces in or yields to A's initiatives"); and Gardner (2018: 58: "by consenting, one is placing oneself in the position of patient and the other in the position of agent, so far as what is consented to is concerned. From there, one can quickly see that the question 'was there consent?' presupposes an asymmetry of exactly the kind that […] is not to be found in good (teamwork) sex").

What's more, this asymmetrical scenario combines with a cultural aspect where, typically, it is men actively asking and women passively consenting or refusing.[13] This default scenario tends to reinforce the stereotype of the active man and the passive woman—a stereotype that is part of the ideology legal scholars and philosophers are trying to challenge.[14] As Kukla summarises the objection, "All these authors presume a default scenario in which men want sex, women want to refuse sex, and refusal is, for one reason or another, pragmatically difficult" (Kukla 2018: 78).

4.2 Requests: The Benefit Objection

The Request Model faces another problem—one connected to the pragmatic structure of requests. Not only are they attempts to commit the H to doing something, but they also frame the desired action as being to the *advantage of S* and at the cost of H.[15] A similar objection may be raised as far as requests of permission are concerned: again, the action to be performed by S is framed as beneficial for

[12] Note, however, that for Leech, requests for permission are requests for an action *by H*, "namely the verbal act of giving of permission for S" to do something (2014: 142). For a detailed analysis of requests and requests of permission, see Caponetto *forth*. Caponetto observes that there is a substantial consensus on the idea that consents and refusals are replies to requests for permission—and quotes McGowan 2009: 489: "it seems intuitively clear (and perhaps even obvious) that refusals concern permission"; Cowart 2004: 514: "The act of giving your consent revolves around willingly giving permission to someone to do something that they do not have a right to do without asking for your permission»; Dougherty 2015: 226: "Consent is "morally valid" when, all else equal, it succeeds in generating a moral permission".

[13] See McGowan 2017: 44: "This consent model is problematic since it seems to presuppose that one person (typically a male) is the initiator or proposer of sexual activity and the other person (typically a female) accepts or declines that proposal".

[14] See also Gardner 2018: 68: "[the] overwhelming emphasis on consent—might have helped to reinforce the very ideology that the attempts were supposed to be challenging".

[15] Actually, things are far more complex than that: see Caponetto *forth*. for the hybrid nature (directive and commissive) of many illocutions.

S but costly for H. In both cases (requests and requests for permission), the illocution presents the desired activity as an activity to be performed without taking into account (or without acknowledging) H's desires.

What is more, requests for permission seem to presuppose a model of woman's sexuality as a piece of *property*, owned by one part and desired by another, and used with or without the owner's permission (cf. Du Toit 2008: 151; Du Toit 2009).

5. Collaborative Models

5.1 Invitations and Offers

In light of the objections raised against the Request Model, Kukla suggests a different, more *collaborative* model of initiations of sex in terms of invitations and gift offers. The general idea is that invitations, to a greater degree than requests, enhance our sexual agency because they are *welcoming*, rather than demanding, illocutions: they create a hospitable space for the invitee to enter and present the activity to be performed as beneficial to both S and H. Moreover, invitations leave the invitee free to accept or turn down the invitation, but at the same time they do not propose a neutral choice:

> Invitations open up the possibility of sex, and not just as a neutral possibility; the invitation makes clear that the one issuing it hopes for acceptance from the invitee. They are welcoming without being demanding. Accepting them is not a favor to the one issuing the invitation, as granting a request would be. Although we are generally pleased when people agree to have sex with us, we generally don't want people to agree to sex with us as a favor to us. While a rejection may well be disappointing, the inviter has no license to feel aggrieved if the invitation is turned down (Kukla 2018: 82).[16]

Kukla suggests an additional "ethical" model, conceiving initiations of sex in terms of *gift offers*: "generous offers of sexual gifts, designed first and foremost to please one's partner rather than to directly satisfy one's own sexual desires, are a normal part of an ongoing healthy relationship" (Kukla 2018: 86). According to Kukla, offering to have sex out of generosity rather than desire is an ethical and sensible option—at least as far as long-time partners are concerned.[17]

Invitations and gift offers are more ethical initiations of sex in virtue of their pragmatic structure, that is because they are "welcoming", "hospitable" and "generous" illocutions. Indeed, when performing an invitation, the speaker presents the action to be performed by H as being to the advantage of H (or of H and S) and at a cost to S—as when S invites H, say, to dinner and commits themself to arranging things for the meal, to H's expected benefit. Similarly, when S offers H a box of chocolates, or a flower bouquet, S is presenting the action as being at

[16] Note that in an Invitation Model, consent and refusal are no longer the appropriate replies: "One can't consent to an invitation—one accepts it or turns it down" (Kukla 2018: 82).

[17] Kukla 2018: 84: "Not all sex or all parts of sex have to be enthusiastically desired by all parties in order to be ethical and worthwhile". Interestingly, this last opinion runs contrary to many sex education courses, requiring "undivided enthusiasm on everyone's part as an ethical precondition of sexual activity": Kukla 2018: 84. See, as an example, the video series "Consent 101", created by Planned Parenthood: https://www.youtube.com/watch?v=qNN3nAevQKY.

a cost to themself and, in principle, to the advantage of H: "A gift must be designed to please the recipient. It might not actually succeed in pleasing, but an offer of something that is not expected to please is not in fact a gift" (Kukla 2018: 85).

5.2 Proposals

Two objections may be raised against the Request Model:

1. the *asymmetry objection*: requests and requests for permission involve one-sided activities;
2. the *benefit objection:* requests and requests for permission present the activity as beneficial for the S and costly for the H.

The Invitation Model and the Offer Model are less problematic, at least as far as the *benefit objection* is concerned, for the cost/benefit configuration is reversed. Indeed, when inviting H to a certain activity, the speaker S presents the action to be performed by H as to the advantage (also) of H and as at a cost to S. Likewise, when offering H a certain activity, the speaker S presents the action to be performed by S as to the advantage (also) of H and at a cost to S.

However, as Caponetto has pointed out, neither the Invitation Model nor the Offer Model can overcome the *asymmetry objection*: they both seem to involve one-sided activities: "Offers (e.g. "I can lend you some money, if you wish") are promises that the speaker commits to keep on condition that the hearer accepts. […] The bringing about of the state of affairs at stake (e.g. […] lending money) requires the active contribution of one party ([…] the offerer), but not necessarily a contribution from the other" (Caponetto 2021b).[18] A similar observation goes for invitations. True, the activity to which S invites H may involve some agency on H's part, but there is still an asymmetry between the two agents: "although an invitation, once accepted, calls for action on both sides (the inviter will have to throw the party and the invitee will have to show up), the details of the event are appanage of the inviter" (Caponetto 2021b).[19] Again, this asymmetry combines with a stereotypical scenario where an active individual (typically a man) not only initiates but also sets up the details of the sexual activity, and a passive individual (typically a woman) either accepts or declines the invitation to (or the offer of) a sexual activity.

Hence, Caponetto suggests a *Proposal* Model of initiations of ("good, agency-enhancing") sex. In her model, what S proposes to H is a *joint activity*, with both S and H in charge:

> A genuine proposal is an attempt to get another person to take part in some joint, fully collaborative activity. Conceiving of sex as something initiated by a proposal means, I claim, conceiving of it as an agent-agent symmetrical activity (Caponetto 2021b).

[18] Note that Caponetto is also sceptical as far as objection 2. is concerned: according to her, just as it is intuitively wrong to say that one who approaches someone for sex is asking her to do something solely or primarily for one's own benefit; it seems utterly wrong too to say that, in making sexual advances, one is offering to do something that would please the other person but involve a cost for oneself (Caponetto, personal communication).

[19] According to Leech, an invitation "is an offer taking place in a hospitality frame; it means that S, in the role of host, offers to provide something nice for [H] in the role of guest" (Leech 2014: 180).

The activity proposed is not only symmetrical (overcoming objection 1.), but also presented by S as beneficial for both S and H (overcoming objection 2.). In this way, the Proposal Model emphasizes the *collaborative* nature of sex, that is it construes sex as an activity "that one does *with* the other person, and over which each partner has an equal say" (Caponetto 2021b).

John Gardner holds a similar opinion: the speech acts of consent and refusal belong to an individualistic framework, to an idea of sexual activity as something that an individual does not *with*, but rather *to* another individual. ("Good") sexual activity must be conceived as a joint activity or even as "teamwork": "There have to be three agents in the room at least: the me, the you and the we. The actions of the me and the you have to contribute constitutively to the actions of the we. In this situation, nothing is being done *to* anybody. What is done, including what is done constitutively by me or you, is now being done *with* somebody" (Gardner 2018: 56).

6. Collaborative Models: A Critical Assessment

Let's take stock. The Request Model faces two objections: the *asymmetry objection* (requests involve one-sided activities) and the *benefit objection* (requests present the activity as beneficial for the S and costly for the H). While requests are exercitives (or directives), namely acts designed to influence the H's behaviour for the S's benefit, regardless of, or even contrary to, H's own desires—invitations, offers and proposals have a *commissive* component, namely they are acts designed to commit the S to a certain course of action in principle for the H's benefit.[20] It is this very structure that supporters of Collaborative Models deem more appropriate to a "regulative ideal" of initiations of sex: when S initiates sex, S is suggesting an activity for the expected benefit of both S and H. In this sense they are inherently positive ("welcoming and generous") illocutions. The Invitation Model and the Offer Model, then, overcome the *benefit objection*, but not the *asymmetry objection*, for they involve one-sided activities to be performed or planned mainly by the S. As far as the asymmetry objection is concerned, the Proposal Model seems to fare better. Proposals present the suggested activity not only for the *benefit* of both S and H (overcoming the benefit objection), but also to be planned and performed by both parties together (overcoming the asymmetry objection).

In the remainder of this paper I will assess the Collaborative Models, and challenge the very idea of *inherently positive* illocutions.

6.1 The Cost Objection

The different pragmatic structure characterizing requests on the one hand, and invitations, offers and proposals on the other, is reflected in politeness theory. As requests are exercitives (or directives)—namely acts designed to influence the H's behaviour to the S's benefit—they are often conceived as *FTA*, face-threatening acts, inherently *negative* illocutions, for they are a threat to H's negative face (her

[20] Strictly speaking, also requests have a commissive dimension—i.e. they aim at directing the audience's conduct, but also commit the speaker to a future course of action: when I request you to do something, I am not only trying to make you do something, I am also committing myself to *let* you do something. I thank Laura Caponetto for pointing out this aspect.

desire to be independent and to have freedom of action and freedom from imposition). Invitations, offers and proposals, instead, have a commissive component: they are acts designed to commit the S to a certain course of action—in principle to the H's benefit. As such they are often conceived as *FEA*, face-enhancing acts, inherently *positive* illocutions, for they preserve H's positive face (her desire to be appreciated and approved of).[21]

It is obvious, however, that invitations, offers and proposals are at the same time a potential threat to H's negative face—for they menace her personal space and freedom of action. This explains why, for example, we ask for permission for performing an offer, as in

(1) May I buy you a drink?

or an invitation, as in

(2) May I invite you to dinner?

It would be odd to ask H for permission to do something to her *benefit*: with (1) and (2), S is showing awareness of the menacing aspect of his invitation or offer.

Moreover, invitations and offers are not only a threat to H's personal space and freedom of action: they are also a demonstration of *immodesty* on S's part—for S is presupposing that his invitation or his offer would be welcomed by H (cf. Leech 2014: 183). The same goes for proposals, where illocutions such as

(3) Let's start a musical duo

or

(4) Let's have sex

typically present the activity suggested as beneficial for both parties: again, such presupposition is a potential breach to the Modesty Maxim (Leech 2014: 94).

Invitations, offers and proposals, in this sense, far from being intrinsically FEA, are potentially FTA. They may become unwelcome, even *predatory* acts for they constitute i) potential impositions, intromissions, interferences with H's privacy (a threat to H's negative face) and ii) a violation of general strategies of politeness, requesting S to express or imply "meanings that associate […] an unfavourable value with what pertains to S" (Leech 2014: 90).

Indeed, abusive agents take advantage of social norms and politeness norms in order to achieve their goals. On the one hand, they exploit positive politeness norms to impose themselves, by presenting certain activities as beneficial (also) for the H. On the other, they exploit H's tendency to comply with politeness norms, according to which refusals are always dispreferred options (for they threaten the positive face of the interlocutor: cf. Brown and Levinson 1987). This is especially true for women—who, more than men, are socialized to be polite and compliant. A compelling example of the interaction between politeness norms and stereotypical gender expectations is provided by North American or Western European bars—*sexualized* places where women's polite refusals of men's offers and invitations often turn out to be problematic, as they are typically perceived as rude and frequently disregarded. When a man offers a woman a drink, or invites her to dinner, he frames offer or invitation as pleasurable *for both of them*: presenting the suggested activity as beneficial for both partners places the woman

[21] See Brown and Levinson 1987, Leech 2014. On *FEA*, see Kerbrat-Orecchioni 1997.

in the awkward position not only of avoiding an unwanted activity, but also of having to justify an "irrational" choice.[22]

Conceiving of initiations of sex in terms of invitations, offers and proposals (illocutions that present the sexual activity as beneficial for both parties) and of sex as a collaborative activity presupposes that the suggested sexual activity has not only the same *benefit* for both partners, but also an analogous *social cost* for men and women, boys and girls. However, the collectively defined value of what is being transacted (still) has an importantly different weight for the two parties, for the social expectations in the sexual domain are tenaciously gendered—in terms of reputation, emphasis on virginity or chastity, unwanted pregnancy, and exposure to sexually transmittable diseases. In this sense, conceiving of initiations of sex in terms of invitations, offers and proposals does not eliminate but actually *conceals* the asymmetry. I will return to this point in my conclusion.

6.2 Real Life or Ideal Situations?

In this concluding section, I will comment on the general project we are engaged in when discussing initiations of sex. Are Kukla, Caponetto and Gardner dealing with the illocutionary acts actually performed by individuals approaching other individuals for sex in *real life* contexts, or are they rather dealing with the illocutionary acts that should be performed by individuals willing to initiate sex in *ideal* contexts—where sex is conceived in ethical terms, as an agency-enhancing activity?

While Gardner and Caponetto are quite explicit in outlining their enterprise as *normative*, Kukla is more ambiguous. It is unclear if Kukla's Invitation Model is a descriptive account ("invitations are a more common and typically more appropriate way of initiating sex than are requests": Kukla 2018: 82) or a normative account ("typical initiations of sex—particularly of agency-enhancing, ethical, good sex—are not requests or imperatives, but rather invitations and gift offers": Kukla 2018: 80-81).[23]

[22] Elinor Mason gives a nice analysis of such interaction in her *forth*.: 12-13: "A man sends over a drink, or offers to buy her a drink, and comes over to sit with her. She attempts refusal, politely but firmly. If she is lucky the encounter ends there, but that is rare. There is usually a period of negotiation, the man insists that he would like to get to know her, that it would be nice for both of them. She refuses again, politely, and he may stop there, or he may go on. One way the women can end the interaction is by giving a reason, by saying, 'I'm married', or, 'I am on a date'. But if the woman insists on politely saying no without giving a reason, very often, the man reacts as if the woman has been suddenly and inexplicably rude to him". Note, however, that Mason frames her explanation of such interactions in terms of *lack* of authority: "If the woman had robust authority here, she would not have to give justifying reasons" (Mason *forth*.: 13). A more suitable way, I believe, would be to frame the explanation in terms of authority *undermined*, *denial* of authority, lack of authority *recognition*, or else failure to *recognize* that the speaker has the authority to refuse. The choice is not only terminological, especially as far as sexual refusals are concerned: it matters for many victims of rape and sexual assault that they *did* have the authority to refuse, even if their authority was not acknowledged, and there was a failure of some sort (see Bianchi 2021b).

[23] Similarly, at times Kukla's objections to the Request Model are almost *empirical* in character: "Contrary to the consent model, requesting sex, while it is certainly something that we sometimes do, is not really the typical way we enter into sex, at least not when things are going well. (Requests along the way once sex is initiated are more common)" (Kukla

Moreover, the normative interpretation of Kukla's project is in tension with two claims, suggesting that Kukla has in mind non-ideal situations:

1) standard invitations and sexual invitations bear important differences;
2) the use of safe words should be extended as standard practice to "traditional" sexual negotiations.

1) According to Kukla there are important differences between standard invitations and sexual invitations: "One peculiarity of sexual invitations is that, unlike standard invitations, I do not owe you regret if I turn down your invitation. Another more important peculiarity is that I can back out of my acceptance of a sexual invitation at any time, for any reason at all" (Kukla 2018: 83). An individual may revoke their acceptance of a sexual activity at any time without any *normative residue*, such as justifications, excuses or expressions of regret.[24] Moreover, unlike standard invitations, just as I do not owe someone *regret* if I turn down their sexual invitation, similarly I do not owe someone *gratitude* for being invited to have sex with them. Kukla acknowledges that this is due to the non-ideal nature of so many sexual contexts, particularly as far as women are concerned:

> in our culture showing gratitude for a sexual invitation is often unacceptably risky, especially for women, because it carries with it all sorts of extra meanings and expectations and triggers various problematic social norms […] we live in a world filled with so many inappropriate sexual invitations, and so many men who refuse to take no for an answer if they sense any possible weakness or opening, that we often have good reason to forego showing gratitude, even if it is called for in some sense (Kukla 2018: 83-84).[25]

2) Safe words are discursive tools typical of the BDSM scene, designed to create a safe framework for sex: "they offer a tool for exiting an activity cleanly and clearly, with no real room for miscommunication […] and allow people to engage in activities, explore desires, and experience pleasures that would be too risky otherwise".[26] Kukla suggests extending the use of safe words as a standard practice to traditional ("vanilla") sexual encounters, especially when young and inexperienced people are involved—in order to help explore and understand their and their partners' desires, pleasures, boundaries and fantasies: "in my view it would be fantastic if the use of safe words became standard practice, and in particular if training on the use of safe words became a completely standard part of sex education for teens" (89).

2018: 80). At other times their objections are more *ethical* in character: requests don't seem to be the ideal way to initiate sex because they frame the action to be performed by H as an action that does not take into account (or does not acknowledge) H's desires.

[24] An individual may back out of their acceptance of a sexual activity at *any time* "including moments before we begin" (83): Kukla underlines that this pattern characterizes all invitations to participate in intimate bodily activities, such as invitations to donate an organ or gametes, gestate a child, or participate in medical research.

[25] According to Kerbrat-Orecchioni 1996: 83, the situations where it is appropriate to thank others allow us to identify what, in our society, is conceived as a gift (or, more generally, as a benign action). If we do not owe someone gratitude for some action—this action cannot be conceived as a gift or an offer.

[26] Kukla 2018: 88-89. BDSM is an acronym for "Bondage, Domination, Sadism, and Masochism": "It roughly refers to any consensual sexual practice involving the intentional infliction of pain or discomfort, restriction of motion, or asymmetric power play" (73n).

I argue that this proposal results from the non-ideal nature of many (traditional) sexual contexts. In ideal, ethical situations, all parties would be attuned to the others' desires and pleasures, and responsive to discursive cues, especially if highly conventional and standard, such as the use of "no" in order to exit from an activity at any time "without having to explain themselves or accusing anyone of transgression or any other kind of wrongdoing" (88). If, in a "traditional" sexual context (that is a sexual context with no role-playing coercion or domination and submission), we need to establish safe words in order to exit from an activity without pressure, coercion, or ambiguity, it is because we are navigating in non-ideal contexts, where collaboration and sensitivity to discursive cues are assumptions that interlocutors can no longer reasonably make—and where even standard language conventions are either no longer in place or lack their usual application. Only in non-ideal (and sometimes even strategic or conflictual) contexts could discursive tools such as "no", "I don't want to", "Stop it", "I don't feel we should continue anymore", and so on, be considered ambiguous, unclear, non-literal or in need of interpretation.

7. Conclusion

The philosophical and legal debate concerning the context of sexual communication is dominated by discussions of consent and refusal, focusing massively on how to avoid harm, and to prevent unwanted sexual activities, and rape. True, this casts a negative light on sexual negotiation, and obfuscates all kinds of illocutions occurring before, during and after sex and designed to communicate desires, boundaries and conditions to start, continue, and stop sex. Communication in sexual contexts undeniably has the power to enhance our sexual agency and autonomy, and to lead us to non-abusive—and arguably better, more pleasurable—sex. While non-abusive sexual interactions indeed require the recognition of the relevance of the desires of all parties involved, the Collaborative Models, *per se*, run the risk of masking prevarication and abuse. In these models, women's sexual agency is presupposed and presented as a *given*, while in real life cases its efficacy is limited in a variety of ways, all of which refer to the man's sexual preferences, benefits, and desires. Policies and laws (and much philosophical analysis) must too often deal with non-ideal, real-life contexts, where the parties involved are not necessarily engaged in ethical undertakings, with more or less the same goals, and more or less the same price to pay for the activities performed. Presupposing otherwise has the unwelcome consequence of camouflaging the different takes on sex that men and women, boys and girls still have—by favouring "hypocritical" models of initiations of sex over realistic ones. Realistic contexts include cases where women feel significant pressure to have sex with their partners and where men actively ask for activities presented (more sincerely) as beneficial for themselves. Sometimes the benefit for women (and more generally, for disadvantaged individuals) cannot be cast in terms of pleasure or personal flourishing, but in more mundane terms as ways to be included in a group, prevent conflict with the partner, or even avoid additional violence.[27]

[27] On unjust sex vs. abusive sex, see Cahill 2016. On sexual consent in *real life* as opposed to *ideal* contexts, see Garcia 2021.

I am deeply indebted to Laura Caponetto and Bianca Cepollaro for countless discussions and comments on this and related topics.

References

Anderson, M.J. 2005, "Negotiating Sex", *Southern California Law Review*, 41, 101-40.

Anderson, M.J. 2010, "Sex Education and Rape", *Michigan Journal of Gender and Law,* 17, 83-110.

Brown, P. and Levinson, S.C. 1987, *Politeness: Some Universals in Language Usage*, Cambridge: Cambridge University Press.

Bianchi, C. 2021a, *Hate speech: Il lato oscuro del linguaggio*, Roma-Bari: Laterza.

Bianchi, C. 2021b, "Discursive Injustice: The Role of Uptake", *Topoi*, 40, 181-90.

Bird, A. 2002, "Illocutionary Silencing", *Pacific Philosophical Quarterly*, 83, 1-15.

Caponetto, L. 2017, "On Silencing, Authority, and the Act of Refusal", *Rivista di estetica*, 64, 35-52.

Caponetto, L. 2021a, "A Comprehensive Definition of Illocutionary Silencing", *Topoi*, 40, 191-202.

Caponetto, L. 2021b, "Modeling Sex as a Joint Activity", *Open for Debate*, https://blogs.cardiff .ac.uk/openfordebate/modeling-sex-as-a-joint-activity/.

Caponetto, L. *forth*., "The Pragmatic Structure of Refusal", *Synthese*.

Cahill, A. 2016, "Unjust Sex vs. Rape", *Hypatia*, 31, 4, 746-61.

Cowart, M. 2004, "Understanding Acts of Consent: Using Speech Act Theory to Help Resolve Moral Dilemmas and Legal Disputes", *Law and Philosophy*, 23, 495-525.

Dougherty, T. 2015, "Yes Means Yes: Consent as Communication", *Philosophy and Public Affairs*, 43, 3, 224-53.

Du Toit, L. 2008, "The Contradictions of Consent in Rape Law", *South African Review of Sociology*, 39, 1, 140-55.

Du Toit, L. 2009, *A Philosophical Investigation of Rape. Making and Unmaking the Feminine Self*, New York: Routledge.

Estrich, S. 1987, *Real Rape*, Cambridge, MA: Harvard University Press.

Garcia, M. 2021, *La Conversation des Sexes: Philosophie du Consentement*, Paris: Flammarion.

Gardner, J. 2018, "The Opposite of Rape", *Oxford Journal of Legal Studies*, 38, 1, 48-70.

Hornsby, J. and Langton, R. 1998, "Free Speech and Illocution", *Legal Theory*, 4, 21-37.

Kerbrat-Orecchioni, C. 1996, *La Conversation*, Paris: Seuil.

Kerbrat-Orecchioni, C. 1997, "A Multilevel Approach in the Study of Talk-in-interaction", *Pragmatics,* 7, 1, 1-20.

Kukla, R. 2018, "That's What She Said: The Language of Sexual Negotiation", *Ethics*, 129, 1, 70-97.

Langton, R. 1993, "Speech Acts and Unspeakable Acts", *Philosophy and Public Affairs*, 22, 4, 292-330.

Leech, G. 2014, *The Pragmatics of Politeness*, Oxford: Oxford University Press.

MacKinnon, C. 1987, *Feminism Unmodified: Discourses on Life and Law*, Cambridge, MA: Harvard University Press.

MacKinnon, C. 2016, "Rape Redefined", *Harvard Law & Policy Review*, 10, 431-77.

Maitra, I. 2004, "Silence and Responsibility", *Philosophical Perspectives*, 18, 189-208.

Mason, E. *forth*., "Sexual Refusal: The Fragility of Women's Authority", *Hypatia*.

McGowan, M.K. 2009, "On Silencing and Sexual Refusal", *The Journal of Political Philosophy*, 17, 4, 487-94.

McGowan, M.K. 2017, "On Multiple Types of Silencing", in Mikkola, M. (ed.), *Beyond Speech: Pornography and Analytic Feminist Philosophy*, New York: Oxford University Press, 39-58.

Möller, A., Söndergaard, H.P., and Helström, L. 2017, "Tonic Immobility During Sexual Assault: A Common Reaction Predicting Post Traumatic Stress Disorder and Severe Depression", *Acta Obstetricia et Gynecologica Scandinavica*, 96, 8, 932-38.

Schulhofer, S. 1998, *Unwanted Sex: The Culture of Intimidation and the Failure of Law*, Cambridge, MA: Harvard University Press.

Advisory Board

The Debate on the Problem of For-Me-Ness: A Proposed Taxonomy

Alberto Barbieri

NETS – IUSS Pavia Center for Neurocognition, Epistemology and Theoretical Syntax

Abstract

Several philosophers claim that a mental state is phenomenally conscious only if it exhibits so-called *for-me-ness,* or *subjective character,* i.e., the fact that there is something it is like to be in a conscious state not just for everyone but only *for the subject who undergoes it*. Consequently, they stress, a proper explanation of consciousness requires to address the question of what the nature of for-me-ness is. This question forms what I call the problem of for-me-ness. Although the debate on the problem of for-me-ness has assumed a centre stage within philosophy of consciousness, relatively scant attention has been paid to systematize it. In this paper, I propose to fill this gap by developing a taxonomy of the existing responses to the problem at stake. I start by claiming that for-me-ness—the phenomenon to be explained—is best thought of as a minimal form of self-consciousness. Answering the problem of for-me-ness, hence, means to provide an account of the metaphysical structure of such a phenomenon. Next, I claim that the nature and the structure of minimal self-consciousness can be established by considering five conceptual distinctions. Based on such distinctions, finally, I classify the existing responses to the problem of for-me-ness into five positions. Scholars interested in the debate at issue should find this taxonomy useful not only to recognise and assess the core theses of the existing answer to the problem of for-me-ness but also to develop their own response.

Keywords: For-me-ness, Subjective character, Phenomenal consciousness, Minimal self-consciousness, Phenomenal character.

1. Introduction

Consciousness is the property mental states have when, and only when, they exhibit a so-called phenomenal character, namely, the fact that there is something it is like for their subject to be in these mental states (Nagel 1974). Accounting for the nature of this character forms the problem of consciousness—one of the most central issues in contemporary philosophy of mind.

An important first step in approaching such a problem is to pin down the constitutive features of the phenomenon in need of explanation, that is, the phenomenal character of experience. For many years, philosophers of mind have made this step by equating phenomenal character with the qualitative character

Argumenta 8, 1 (2022): 155-171
ISSN 2465-2334

DOI 10.14275/2465-2334/20210.bar

(or dimension) of experience, usually spelled out in terms of the characteristic qualitative properties, or *qualia*, that conscious mental states have and that unconscious states do not.

Recently, however, several philosophers have argued that this conception of the structure of phenomenal character is inadequate (cf. Kriegel 2009; Levine 2001; Zahavi 2014; Zahavi & Kriegel 2015, among others). Conscious mental states, it is claimed, are also essentially characterised by another aspect, corresponding to the fact that they affect their subject in a special way and that no one else is affected by them in the same way. In other words, conscious mental states are not just held by me, but they are also given to me in a first-personal way—they are like something *for me*. Such an aspect of conscious mental states is usually called the *for-me-ness*, or *subjective character*, of the experience, but terms such as *subjectivity* (Levine 2001), *me-ness* (Block 1995), and *first-personal givenness of experience* (Zahavi & Kriegel 2015) are sometimes employed to designate it. In this view, thus, phenomenal character necessarily has two constitutive features: qualitative character and for-me-ness. The former captures the changing aspect of a conscious mental state—that is, what makes a conscious state the conscious state it is—and the latter captures its invariant aspect—that is, what makes a mental state conscious at all (Kriegel 2009).

Putting aside concerns about the correctness of such a view, one significant implication of it has been that, in addition to the traditional issue of qualia, a new problem has emerged as pivotal for understanding consciousness and its nature. This is the problem of explaining the nature of the conscious mental states' being for their subjects. We might call this the *problem of for-me-ness*.

This paper focuses on the problem of for-me-ness. More precisely, this paper aims to develop a taxonomy of the existing responses to the problem at stake. Although the nature of for-me-ness has become a hot topic within the philosophy of consciousness, to date no substantive works have been devoted to systematising the debate that grew up around it. The relevant literature prevalently contains either attempts to address the problem of for-me-ness or defences of the phenomenon at its heart from several forms of scepticism. However, virtually no studies have aimed to take stock of what has become an extended, but increasingly murky, debate. Some recent analyses on the concept of for-me-ness and its cognates might be viewed in this regard (cf. Farrell & McClelland 2017; Guillot 2017; Howell & Thompson 2017), but they, in short, have questioned the existence of a real unified problem of for-me-ness and have suggested that different authors address different questions under the same label. As such, these studies do not really aim at mapping the debate at issue but rather at warning it, or better, criticising it.

I proceed as follows. In Section 2, I clarify what phenomenon is supposed to constitute (or ground) the for-me-ness of experience. In doing so, I provide a description of the phenomenon at issue that is both theoretically illuminating and as free of metaphysical assumptions as possible. I argue that for-me-ness, in its last committal form, is constituted by (or grounded in) a *minimal form of self-consciousness* (or self-awareness),[1] allowing me thus to equate the debate on for-me-ness with the discussion on minimal self-consciousness. In Section 3, then, I outline five distinctions of self-consciousness that are relevant for the phenomenon constituting for-me-ness, or so I suggest. Finally, in Section 4, after having argued that these distinctions are conceptually independent of each other, I classify the

[1] In what follows, I will consider self-consciousness and self-awareness as synonyms.

existing responses to the problem at issue based on the distinctions individuated in the previous section. The final result, it is hoped, will possess sufficient clarity to serve both as a good description of the current state of the debate in question and as a useful theoretical framework for further approaches to the problem of for-me-ness.

2. For-Me-Ness as Minimal Self-consciousness

A preliminary task to develop a taxonomy of the main responses to the problem of for-me-ness is to fix firmly on the phenomenon at its heart, that is, the for-me-ness of experience. Above I noticed that the notion of for-me-ness is intended to refer to the fact that conscious mental states are not just *held* by the subject but also *experientially given to the latter*. As Kriegel (2011) puts it, "when I have a conscious experience, the experience does not occur only *in me*, but also *for me*. There is some sort of direct presence, a subjective significance, of the experience to the subject" (Kriegel 2011: 53).

To a first approximation, this is already sufficient to get a good grasp of what for-me-ness amounts to. To a second approximation, a more precise description is needed. In fact, while characterising for-me-ness as the mental state's property of *being for the subject who undergoes it* is surely a suggestive—and arguably phenomenologically apt—way of introducing the phenomenon at stake; nevertheless, it maintains a certain degree of ambiguity and conceptual blurriness. It seems appropriate, then, to raise the following further question: What exactly does the phenomenally conscious state's *being for the subject* consist in?

One might answer by appealing to the conceptual division of labour between qualitative and subjective character and might claim that the latter consists in the substantive commonality among all phenomenal characters. But, still, a sense of vagueness lurks here. What is such a commonality? What phenomenon are philosophers pointing to when they talk about the component that remains constant in the experience? If we are to build a taxonomy of the problem of for-me-ness, an answer to such questions is required, which essentially means arriving at a clear and minimal description of what for-me-ness amounts to. As Levine (2016) effectively points out, before starting any theorising on for-me-ness, "it is important to clarify just what it is for a mental state to be subjective in this way" (Levine 2016: 342).

Let me emphasise that by "clear and minimal", I mean a description of for-me-ness that satisfies the following two conditions (or desiderata, if you prefer): first, advancing our theoretical understanding of the phenomenon at stake, which means offering a suitable answer to the questions raised above; second, avoiding incorporating too many substantial assumptions, which means to avoid providing a description that says something about the metaphysics of the phenomenon in need of explanation.[2]

Doing so risks leading to the arbitrariness and accusation of misconstrual, but what motivates me to set such conditions is the ambition to achieve a

[2] With this I do not want to claim that there are no "metaphysical" assumptions built into the notion of for-me-ness. I only claim that these assumptions are less demanding than an explanation of the ultimate nature of such phenomenon. In other words, this way of describing *for-me-ness* provides an answer to what Van Gulick (2018) calls the 'descriptive question'.

characterisation of for-me-ness similar to that of qualitative character. As I have noted above, the latter is typically spelled out in terms of the various qualitative properties (or *qualia*) involved in the experience. As far as I can see, this way of characterising qualitative character is both something that advances our theoretical understanding of the phenomenon under investigation and something that does not incorporate (at least too much) metaphysical assumptions. If, on the one hand, it claims that qualitative character is constituted by (or grounded in) the qualitative properties of conscious mental states, on the other hand, it is silent on the nature of these properties. Put differently, such a characterisation sheds light on the phenomenon without hiding an answer to what—in the introduction—I called the problem of qualitative character. Similarly, I take a description of for-me-ness ought to do the same theoretical work.

A promising starting point to arrive at this characterisation is to analyse the nature of the property picked out by the term 'for-me-ness'. Unlike the qualitative character—which seems to pick out a non-relational property of mental states[3]—the latter, as suggested by the term itself, seems to be a relational property: it is the mental state's property of *being given to the subject*. Such property, hence, pinpoints a relation between the subject and one of their mental states. Since, arguably, a relational property is grounded in the relation it picks out, it follows that the mental state's being for me essentially involves the subject bearing a certain relation with their mental state.

What is the nature of such a relation? Interestingly, a common pattern exists that can be recognised in the literature. Consider, for instance, the following passages:

> If there is something that makes a conscious experience "for me", then by having the experience, I must be *somehow aware of having it*. For if I am wholly unaware of my experience, there is no sense in which it could be said to be "for me" (Kriegel 2005: 25, emphasis added).
>
> To say that there is something it is like for a creature to be in a certain mental state is to say that the creature is implicitly *aware of herself being in that mental state* (Janzen 2008: 156, emphasis added).
>
> [The for-me-ness of experience] concerns the fact that when one is directly and noninferentially conscious of one's own occurrent thoughts, perceptions, or pains, they are characterized by a first-personal givenness that immediately reveals them *as one's own* (Zahavi 2005: 26, emphasis added).

When abstracted from the details,[4] Kriegel (2005), Janzen (2008), and Zahavi (2005) agree in claiming that the relation at stake in the for-me-ness of the experience is an epistemic relation in which the subject is somehow aware of the experience. In other words, the idea seems to be that phenomenally conscious mental

[3] In describing qualitative character as intrinsic, I only mean to claim that this property is *prima facie* intrinsic, that is, its surface grammar suggests an intrinsic modification of mental states. However, it might well be that—*ultima facie*—qualitative character is relational.

[4] To be clear, I am not saying that there are not important differences. When one embarks in the project of explaining the nature of the phenomenon in question, such fine-grained "details" play a key role; however, for our purposes here, we can set them aside.

states are not (just) states which give us awareness of the environment around us[5] but (most of all) states we *are somehow aware of*. And the for-me-ness picks out this awareness exactly.

My contention, therefore, is that, by introducing the for-me-ness of experience, philosophers are pointing to a special *awareness* which—in being not directed outward, to the world and its features, but inward, to something that concerns the mental life itself—can be suitably regarded as a kind of *inner awareness* or *self-awareness*. Moreover, since it accompanies all possible experiences—even in the simplest ones and in the simplest conscious entities—it is reasonable to claim that such self-awareness is the most fundamental or minimal form, namely the one that grounds higher forms. The following is a somewhat loose formalisation of the idea:

MSA: For any mental state M of a subject S, such that S is in M, M has for-me-ness if, and only if, S is (suitably) aware of M.

Call this the *minimal self-awareness principle.*[6] [7]

MSA is meant to capture what for-me-ness is, what the "being for me" of mental states picks out at the phenomenological level. The task of subjectivists, then, is to explain what the nature of such a phenomenon is and how it is instantiated in us. In other words, MSA fixes the data that proponents of for-me-ness aim to explain.

A qualification about the kind of self-awareness at stake here is required. The expression "most fundamental or minimal form" does not endorse any specific thesis about the nature of such a form of self-consciousness. Indeed, one can claim that minimal self-consciousness is a sophisticated cognitive phenomenon that is available months after birth. By contrast, another could maintain that it does not demand any cognitive achievement, but it is relatively simple. Whether these are correct accounts of the phenomenon, they are not *definitional* of "minimal self-consciousness", but rather they are substantive claims about it. What is definitional of "minimal self-consciousness" is that it is the most fundamental, or minimal, occurrence of self-consciousness.

If what I have said here is on the right track, it follows that for-me-ness in its least committal way is characterizable as the minimal form of self-awareness. Consequently, the problem of for-me-ness can be refined as follows:

The Problem of for-me-ness: What is the (metaphysical) nature of minimal self-consciousness, which constitutes (or grounds) the phenomenally conscious state's being for me?

[5] This is the concept of state consciousness endorsed by those philosophers who consider phenomenal character wholly constituted by the qualitative character. To borrow from Drestke (1993), the idea is that phenomenal conscious states are states we "are conscious *with*" (Drestke 1993: 281).

[6] MSA goes by different names in the literature, such as "inner awareness thesis" (McClelland 2015), "awareness principle" (Kriegel 2018), and "transitivity principle" (Rosenthal 2005). I think, however, that these labels incorporate (or at least risk of incorporating) some bold claims on the phenomenon to be explained.

[7] The "suitably" clause is just to remark that the inner awareness at stake here is of some peculiar form; that is, not every awareness of M is suitable to constitutes the for-me-ness of experience. One way to understand the debate on the nature of for-me-ness is precisely to see it as an attempt to unpack this clause (more on this attempt in Section 4).

To build an appropriate taxonomy of the problem in question, then, we need to understand which features mostly characterise the phenomenon of self-consciousness and, more specifically, the phenomenon of minimal self-consciousness. I do this in the next section.

3. Five Distinctions Characterising Minimal Self-Consciousness

What is the nature of the minimal form of self-consciousness? Is it a cognitively complex phenomenon that requires the acquisition and mastery of conceptual abilities, or is it a mental feature possessed since birth? Is it wholly determined by introspection, or is it present in every moment of our experiential life? Is it a property of a subject, or is its most fundamental occurrence explainable wholly in terms of a property (or relation) of mental states? These and other related questions characterise the problem of minimal self-consciousness, and answering them—for the reasons stated above—means addressing the problem of for-meness.

The literature on self-consciousness is as vast as the literature on consciousness. It thus makes sense to find some criteria to distinguish only theories with substantial differences, collecting under a common label those that diverge only in detail and terminological issues. To accomplish this task, I employ five distinctions that are commonly drawn within the self-consciousness debate and that are both independent and parallel to each other—as I hope will become clear by the end of the discussion. The distinctions are as follows:

1. Creature versus state self-consciousness
2. Egological versus non-egological self-consciousness
3. Pre-reflective versus reflective self-consciousness
4. Representational versus non-representational self-consciousness
5. Conceptual versus non-conceptual self-consciousness.

Within this framework, these distinctions are intended to identify the properties responsible for the substantial disagreement between the scholars involved in the debate about the nature of minimal self-consciousness.[8]

To be sure, additional distinctions can certainly be drawn, but I have restricted myself to these because I consider that they capture the salient features of the phenomenon at stake. Indeed, they concern the status of all three constituents of minimal self-consciousness: the two *relata* and their putative relation. In particular, (1) can be taken to be about the nature of the subject-relatum, (2) can be taken to be about the nature of the object-relatum, and the remaining three can be taken to be about the nature of the awareness relation.

Before introducing them, I would like to clarify some terminological and methodological choices. First, even if we are concerned with self-consciousness, some notions have been introduced and developed independently of this debate. For this reason, a notion is explained at first without referring to the self-consciousness where necessary.

[8] Worth noticing is that the existence of (at least some) properties putatively picked out by the distinctions in question has been variously contested. It seems thus appropriate to take these distinctions primarily as conceptual ones. However, nothing worrisome for our purposes follows from this, as theories I discuss in Section 4.2 are—at the very least—committed to the existence of the features they ascribe to minimal self-consciousness.

Second, in the case of most distinctions, I talk about "forms" of self-consciousness. Each distinction identifies two different characteristics of self-consciousness. It might therefore be natural to conclude that in choosing one component of the pair, one automatically rejects the existence of the other one. However, if we accept the idea that self-consciousness is a complex phenomenon manifested in different forms and to different degrees, the two features are not necessarily mutually exclusive; they can be features that relate to the different degrees of the phenomenon. For these reasons, I treat most of the distinctions as claims about the existence of forms of self-consciousness, leaving the discussion about minimal self-consciousness to the following section. With these clarifications in place, let us now consider the distinctions in detail.

3.1 A Distinction about the Subject-Relatum of Self-Consciousness

Let us begin with an important distinction concerning the types of entities that can be self-conscious, that is, the one between *creature* and *state self-consciousness*. That self-consciousness is a property borne by certain creatures or, more generally, certain subjects of experience seem uncontroversial. Which creatures are self-conscious is a matter of a vibrant debate both in the philosophical and the scientific field, but it does not undermine the unquestionable claim that subjects of experience—among which normal adult humans are paradigmatic cases—are a type of entity that can entertain self-consciousness.

However, are subjects of experience the *only* type of entity that can be self-conscious? Several philosophers of mind answer this question negatively. In particular, they claim that the relation of self-consciousness can also be borne by subject's mental states.[9] To clarify the claim, these authors appeal to the distinction between creature and state consciousness. It is widely held that consciousness is a property applied both to creatures, or subjects, and to particular subjects' mental states. Since self-consciousness is a kind of consciousness, the reasoning goes that such a distinction can also be applied to self-consciousness. The upshot, therefore, is that, when we talk of self-consciousness, we should distinguish between self-consciousness as a relation borne by the whole individual, the subject of experience, and as a relation borne by the subject's mental states.

To dispel some initial worries, however, some specifications are in order. First, most advocates of state self-consciousness do not seem to imply that mental states are self-conscious in the same way as creatures are self-conscious.[10] When it comes to being ascribed to mental states, the self-consciousness relation, therefore, tends to be cashed out more metaphorically. Second, those who advocate for state self-consciousness typically do not consider the latter entirely independent from creature self-consciousness,[11] but rather, they use it to explain creature self-consciousness. That is, they claim that a subject is self-conscious—as in the self-consciousness at stake in MSA—by virtue of the fact that one of their mental

[9] See, among others, Kriegel 2004: 3-4; Musholt 2015: introduction; Zahavi 2018: 5.
[10] If for a creature to be self-conscious, it is likely for her to be conscious either of herself or her mental states, what it is for a mental state to be self-conscious is probably different, since mental states, at least *prima facie*, are not the kind of entity that can be *conscious of* something.
[11] After all, MSA is formulated in terms of creature self-consciousness.

states is somehow self-conscious. Impressionistically put, the idea is that *state self-consciousness is the source of creature self-consciousness.*[12]

3.2 A Distinction about the Object-Relatum of Self-Consciousness

A second important distinction is between the *egological* and *non-egological* forms of self-consciousness. If the previous distinction focuses on who or what is self-conscious, namely who, or what, is the subject-relatum of the self-awareness relation, the egological/non-egological distinction focuses on what is presented in self-consciousness, that is, what is the object-relatum of the self-awareness relation.

Let me explicate the distinction by the use of an example. Compare "I am conscious of myself seeing a blue sky" to "I am conscious of seeing a blue sky". According to most philosophers involved in the debate, the awareness relations in the two sentences are both instances of self-consciousness; both are directed inwards to internal mental goings-on. However, they differ significantly with respect to what they are an awareness of. The latter involves merely an awareness of a particular mental state of mine, while the former also involves an awareness of myself and, to this extent, it is a stronger form of self-consciousness. To distinguish between those forms of self-consciousness, philosophers sometimes refer to the stronger variety as egological self-consciousness and the weaker as non-egological self-consciousness.[13]

The above example suggests that the egological/non-egological distinction applies just to creature self-consciousness. However, the distinction applies to state consciousness as well, as is shown by the common practice of using the egological and non-egological categories to refer to the latter.[14] To avoid confusion, we can regiment such a distinction by claiming that egological self-consciousness, in having its content partially constituted by the subject of experience, is a *subject* involving form of self-consciousness, while non-egological self-consciousness, in having its content fully constituted by a particular state that occurs within oneself, is a *mental-state* involving form of self-consciousness.

3.3 Three Distinctions About the Very Self-Consciousness Relation

Let us now turn our attention to some distinctions that do not focus (at least directly) on the two relata of self-awareness but on the very relation. For reasons of clarity, I present them using self-consciousness as a property of subjects. This choice is only for presentation purposes and nothing substantial follows from it. I begin with the distinction between *non-reflective* and *reflective* forms of self-

[12] Guillot (2017) highlights the same grounding claim that I point out here. According to her, several authors endorse what she calls "the state self-awareness" view of for-me-ness, according to which "subjective character, construed as the property mental states exhibit when their subject is aware of them, is constituted by a more primitive awareness those experiences have of themselves" (Guillot 2017: 7). However, I find Guillot's definition of the "state self-awareness" view too restrictive, insofar as it leaves out those theories of for-me-ness that appeal to an egological form of state self-awareness (more on this notion in Section 4.2).

[13] The terms "egological" and "non-egological" were originally introduced within the phenomenological tradition, but they have become very popular even in contemporary analytic philosophy of mind.

[14] For a review of the egological and non-egological theories of state self-consciousness, see Sebastián 2012.

consciousness,[15] or—which is typically considered to be conceptually equivalent—*intransitive* and *transitive* forms of self-consciousness.[16]

The non-reflective/reflective dichotomy applies primarily to consciousness as such, and it lies in the target of attention. Non-reflective consciousness is the kind of consciousness a subject maintains when their attention is directed outward. The focus of their conscious awareness, thus, is on the external world or, more generally, on objects that are not themselves parts of conscious experiences. By contrast, reflective consciousness is the kind of consciousness a subject maintains when their attention is directed inward by focusing on themselves and their mental states—what is usually called introspection.

There is no doubt that self-consciousness can be present in reflective consciousness. In fact, some would say that introspection is always a kind of self-consciousness. However, several philosophers claim that self-consciousness can also be present in non-reflective consciousness, that is, when the subject's attention is directed outward. Here, for instance, is Goldman:

> [Consider] the case of thinking about x or attending to x. In the process of thinking about x there is already an implicit awareness that one is thinking about x. […] When we are thinking about x, the mind is focused on x, not on our thinking of x. Nevertheless, the process of thinking about x carries with it a *non-reflective self-awareness* (Goldman 1970: 95, emphasis added).

Thus, self-consciousness, besides an introspective (or attentive) form, which involves our reflecting upon ourselves or ours own mental states, admits a *non-introspective* (or inattentive) variety, in which we do not entertain any reflective state but, rather, we are self-conscious in a more implicit manner. The distinction between reflective and non-reflective self-consciousness aims at keeping track of this fact.[17]

Another important distinction is between representational and non-representational forms of self-consciousness. Since this distinction applies primarily to consciousness rather than self-consciousness, it is expedient to first clarify it in relation to consciousness and, second, to understand its application to self-consciousness.

[15] In the relevant literature, actually, the concept of non-reflective self-consciousness often goes under the label of *pre-reflective* self-consciousness (cf. Miguens, Morando & Preyer 2016; Zahavi 2014, 2019). However, although this term reflects better the context of its introduction, it is often employed to deliver a reading of the distinction at issue that embeds the distinction between representational and non-representational self-consciousness. Here, as a descriptive context, I prefer to stick to a weaker and less theory-loaded reading of the distinction under discussion.

[16] This conceptual equivalence is acknowledged by philosophers of both traditions (cf. Zahavi 2004: 73; Kriegel 2004: 186).

[17] Some readers might question the relevance of this distinction for the nature of for-me-ness, as there seems to be a straightforward connection between the subject's awareness at stake in MSA and non-reflectivity. Straightforward this link might be, however, it does not seem to take the form of a *logical entailment*. Moreover, this distinction is supposed to discriminate different forms of *state* self-consciousness too—that is, specifically, those involving a higher-order state and those involving a same-order state (see Kriegel 2004). And some philosophers consider the awareness at stake in MSA ultimately grounded in a form of state self-consciousness.

Being in a conscious mental state often—if not always—involves an awareness-of relation connecting the subject of experience with what it is experienced. Take, for instance, perceptual experiences of external objects: consciously seeing a blue sky is to be presented with the bluishness of the sky, and, to this extent, to be aware of (at least some features of) the sky. A straightforward way of construing such an awareness relation is in terms of representation. The idea, very roughly, is that "being aware of x" is a matter of representing x as being in a certain way, that is, harbouring a mental representation that "stands for" x. Thus, when I undergo a conscious experience of a blue sky, I harbour a mental representation that "stands for" the blue sky, and it is by virtue of harbouring such a mental representation that I am aware of the sky.

Given the distinction between representation and what it is represented, construing awareness as a representational relation entails that the connection between the subject and the object of awareness is mediated, and consequently indirect. However, some philosophers have argued that some ways of being conscious of things are so immediate and direct that they cannot be accounted for in representational terms.[18] Accordingly, the relation of awareness has to be construed—at least in some cases—as non-representational in nature. The term "acquaintance" is usually applied to such a non-representational awareness relation, a relationship similar to representation but lacking the typical mediation between the subject and the object of awareness. As stressed by Levine, acquaintance is just not acquaintance without directness and cognitive immediacy (cf. Levine 2019: 35).

If we employ the representational/non-representational distinction to the phenomenon of self-consciousness, the subject of the matter is that in self-consciousness reports, such as "I am conscious of myself", the "of" of such an awareness could stand for either the "of" of representation or the "of" of acquaintance. In the former case, we are faced with a representational form of self-consciousness, whereas in the latter, we are faced with a non-representational form of self-consciousness.

The last distinction I want to consider is the one between conceptual and non-conceptual forms of self-consciousness. One typical way of characterising self-consciousness is in terms of the subject's ability to entertain thoughts that are non-accidentally self-related, that is, in terms of the subject's thinking about themselves *as themselves*. It is widely acknowledged that the canonical expression of the thoughts in question ("I-thoughts") involves the first-person pronoun "I". Moreover, since, arguably, the mastery of the first-person pronoun requires the possession of the first-person concept, self-consciousness is typically conceived as a phenomenon that involves the deployment of concepts.

It seems undeniable that such a conceptual self-consciousness exists. Several philosophers, however, claim that more basic forms of self-consciousness, which do not involve the deployments of concepts, need to be recognised.[19] According to Bermúdez, for instance, both visual perception and somatic proprioception are experiences that possess "non-conceptual first-person contents" (Bermúdez 1998:

[18] See, for instance, proponents of so-called 'naïve realist' or 'relational' theories of perceptual experience, such as Campbell (2002) and Fish (2009), according to whom our perceptual awareness of external object is direct and unmediated.

[19] For an in-depth exposition of this theory, see Bermudéz 1998: ch. 3, Musholt 2015: ch. 2, and Smith 2017: §3.

131) and, as such, they are genuine forms of self-consciousness. The upshot, thus, is that when we talk of self-consciousness, we need to distinguish between those forms that possess conceptual content and those forms that possess non-conceptual content. The former are conceptual forms of self-consciousness, whereas the latter are non-conceptual forms.

4. Ways to Answer the Problem

4.1 Five Pairs of Opposing Theses about Minimal Self-Consciousness

Let us recap the five distinctions schematically by expressing the thesis concerning the minimal form of self-consciousness. Up until now, I have presented most distinctions by talking about the existence of forms of self-consciousness; however, we are not concerned with all forms but only with the minimal (the most fundamental) occurrence of self-consciousness. In this case, for each distinction, only one component of the pair applies to minimal self-consciousness. As a result, each distinction identifies two opposing theses about minimal self-consciousness. They are as follows:

1. Creature self-consciousness (CSC) versus state self-consciousness (SSC)
   a. (CSC) Minimal self-consciousness is a property of subjects.
   b. (SSC) Minimal self-consciousness is a property of mental states.
2. Egological (E) versus non-egological (NE)
   a. (E) Minimal self-consciousness is a subject-involving awareness.
   b. (NE) Minimal self-consciousness is a mental-state-involving awareness.
3. Pre-reflective (PR) versus reflective (R)
   a. (PR) Minimal self-consciousness is a pre-reflective (intransitive) awareness.
   b. (R) Minimal self-consciousness is a reflective (transitive) awareness.
4. Representational (RSC) versus non-representational (NRSC)
   a. (RSC) Minimal self-consciousness is a representational awareness.
   b. (NRSC) Minimal self-consciousness is a non-representational awareness.
5. Conceptual (CMSC) versus non-conceptual (NCMSC)
   a. (CMSC) Minimal self-consciousness requires the mastery of certain conceptual abilities.
   b. (NCMSC) Minimal self-consciousness does not require the mastery of certain conceptual abilities.

This schema might raise some questions. Specifically, one might wonder if there are redundant pairs of opposing theses. To remove any doubt, let us analyse the conceptual independence between the five distinctions. For some of them, the independence seems quite pacific (e.g., between 1 and 4, 1 and 3, or 2 and 3).[20] For other distinctions, independence seems less obvious. For instance, one could claim that 3 and 5 are coextensive. There seems to be a *prima facie* relationship between the mastery of conceptual abilities and the capacity to reflect upon oneself and/or one's own mental states. One natural way to account for such a relationship is in terms of identity. After all, what might introspection ever be, if not a self-direct thought? Moreover, what could the content of a thought be, if not a concept?

[20] The distinctions 3 and 4 are also independent unless one endorses the bold reading of 3 that I have discussed in fn15.

Despite its initial appeal, however, several philosophers engaged in the debate on self-consciousness have put pressure on such a claim, particularly those who advocate for thoughts with *nonconceptual content* and those who advocate for a kind of *introspection* that *does not deploy concepts.*[21] Whether they are correct views is not our concern here; both views imply that the pre-reflective/reflective distinction is not coextensive with the non-conceptual/conceptual distinction, and this is all we need to make a (*prima facie*) case for the independence between 3 and 5.

4.2. Five Theories on Minimal Self-Consciousness

At our disposal now are five independent pairs of opposing theses on minimal self-consciousness. With these pairs in place, we can classify the various responses to the problem of for-me-ness into five different positions—or classes of positions—which I call to highlight their characteristic claims: (a) *state reflectivism*, (b) *state non-conceptualism*, (c) *state non-reflectivism*, (d) *state non-representationalism*, and (e) *creature non-representationalism.*[22] [23] While they do not exhaust the whole space of logically possible approaches to the problem, this should not be concerning, as my goal here is just to capture the main positions advocated in the literature.

(a) *State reflectivism* considers minimal self-consciousness to be ultimately a state self-conscious phenomenon, where a mental state is directed at ("is aware of") another mental state, by representing it *as owned by the subject* and by making it a state the subject is conscious of—that is, a state that is "for me". According to this view, minimal self-consciousness is a property of mental states because it is a subject's mental state—which we can call, by adopting the logical jargon, the second-order state—that targets another subject's mental state—which we can the first-order state—and not (at least directly) the subject themselves.[24]

It is worth noting two things here: first, the second-order state has to be called a self-conscious state because the subject of experience is represented in its content—it is a form of egological state self-consciousness; second, since the content of the second-order state has a quite complex form—viz "*I am in this mental state"*—it is taken to be a *conceptual* mental state (typically a thought). Examples of *state reflectivism* are HOT theories of consciousness, according to which a conscious mental state is a state which is the object of a conceptual higher-order thought. Whether such conceptual high-order thoughts are innate or acquired is controversial among the supporters of HOT; for example, Gennaro (2012) argues for an innatist view, while Carruthers (2000) claims that conceptual abilities to exhibit consciousness develop in tandem with the Theory of Mind (ToM). In conclusion, *state reflectivism* addresses the problem of for-me-ness by claiming that for-me-ness is constituted by (or grounded in) a *reflective, egological, representational,* and *conceptual* form of state self-consciousness. Philosophers who endorse some

[21] For a review of philosophers who argue for the existence of a thought's non-conceptual content, see Muscholt 2015: ch. 3. For a defense of a nonconceptual form of introspection, see Giustina & Kriegel 2017.

[22] Remember that, at the end of Section 2, we defined the problem of for-me-ness as the problem of understanding the nature of minimal self-consciousness.

[23] These positions might sound unfamiliar to the reader. Actually, they are unfamiliar in terminological terms but, arguably, not in substantive terms. In fact, as we will see in a moment, some of the leading accounts along their lines are quite familiar in the debate over the nature of consciousness.

[24] Cf. Gennaro 2012; Rosenthal 2005; Carruthers 2000.

version of this theory include Carruthers (2000), Gennaro (2012), and Rosenthal (1998, 2005).

The second position is the one I call (b) *state non-conceptualism*, according to which minimal self-consciousness is ultimately thought to be a kind of state self-consciousness, where a perceptual mental state is directed at ("is aware of") another mental state, by representing it and, thus, by transforming it into a phenomenally conscious state. As the reader has likely already noticed, state non-conceptualism shares the spirit of (a) but differs from the latter because the higher-order state is taken to be non-conceptual. This non-conceptual view is the reason why the content of the higher-order state is usually considered not to be subject-involving, but only mental state-involving. It is a non-egological kind of state self-consciousness. Paradigmatic cases of *non-conceptualism* are HOP (high-order perception) theories of consciousness.

To sum up, state non-conceptualism addresses the problem of for-me-ness by claiming that for-me-ness is constituted by (or grounded in) a *reflective, non-egological, representational,* and *non-conceptual* form of state self-consciousness. Philosophers who endorse some version of this view include Armstrong (1968) and Lycan (1996).

A third position is (c) *state representationalism*, according to which minimal self-consciousness consists ultimately in a particular reflexive awareness, that is, a subject's mental state representation of itself. In this view, the minimal self-consciousness constituting the for-me-ness of experience is an occurrence of a representational and non-egological form of state self-consciousness.

State representationalism is similar to the previous views in considering the awareness constituting the for-me-ness of experience ultimately as a relation borne by a subject's mental states, but it is dissimilar to them because it denies the idea that minimal self-consciousness is a reflective phenomenon—that is, it is not a matter of a mental state to be "aware of" *another* mental state. In addition, in contrast with (a), this theory denies that the self-representing state necessarily has a conceptual nature. Paradigmatic cases of state representationalism are self-representational theories of consciousness, according to which a mental state is conscious if it represents itself (in the right way).

To sum up, state representationalism is a position that considers for-me-ness ultimately constituted by a *pre-reflective, non-egological, representational,* and *non-conceptual* form of state self-consciousness. Philosophers who endorse some version of this view include Kriegel (2009), Caston (2002), and Williford (2006).

A fourth common position is (d) *state non-representationalism*, which considers minimal self-consciousness an intrinsic property of mental states that is not accountable in representational terms. As with (c), state non-representationalism claims that experiences have a reflexive structure, owing to which any experience is immediately aware of itself and thereby given to the subject who undergoes it. But supporters of (d) deny the idea that the experience represents itself. By contrast, they often account for the relation that the experience entertains with itself in terms of self-acquaintance (cf. Williford 2019). State non-representationalism therefore construes the minimal self-consciousness that ultimately grounds the for-me-ness of experience as a *pre-reflective, non-egological, non-representational,* and *non-conceptual* form of state self-consciousness. Philosophers who embrace some version of this view include Smith (1989), Williford (2019), and Zahavi (2005, 2014).

The fifth, and last, position I consider is (e) *creature non-representationalism*, according to which minimal self-consciousness is constituted by (or grounded in) a non-intentional relation ultimately borne by the subject of experience. The idea, here, is that the subject, by virtue of being aware of having the experience, is immediately aware of themselves as the bearer of the experience, and this awareness makes the experience precisely "for me". Such a form of self-consciousness is not taken to be representational and, usually, it is accounted for in terms of acquaintance. According to supporters of (e), any conscious phenomenon, from the simplest to the most complex one, displays such a primitive subject's self-awareness. As Nida-Rümelin puts it,

> [B]eing presented with something necessarily involves being pre-reflectively and pre-conceptually aware of being the subject to whom something is presented. […] According to the view here proposed, pre-reflective self-awareness is an awareness of oneself as an experiencing subject (Nida-Rümelin 2017: 12).

As the reader may have already noticed, *creature non-representationalism,* by claiming that for-me-ness is ultimately constituted by (or grounded in) a kind of creature self-consciousness, differs from all the previous positions. It does not explain the subject's self-consciousness in terms of a more primitive self-consciousness relation of some subject's mental state. We can thus conclude that creature non-representationalism is a theory that considers for-me-ness grounded in a pre*reflective*, *egological*, *non-representational*, and *non-conceptual* form of creature self-consciousness. Such a view is much less popular in the literature, but it is not without its defenders, such as Duncan (2018) and Nida-Rümelin (2017).

5. Conclusion

In this paper, I made three main claims about the problem of for-me-ness to systematise the philosophical debate that grew up around it. In Section 2, I claimed that for-me-ness—the phenomenon to be explained—is best thought of as minimal self-consciousness. In Section 3, I claimed that the nature and the structure of the phenomenon at stake can be established by considering five distinctions: (i) creature versus state self-consciousness, (ii) egological versus non-egological self-consciousness, (iii) reflective versus pre-reflective self-consciousness, (iv) representational versus non-representational self-consciousness and (v) conceptual versus non-conceptual self-consciousness. In Section 4, based on such distinctions, I claimed that the following five positions capture the main existing accounts of the nature of for-me-ness: (a) state reflectivism, (b) state non-conceptualism, (c) state representationalism, (d) state non-representationalism, and (e) creature non-representationalism.

Although the primary goal of the proposed taxonomy is to take a step towards a proper systematisation of the debate on the problem of for-me-ness, its relevance goes beyond a mere description of the current state of this debate. By identifying those aspects that a theory needs to take into account to qualify itself as a response to the problem of for-me-ness, the taxonomy provides us with an analytical framework for better approaching the problem under discussion. According to the framework proposed herein, addressing the problem of for-me-ness means taking a stance on (at least) one pair of opposing theses among those stated in Section 4.1, fixing the constitutive aspects of minimal self-consciousness. More precisely, if we take a

stance on just some pairs, we provide a *partial* response to the problem, but if we take a stance on all of them, we provide a *full* response to it.

To be sure, this is no to say that it is the only viable framework; alternative ways to frame the problem of for-me-ness are possible and, plausibly, they will reflect the same actual space of debate captured by my taxonomy. As far as I know, however, such alternative frameworks are not present. Furthermore, framing the problem of for-me-ness in this way is particularly profitable for at least two reasons. First, by making transparent the definitional theses of the main theories of for-me-ness, it enables us to get a grip on the most popular theses about the nature of minimal self-consciousness. Second, it enables us to develop different strategies to address the problem. One can, for instance, takes a stance on one thesis about the nature of minimal self-consciousness either by arguing for it or by arguing against its opposite thesis.

Scholars involved in the debate should therefore find this taxonomy useful not only to recognise and assess the core theses of the existing answer to the problem of for-me-ness but also to develop their own response.[25]

References

Armstrong, D.M. 1968, *A Materialist Theory of Mind*, New York: Humanities Press.

Bermudez, J.L. 1998, *The Paradox of Self-Consciousness*, Cambridge, MA: MIT Press.

Block, N. 1995, "On a Confusion about a Function of Consciousness", *Behavioral and Brain Sciences*, 18, 227-87.

Campbell, J. 2002, *Reference and Consciousness*, Oxford: Oxford University Press.

Carruthers, P. 2000, *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.

Caston, V. 2002, "Aristotele on Consciousness", *Mind*, 111, 751-815.

Di Francesco, M., Marraffa, M. & Paternoster, A. 2016, *The Self and its Defenses: From Psychodynamics to Cognitive Science*, London: Palgrave Macmillan.

Dretske, F. 1993, "Conscious Experience", *Mind*, New Series, 102, 406, 263-83.

Duncan, M. 2018, "Subjectivity as Self-Acquaintance", *Journal of Consciousness Studies*, 25, 3-4, 88-111.

Farrell, J., & McClelland, T. 2017, "Editorial: Consciousness and Inner Awareness", *Review of Philosophy and Psychology*, 8, 1, 1-22.

Fisch, W. 2009, *Perception, Hallucination, and Illusion*, Oxford: Oxford University Press.

Gennaro, R.J. 2012, *The Consciousness Paradox: Consciousness, Concepts, and High-Order Thoughts*, Cambridge, MA: MIT Press.

Giustina, A. & Kriegel, U. 2017, "Fact-Introspection, Thing-Introspection, and Inner Awareness", *Review of Philosophy and Psychology*, 8, 1, 143-64.

[25] Materials of this paper were presented at the "Philosophy of Mind and Subjectivity" symposium, hosted within the 13th conference of the Italian Society for Analytic Philosophy (SIFA) in Novara. I am grateful to the organisers of this symposium, as well as the audience there. I am especially grateful to Alfredo Tomasetta for discussing at length a previous version of this paper with me. I would also like to thank two anonymous referees of *Argumenta* for their useful comments and remarks.

Goldman, A.I. 1970, *A Theory of Action*. Princeton: Princeton University Press.

Guillot, M. 2017, "I Me Mine: On a Confusion Concerning the Subjective Character of Experience", *Review of Philosophy and Psychology*, 8, 1, 103-27.

Howell, R.J. & Thompson, B. 2017, "Phenomenally Mine: In Search of the Subjective Character of Consciousness", *Review of Philosophy and Psychology*, 8, 1, 103-27.

Janzen, G. 2008, *The Reflexive Nature of Consciousness*, Amsterdam: John Benjamins.

Kriegel, U. 2004, "Consciousness and Self-Consciousness", *Monist*, 87, 182-205.

Kriegel, U. 2005, "Naturalizing Subjective Character", *Philosophy and Phenomenological Research*, LXXI, 1, 23-57.

Kriegel, U. 2009, *Subjective Consciousness*, New York: Oxford University Press.

Kriegel, U. 2011, "Self-Representationalism and the Explanatory Gap", in J. Liu & J. Perry (eds.), *Consciousness and the Self: New Essay*, Cambridge: Cambridge University Press, 51-75.

Kriegel, U. 2018, "Brentano's Dual-Frame Theory of Consciousness", *Philosophy and Phenomenological Research*, 97, 79-98.

Levine, J. 2001, *Purple Haze. The Puzzle of Consciousness*, Oxford: Oxford University Press.

Levine, J. 2016, "A 'Quasi-Sartrean' Theory of Subjective Awareness", in S. Miguens, G. Preyer, & B.C. Morando (eds.), *Pre-Reflective Consciousness: Sartre and Contemporary Philosophy of Mind*, New York: Routledge, 342-61.

Levine, J. 2019, "Consciousness is Acquaintance", in J. Knowles & T. Raleigh (eds.), *Acquaintance: New Essays*, Oxford: Oxford University Press, 33-48.

Lowe, E.J. 1996, *Subjects of Experience*, Cambridge: Cambridge University Press.

Lycan, W. 1996, *Consciousness and Experience*, Cambridge: Oxford University Press.

McClelland, T. 2015, "Affording Introspection: An Alternative Model of Inner Awareness", *Philosophical Studies*, 172, 9, 2469-92.

Miguens, S., Morando, C. & Preyer, G. (eds.), 2016, *Pre-reflective Consciousness: Sartre and Contemporary Philosophy of Mind*, London: Routledge.

Musholt, K. 2015, *Thinking about Oneself: From Nonconceptual Content to the Concept of a Self*, Cambridge, MA: MIT press.

Nagel, T. 1974, "What Is It Like to Be A Bat?", *The Philosophical Review*, 83, 4, 435-50.

Nida-Rümelin, M. 2017, "Self-Awareness", *Review of Philosophy and Psychology*, 8, 1, 55-82.

Rosenthal, D.M. 1998, "A Theory of Consciousness", in N. Block, O.J. Flanagan & G. Guzeldere (eds.), *The Nature of Consciousness*, Cambridge, MA: MIT Press, 729-53.

Rosenthal, D. M. 2005, *Consciousness and Mind*, London: Oxford University Press.

Sebastián, M.Á. 2012, "Experiential Awareness: Do You Prefer 'It' to 'Me'?", *Philosophical Topics*, 40, 2, 155-77.

Smith, D.W. 1989, *The Circle of Acquaintance*, Dordrecht: Kluwer.

Smith, J. 2017, "Self-Consciousness", in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2017 Edition.

Van Gulick, R. 2018, "Consciousness", in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2018 Edition.

Williford, K. 2006, "The Self-representational Structure of Consciousness", in U. Kriegel & K. Williford (eds.), *Self-representational Approaches to Consciousness*, Cambridge, MA: MIT Press, 111-42.

Williford, K. 2019, "Self-Acquaintance and Three Regress Arguments", *ProtoSociology*, 36, 368-412.

Zahavi, D. 2005, *Subjectivity and Selfhood: Investigating the First-Person Perspective*, Cambridge, MA: MIT press.

Zahavi, D. 2014, *Self and Other: Exploring Subjectivity, Empathy, and Shame*. London: Oxford University Press.

Zahavi, D. 2018, "Consciousness, Self-Consciousness, Selfhood: A Reply to Some Critics", *Review of Philosophy and Psychology*, 9, 3, 703-18.

Zahavi, D. 2019 "Consciousness and (Minimal) Selfhood: Getting Clearer on For-me-ness and Mineness", in U. Kriegel (ed.), *The Oxford Handbook of Philosophy of Consciousness*, Oxford: Oxford University Press, 1-19.

Zahavi, D. & Kriegel, U. 2015, "For-me-ness: What It Is and What It Is Not", In D. Dahlstrom, A. Elpidorou & W. Hopp (eds.), *Philosophy of Mind and Phenomenology*, Oxford: Routledge, 36-53.

Kant on the Analyticity of Logic

Costanza Larese

Scuola Normale Superiore, Pisa

Abstract

This paper calls into question the traditional interpretation that logic is, according to Kant, analytic. On the basis of a reconstruction of the salient features of both Kant's theory of analyticity and conception of pure general logic, it is shown that Kant does not apply the analytic-synthetic distinction to logical judgments at all. Moreover, applying Kant's definitions beyond his reasons for leaving the matter unsolved leads to the result that many logical judgments are neither analytic nor synthetic.

Keywords: Kant, Analytic-synthetic distinction, Pure general logic, Formality of logic, Containment criterion.

1. Introduction

In his *Wissenschaftslehre*, Bolzano writes: "Concerning logic, K.[ant] claimed that it (i.e., pure, general logic) consisted of nothing but analytic judgments" and adds "I cannot agree with this finding: rather, it seems to me that logic contains a considerable number of synthetic propositions" (Bolzano 2014, §315, vol. III: 161-62). The latter claim results from his criticism of Kant's notion of analysis together with his conception about the nature of logic. On the contrary, the idea that logic, according to Kant, is analytic is something that is taken for granted and Bolzano does not justify this assumption.

A similar position is taken by Frege. His work is devoted to the program of reducing arithmetic to logic, which amounts to showing, against Kant, that arithmetic is analytic. But Frege neither feels the need to specify that, according to his own definition of analyticity (Frege 1960, §3: 4), logical truths turn out to be analytic, nor to compare this outcome of his theory with Kant's position. That logic, according to Kant, is itself analytic is an unspoken assumption that works behind the scenes of the logicist program. This heritage is accepted by the Vienna Circle. Again, the critical target is Kant's synthetic *a priori*: this category, which has already been impoverished by Frege's thesis that arithmetic is analytic, is now rejected in toto (Carnap, Hahn and Neurath 1973: 308). But, once more, neither the idea that logic is analytic, nor the alleged Kantian origin of this thesis is called into question.

Argumenta 8, 1 (2022): 173-187
ISSN 2465-2334

DOI 10.14275/2465-2334/20200.lar

What is perhaps more surprising is that this belief is not shaken even in a work like Hintikka's *Logic, Language-Games and Information* that aims at vindicating Kant's position against the attacks of the Vienna Circle. Here (1973: 182), Hintikka argues that some polyadic first-order inferences are synthetic *a priori* and that Kant would have considered these modes of reasoning mathematical, rather than logical. This is indeed a vindication of Kant's claim that mathematics is synthetic *a priori*. However, the flip side of Hintikka's reasoning is that analytic first-order arguments correspond with arguments that Kant treated in logic. Again, Hintikka does not justify his claim that Kant considered the logic of his days to be analytic.

Despite rare notable exceptions,[1] this interpretative attitude has survived until now. For example, Hanna (2001: 140) states that "Kant also holds that all the truths of logic—that is, all the truths of what he regarded as logic—are analytically true" and Anderson (2015: 103) states Kant's alleged suggestion that formal general logic is analytic.

The aim of this paper is to shed light on this thorny issue: is it really the case that logic is, according to Kant, analytic? In order for this question to make sense, it is necessary to specify, first of all, at least some defining features of Kant's theory of analyticity.

2. The Containment Criterion of Analyticity

In a famous passage of the Introduction to the first *Critique*, Kant presents his analytic-synthetic distinction in the following terms:

> In all judgments in which the relation of a subject to the predicate is thought (if I consider only affirmative judgments, since the application to negative ones is easy) this relation is possible in two different ways. Either the predicate *B* belongs to the subject *A* as something that is (covertly) contained in this concept *A*; or *B* lies entirely outside the concept *A*, though to be sure it stands in connection with it. In the first case I call the judgment **analytic**, in the second **synthetic**.[2]

This excerpt makes clear that the containment criterion of analyticity cannot apply to judgments whatsoever. First, it is restricted to true judgments. Analyticity in terms of containment is a sufficient reason for the truth of judgments: as a result, false judgments cannot be analytic. Second, it applies only to affirmative judgments. Nevertheless, the definition can be easily extended as to contemplate also negative judgments, which might be said to be analytic if the predicate is incompatible with the concept of the subject. Third, it is restricted only to categorical judgments, namely judgments of the subject-predicate form.

Kant's theory of analyticity has been attacked mainly on the ground of the third restriction. The containment criterion soon appeared too narrow. For example, Frege finds in this restriction one of the reasons for what he took to be Kant's misunderstanding of the status of arithmetical judgments: "Kant obviously—as a result, no doubt, of defining them too narrowly—underestimated the value of analytic judgments" (1960, §88: 99-100). Many denied the very fact that Kant intended

[1] See De Jong 2010: 250, and Burge 2005: 388.

[2] (CPR) A6-7/B10. All quotations from Kant's *Critique of Pure Reason* follow the English translation in Kant 1998 and are cited by page numbers in the original first (A) and second (B) editions preceded by the acronym (CPR).

his analytic-synthetic distinction to apply only to categorical judgments.[3] On the one hand, some maintained that the containment criterion was nothing but a proper part of Kant's theory of analyticity, which would be extended thanks to more comprehensive criteria.[4] On the other hand, some tried to balance the weight of the textual evidence given by the very beginning of the quotation above with other Kantian *loci*, that have been read as proofs of Kant's supposed intention to apply his distinction to judgements of any kind.[5]

Two are the main texts that are usually interpreted as saying that Kant intended to apply his analytic-synthetic distinction to any kind of judgments:

> Judgments may have any origin whatsoever, or be constituted in whatever manner according to their logical form, and yet there is nonetheless a distinction between them according to their content, by dint of which they are either merely *explicative* and add nothing to the content of the cognition, or *ampliative* and augment the given cognition; the first may be called *analytic* judgments, the second *synthetic* (Kant 1997: 16).

> Every existential proposition is synthetic ((CPR) A598/B626).

Proops (2005: 592ff.) has persuasively shown that the two passages can be given a different reading. The former does not mean that the analytic-synthetic distinction applies to judgments regardless of their logical form, but rather that any subject-predicate judgment may have any *degree of distinctness* whatsoever and still be appropriately classified as analytic or synthetic. The latter, together with Kant's famous claim that existence is not a predicate, seems to suggest that there are non-categorical judgments that are synthetic. Nevertheless, Proops points out that Kant, in his criticism of the ontological argument, holds that existence is not a "real predicate, i.e., a concept of something that could add to the concept of a thing" ((CPR) A599/B627), but he does not say that existence is not a *logical* predicate. On the contrary, Kant claims that "anything one likes can serve as a logical predicate" ((CPR) A598/B626): existence included.

Therefore, textual evidence is not overwhelming. But conceptual motivations are decisive. Kant's theory of analyticity is restricted to categorical judgements simply because it is not possible to apply the containment criterion to judgments that are not of the subject-predicate form. Only categorical judgments require the relation of thought between concepts, between a subject and a predicate; on the contrary, disjunctive and hypothetical judgments consider the relation of thought of judgment to judgment(s).

Moreover, if Kant had intended to apply his distinction via containment to all kinds of judgments, he could have worked out a strategy to reduce non-categorical judgments to categorical ones along the Leibnizian lines. But Kant did not go down that road. Hypothetical and disjunctive judgments are enumerated, together with categorical ones, under the heading "relation" in Kant's table of judgments and Kant insists that all the twelve forms of judgments must be recognized as primitive.[6]

[3] See e.g. Kneale and Kneale 1962: 357.
[4] See e.g. Hanna 2001: 145.
[5] See e.g. Anderson 2015: 20.
[6] See Kant 1992, §105: 601.

Since it applies only to true, (affirmative), categorical judgments, Kant's analytic-synthetic distinction via the containment criterion is not exhaustive and, as a consequence, there are some judgments that are neither analytic nor synthetic. This characteristic of Kant's classification is surely a disappointment for most of the twentieth-century philosophers, but probably not for his contemporaries. The main critical target of the *Critique* is the metaphysics based on the Leibnizian predicate-in-subject theory[7] and, given the close relationship between containment and categorical judgments, it is sufficient for Kant to focus on judgments of the subject-predicate form. In other words, Kant's "chief concern is to argue for the syntheticity of certain judgments", such as the claims of mathematics, natural sciences and metaphysics, "that in his days would have been assumed to have subject-predicate form" (Proops 2005: 589).

Beside the charge of narrowness, Kant's theory of analyticity has been accused for a long time of both psychologism and obscurity. While the former criticism can be easily dismissed,[8] the latter is more serious. It plays an important role in Bolzano's analysis of the Kantian definition[9] and has become a cliché after Quine's attack in his influential *Two Dogmas of Empiricism*, where he states that Kant's formulation "appeals to a notion of containment that is left at a metaphorical level" (Quine 1951: 21). Only recently some scholars[10] have challenged this interpretative trend by pointing out that the containment criterion, far from being a metaphorical formulation, is instead a precise notion. As Anderson explains, Kant follows the Wolffian tradition and clarifies the standard notion of containment by appealing to the theory of logical division of concepts and Porphyrian concept hierarchies.

According to the traditional theory of concepts, each genus is said to be "contained in" its species and each species is "contained under" its genus. For Kant, containment relations are thus ordered in a hierarchy of genera and species, where each genus is contained in its species and each species is contained under its genus. While admitting a *summum genus*, Kant denies the possibility of lowest concepts,[11] because, since concepts are general, their extension "must at every time contain other concepts, i.e., subspecies, under itself" ((CPR) A656/B684). By virtue of this relation between containment and the theory of genus and species, the rules of logical division can be applied to the standard notion of containment.[12]

[7] See Anderson 2015.
[8] On this point see e.g. Hanna 2001: 155ff.
[9] See Bolzano 2014, §148, vol. II: 61-2.
[10] In particular, Anderson 2015: Part I, and De Jong 1995.
[11] Kant believes that there can be no singular concept. The clash with his table of judgments, which distinguishes between universal, particular and *singular* judgments according to the quantity of the subject concept, is only apparent. Kant argues that "It is a mere tautology to speak of universal or common concepts—a mistake that is grounded in an incorrect division of concepts into *universal*, *particular*, and *singular*. Concepts themselves cannot be so divided, but only *their use*" (Kant 1992, §1: 589). Thus, every concept is general, but might be *used* to think about singular things.
[12] This theory accounts also for the fact that Kant, in the *Groundwork of the Metaphysics of Morals* (Kant 1902, IV 417), takes hypothetical imperatives to be analytic. The point is that the analysis of the concept of "willing the end" contains the concept "ought to will the necessary means" (notice that this formulation preserves the possibility of willing an end and, at the same time, not willing the necessary means, which is indispensable for an imperative to be an imperative).

The divisions, which are based on the Aristotelian definitions, are governed by the rule that the species exhaust the divided genus and exclude one another: divisions are exhaustive and exclusive disjunctions. Therefore, the relation of two concepts is either of complete inclusion or of total exclusion: partial overlaps are not admitted in these concepts' hierarchies. As a result, judgments that connect any two concepts will be either true, in the case of total inclusion, or false, in the case of total exclusion. Containment is not a metaphor, but rather a technical criterion deeply rooted in theories that were widely available in Kant's days.

3. Clarification, Identity and Contradiction

In the first *Critique*, Kant's distinction between analytic and synthetic judgments is defined not only in terms of containment, but also according to three more criteria, namely, clarification, identity and contradiction:

1. The clarification criterion is characterized in the Introduction to the *Critique* combining both a negative and a positive requirement. The former is that analytic judgments "through the predicate […] do not add anything to the concept of the subject"; the latter is that analytic judgments break the concept of the subject up "by means of analysis into its component concepts, which were already thought in it (though confusedly)" ((CPR) A7/B11; see also Kant 1997: 19).
2. Again, in the Introduction, Kant explains that in analytic judgments (affirmative ones) "the connection of the predicate is thought through identity" ((CPR) A7/B10-11).
3. In the second chapter of the Analytic of Principles, Kant seems to suggest that analytic judgments can be known through the only means of the principle of non-contradiction ((CPR) A151-2/B190-1).

But what is the relationship between Kant's four criteria of analyticity?[13]

The clarification criterion can be reduced to the containment definition, which constitutes its fundamental idea. The deep link between the two versions of analyticity can be mostly appreciated considering the positive feature of the definition above: the clarification of the concepts' intensions involved in a certain analytic judgment, which is obtained through conceptual analysis, consists of showing that the predicate concept is contained in that of the subject. Despite of the immediacy of this argument, some scholars have objected that clarification and containment do not have the same extension, because the former would be characterized by an epistemic flavour that the latter would lack.[14]

However, this criticism can be easily dismissed. It cannot be denied that, according to the clarification criterion, analytic judgments are not cognitively empty in so far as the process of analysis explicates the concepts involved by making distinct their conceptual marks. But this feature of analytic judgments emerges

[13] This is one of the major topics in the literature. Since Kant's four criteria of analyticity do not seem to be equivalent, scholars have discussed on whether the set of Kant's formulation is consistent after all. Some, such as Hanna 2001: 124, argued that each definition "merely brings out a different aspect of a single, internally consistent, defensible Kantian theory". Others believed that Kant's criteria cannot be reconciled and identified one of them as the conceptually fundamental or most mature formulation. See e.g. Anderson 2015: 16, Proops 2005, Allison 2004: 89ff.

[14] See Proops 2005: 602, and Allison 2004: 90.

from the containment criterion as well, where the predicate-concept is explicitly said to be "covertly" ((CPR) A6/B10) contained in the subject concept. The distinction via clarification is still a distinction between two kinds of propositional content, as it is for the containment criterion, and not of two kinds of cognitive procedures. Clarification is a characterization in epistemic terms of the same logical distinction based on the containment criterion.

The relation between containment and the identity criterion is problematic when identical judgements, such as "man is man", are taken into account. On the one hand, according to containment (and clarification), analytic judgments are endowed with cognitive content and are not trivial or tautologous. In particular, the predicate concept must at least be different from the subject concept, for otherwise there is no room for any kind of clarification whatsoever. On the other hand, it is obvious that the identity criterion classifies identical judgments as analytic. Kant himself oscillates on this point.[15] While both in the *Critique* and in the *Prolegomena* he clearly holds that "a = a" is analytic,[16] in other *loci* of his work he rejects the thesis of the analyticity of identical judgements.[17]

Identical judgments notwithstanding, containment and identity are strictly connected. In partial identities, such as "all bodies are extended", the predicate concept is partially identical with the subject concept, because the relation of full identity subsists only between the conceptual notes of the predicate concept and a proper part of the conceptual marks of the subject. But this is just a different way of phrasing the containment criterion, because the predicate concept, being a part of the subject concept, is contained in it. Therefore, containment is the fundamental idea at the basis of the identity criterion, although the latter excludes any consideration of epistemic nature and classifies identical judgments as analytic.

The contradiction criterion has been frequently identified as the best among Kant's versions of analyticity.[18] Two are the main reasons that explain its happy fortune. First, it seems more inclusive than the containment criterion, because it is not restricted to categorical judgments. Second, it is closer to contemporary appeals to the class of logical truths in providing a definition of analyticity. Despite this long interpretative tradition, some scholars have shown that the principle of contradiction is not a definition of analyticity at all, but rather an instrument for knowing the truth of analytic judgments.[19] Textual evidence is here determining:

> Now the proposition that no predicate pertains to a thing that contradicts it is called the principle of contradiction, and is a general though merely negative *criterion of all truth* […]
>
> But one can also make a positive use of it, i.e., not merely to ban falsehood and error (insofar as it rests on contradiction), but also *to cognize truth*. For, **if the judgment is analytic**, whether it be negative or affirmative, *its truth must always be able to be cognized* sufficiently in accordance with the principle of contradiction. For the contrary of that which as a concept already lies and is thought *in the cognition of the*

[15] This fact has been explained in different ways. For example, De Jong (1995: 629-30) holds that, strictly speaking, tautological judgments for Kant are neither analytic nor synthetic. Proops (2005) proposes instead a diachronic reading of Kant's position.
[16] (CPR) B17 and Kant 1997: 19. See also Kant 1992, §37: 607.
[17] See e.g. Kant 1902, XX 322.
[18] See e.g. Kneale and Kneale, 1962: 357-58.
[19] See e.g. De Jong 1995 and Proops 2005: 603.

> *object* is always correctly denied, while the concept itself must necessarily be affirmed of it, since its opposite would contradict the object.
>
> Hence we must also allow the **principle of contradiction** to count as the universal and completely sufficient **principle of all analytic *cognition***; but its authority and usefulness does not extend beyond this, as a sufficient *criterion of truth*. For that *no cognition* can be opposed to it without annihilating itself certainly makes this principle into a *conditio sine qua non*, but not into a determining ground of *the truth of our cognition* ((CPR) A151-2/B190-1, emphasis added).

Kant's explanation of the role of the principle of non-contradiction does not find its place in the Introduction, together with containment, clarification and identity criteria, but only later on in the Analytic of Principles. In this passage, Kant is listing the uses of the principle of non-contradiction and he maintains that it is both a "negative criterion of all truth", meaning that it is a necessary condition for the truth of any judgment, and the "principle of all analytic cognition", namely the necessary and sufficient condition for the cognoscibility of analytic judgments. As the emphasised phrases make clear, Kant is careful in stressing its instrumental role as a criterion for establishing the truth of judgments and its epistemological function for determining the possibility of knowing analytic judgments.

Saying that an affirmative analytic judgment is known in accordance with the principle of contradiction does not mean that it is possible to derive an explicit contradiction from the negation of the judgment involved, but rather that the contradiction rests with the concept of the subject and the negation of the predicate. This is because the predicate is "already thought beforehand in the concept of the subject" (Kant 1997: 17), for if the predicate were not thought in that of the subject, then the denial of the former would not contradict the latter. This means that the ultimate reason for the epistemic function of the principle of contradiction in knowing the truth of analyticities is, once more, the relation of containment between the concepts involved in analytic judgments.

To sum up, the containment criterion, which applies only to true, (affirmative) and categorical judgements, is the central notion of Kant's theory of analyticity, not only because it is announced first and has an expositional priority over the other formulations, but also because the remaining criteria are founded on it and might be (completely or partially) reduced to it.

4. Kant's Conception of Logic

The previous sections have specified the main features of Kant's theory of analyticity. But, in order to understand whether logic is really analytic for the author of the *Critique*, it is obviously necessary to delve into another preliminary issue, namely, Kant's conception of logic. In particular, two questions need to find answers. First, what counts as "logical" for Kant and what kind of logical notions did he possess? Second, how is logic conceived and which are the defining features of this discipline according to Kant?

In Kant's writings, the term "logic" refers to a variety of disciplines. In the introduction to the Transcendental Logic of the *Critique*, he first distinguishes between general and special logics. While the former "contains the absolutely necessary rules of thinking, without which no use of the understanding takes place, […] without regard to the difference of the objects to which it may be directed", the latter "contains the rules for correctly thinking about a certain kind of objects"

((CPR) A52/B76). Then, he states that general logic might be pure or applied. The former abstracts "from all empirical conditions under which our understanding is exercised"; the latter undergoes "the subjective empirical conditions that psychology teaches us" ((CPR) A53/B77). At last, Kant introduces the discipline of transcendental logic that, unlike general logic, investigates the origin and the objective validity of the cognition of pure understanding and pure reason, through which we think objects completely *a priori*.

The relationship between pure general logic, which is the discipline that gets closer to both the traditional and the modern conception of logic and might thus be called logic in the strict sense of the term, and transcendental logic, which is Kant's radical innovation and is a metaphysical discipline, is a debated issue.[20] Although there are no doubts that Kant attached the greatest importance to the latter, which covers the largest part of his first *Critique*, it is also fair to recall that the former is a constant presence in Kant's intellectual life. Not only did he take several courses in logic as a student and deepen his logical knowledge while preparing his *venia legendi*, but he also wrote of logical issues and held numerous courses in logic during his forty-years teaching in Königsberg.[21] Nevertheless, a long interpretational tradition has claimed that Kant's knowledge of logic was quite elementary[22] and his esteem for the latest developments of the discipline rather low.[23] A confirmation of this judgment might come from a closer look to what Kant thought belonged to the domain of the logical.

However, identifying in a precise way which topics were proper of pure general logic according to Kant is no easy feat. First, beyond *The False Subtlety of the Four Syllogistic Figures* (1762), the other logical work published during Kant's time and associated with his name, that is to say, the so-called *Jäsche Logic* (1800), must be treated with caution and cannot be taken as a reliable statement of Kant's view.[24] Something similar happens also for the other texts stemming from Kant's logic lectures as well as for his handwritten *Reflexionen* on Meier's handbook. Second, these texts suggest that the content of his lectures included matters that, on his own account, do not belong to logic proper. But it might be safely assumed that Kant accepted the traditional division of logic into three branches: the theory of concepts, of judgments and of inferences.

The theory of concepts provides a complete characterization of the basic unit of thought, through the discussion of crucial distinctions, such as matter and form, empirical and pure, *a priori* and *a posteriori*. Moreover, it includes all the tools needed to formulate the containment criterion, such as the definitions of "content" and "extension" of concepts, the connections between higher and lower concepts on the one hand and genera and species on the other. The theory of judgments concerns instead the relations between concepts. It focuses on the logical forms of judgments, classifying them accordingly to their quantity, quality, relation and modality, and discusses peculiar kinds of judgments. The theory of inferences consists of a restricted version of the Aristotelian syllogistic with a simple theory of disjunctive and

[20] See e.g. Tolley 2012.
[21] See e.g. Capozzi 2002: 59-113.
[22] See e.g. Bocheński 1961: 6, Kneale and Kneale 1962: 354, Young 1992: xvi, Hazen 1999: 92 and Lapointe 2012: 11.
[23] This at least seems to be suggested by Kant's words at (CPR) B viii.
[24] On this point see Young 1992.

hypothetical inferences added on, such as *modus ponens* and *modus tollens*, together with a distinctive treatment of inductive inferences.

But Kant's original contribution to logic must be searched in his conception and philosophy of this discipline.[25] Pure general logic is for Kant a science in the strict sense of the term, namely "an exhaustive and a priori proven system of the merely formal rules of thought".[26] It is characterized by the following qualifying features. First of all, it is pure, because it disregards the empirical circumstances under which the understanding is applied and concerns only *a priori* principles. Second, it is general. This means that the rules of logic are necessary, because they have to be applied no matter what are the objects we are thinking about. On the one hand, they are constitutive for the understanding, in the sense that we cannot think at all without them;[27] on the other hand, they are normative, in the weaker sense that they prescribe how we have to think correctly.[28]

Third, Kant claims, probably for the first time,[29] that logic is formal, in the sense that, like grammar, it abstracts from the semantical content of thought. As a consequence, logic cannot yield any extension of knowledge about reality or objects (see (CPR) A60/B85). The presumption of employing general logic as a tool for extending knowledge, which is its dialectical use, "comes down to nothing but idle chatter, asserting or impeaching whatever one wants with some plausibility" ((CPR) A61/B86). Kant believed that formality was a direct consequence of the generality of logic[30] and many scholars have argued that the two notions, given some Kantian premises, ultimately collapse.[31]

Fourth, Kant holds that logic is a canon for thinking, which is "the sum total of the a priori principle of the correct use of […] understanding and reason in general, but only as far as form is concerned" ((CPR) A796/B824). The thesis that logic is canon for thinking is a consequence of its three features mentioned above.[32] Since logic is pure, the rules for the correct use of understanding and reason are *a priori*. Since it is general, logic is constitutive and normative for thinking: as a result, it prescribes the correct use of understanding and reason and, in this sense, it is a "cathartic". Last, since it is formal, logic cannot be an organon, namely "a directive as to how certain cognition is to be brought about" (Kant 1992, §13: 528-29), but only a canon.

5. Kant on the Relationship between Pure General Logic and Analyticity

The relationship between analyticity and pure general logic consists of two distinct issues that have been frequently confused. First, the function that logic plays in Kant's definition and application of the analytic-synthetic distinction. Second, the question of whether logic itself is really analytic according to Kant.

[25] See Tiles 2004.
[26] See Lu-Adler 2018: 6.
[27] See Kant 1992, §12: 528.
[28] See Kant 1992, §14: 529.
[29] See MacFarlane 2002: 44-46.
[30] This can be easily seen in (CPR) A52/B76 and Kant 1992, §12: 585.
[31] See e.g. Lapointe 2012 and MacFarlane 2002: 32.
[32] Of course, for what has been said before, the premise that logic is formal is redundant in so far as it can be deduced from the fact that it is general.

Consider the first point. For what has been said above, it is clear that logic is the fundamental instrument that Kant employs for drawing the analytic-synthetic distinction. Containment, to which all of Kant's criteria of analyticity can be reduced (see Section 3), is a technical notion based on the theory of logical division of concepts (see Section 2). These were at the core of early modern logic: the theory of concepts is, at the same time, the foundation of the logical doctrine of elements and, in essence, a theory of concepts containment (see Section 4). Moreover, Kant's pure general logic is an instrument not only for defining the analytic-synthetic distinction, but also for applying it. The principle of non-contradiction, which is the logical principle *par excellence*, has the fundamental epistemological function of determining the truth, as well as the analyticity, of analytic judgments (see Section 2).

The role of logic as the fundamental instrument in Kant's thought for both defining and applying the notion of analyticity is probably one of the reasons why most interpreters, from the very beginning of Kant's critical reception until recent days, have concluded that Kant held that logic itself were analytic (see Section 1). Nevertheless, it is obvious that this conclusion is fallacious.

On the contrary, Kant nowhere seems to claim that logic is analytic. While he explicitly argues that judgments of experience, mathematics, natural science and metaphysics are synthetic (see (CPR) B11-8), Kant does not speak about the status of logic. The thesis that *Kant does not apply the analytic-synthetic distinction to logic* might be proven only in one way. Namely, it must be shown that the strongest passages usually taken to support the opposite view are not overwhelming. This is the case[33] for (CPR) A151-2/B190-1, which has already been quoted in Section 3, together with the following excerpts:

> General logic analyzes the entire formal business of the understanding and reason into its elements, and presents these as principles of all logical assessment of our cognition. This part of logic can therefore be called an analytic ((CPR) A60/B84-5).
>
> I understand by an analytic of concepts not their analysis, or the usual procedure of philosophical investigations, that of analyzing the content of concepts that present themselves and bringing them to distinctness, but rather the much less frequently attempted **analysis of the faculty of understanding** itself […] for this is the proper business of a transcendental philosophy; the rest is the logical treatment of concepts in philosophy in general ((CPR) A65-6/B90-91).
>
> General logic abstracts from all content of cognition, and expects that representations will be given to it from elsewhere, whenever this may be, in order for it to transform them into concepts analytically ((CPR) A76/B102).

As explained above, in (CPR) A151-2/B190-1, Kant affirms that the principle of non-contradiction is an instrument for knowing the truth of analytic judgments and nowhere states that logic is analytic. In (CPR) A60/B84-5, Kant suggests calling the formal part of general logic *an* analytic. His aim here is to distinguish the proper use of logic as a canon for judging from the use of it as an organon for the production of seemingly objective knowledge. He names the former part of logic an analytic to contrast it with the latter part of the discipline, that he calls "dialectic". In so doing, Kant himself explicitly states that he is following the

[33] See e.g. Hanna 2001: 140, and Anderson 2015: 103.

tradition. His choice of the term "analytic" is therefore meant only to underline the formal character of logic.

In (CPR) A65-6/B90-1, Kant suggests that the business of the logical treatment of philosophical concepts, unlike the transcendental one, is to analyse the content of concepts so as to clarify them. In (CPR) A76/B102, Kant is again distinguishing transcendental logic from pure general logic on the basis of the formality of the latter as opposed to the manifold of sensibility *a priori* that lies before the former. The point that Kant makes in the three latter passages is the same. In all of them he is simply underlining the formality of logic. But the formality of logic does not amount *per se* to say that the principles of logic are analytic (see Section 4). It only excludes that logic is synthetic (see Section 6).

Therefore, until proven otherwise, it can safely be concluded that in his writings Kant never claims that logic is analytic. A different question regards the theoretical reasons for his choice. Why doesn't Kant apply his analytic-synthetic distinction to logic? The answer to this question, as De Jong (2010: 250) suggests, must be searched in Kant's peculiar conception of logic proper. In particular, in its characterizing features of generality and formality. In so far as it is general, logic develops rules relative to form and, in so far as it is formal, logic abstracts from the content of thinking. As a result, logic cannot extend our knowledge of real objects. It shares with grammar[34] the destiny of being a propaedeutic, rather than a kind of knowledge:

> Hence logic as a propadeutic constitutes only the outer courtyard, as it were, to the sciences; and when it comes to information, a logic may indeed be presupposed in judging about the latter, but its acquisition must be sought in the sciences properly and objectively so called ((CPR) B ix).

Logic is a propaedeutic because it precedes any kind of knowledge: its rules have to be learnt and respected as a *conditio sine qua non* of any cognitive enterprise. In devising his analytic-synthetic distinction, Kant is primarily interested in doctrines, such as mathematics, metaphysics and sciences, that have some content of knowledge and his main purpose is to argue, against the Leibnizian predicate-in-subject theory, that judgments of those disciplines are synthetic. Determining the status of logic with respect to the analytic-synthetic distinction simply does not belong to Kant's *desiderata*.

6. Applying Kant's Definitions beyond Kant's Intentions

Although Kant, as a matter of fact, does not apply his analytic-synthetic distinction to logic, it is still possible to investigate whether logical judgments are analytic or synthetic *a priori* according to Kant's definitions and beyond Kant's reasons for leaving the matter unsolved. In other words, the following analysis is something that Kant did not want to pursue and did not consider a part of his philosophical strategy.

First of all, it can be shown that, according to Kant, logical judgments are not synthetic. This is simply because logical judgments cannot fit Kant's definition of synthetic judgments. In the latter kind of judgments, the concept of the predicate is not contained in the concept of the subject: rather, the former is "outside" ((CPR)

[34] See Mosser 2008.

A6-7/B10-1) or "beyond" ((CPR) A154-5/B193-4) the latter. Nevertheless, in order for grounding and justifying the truth of synthetic judgments, there must be some kind of connection between the two concepts involved, which must be different from the containment relation. This relation cannot be but indirect in that it has to link two concepts to one another by connecting them to a third and distinct element.[35] This third element is, for Kant, an object in which "the synthetic unity of their concepts could establish objective reality" ((CPR) A157/B196). But the appeal to an object for a logical judgment is what is explicitly excluded by the feature of formality that characterizes logic according to Kant (see Section 4).

Another, albeit partial, evidence comes from the following argument. Suppose, *ad absurdum*, that the principle of non-contradiction is synthetic. Since the principle of non-contradiction is the supreme principle of analytic judgments, it follows that analytic judgments can be derived from a synthetic principle. But now consider Kant's thesis, put forward in the Introduction to the first *Critique* ((CPR) B14), that what can be proved from a synthetic judgment is itself synthetic. It turns out that analytic judgments are synthetic and that assuming the syntheticity of the principle of non-contradiction leads to the collapse of the analytic-synthetic distinction.

Notice that the claim that logical judgments are not synthetic is nothing more than a confirmation of the special status of logic in Kant's epistemology. Logic, for Kant, is a science. But while all the other theoretical sciences, such as mathematics, are based on synthetic *a priori* judgments, this discipline, in so far as it is a body of necessary rules and a propaedeutic for thought, must be different. What is perhaps more surprising is that *at least some judgments of logic are not even analytic following Kant's definition*. However, unlike for the question regarding the syntheticity of logical judgments, which can be given a compact (and negative) answer, the issue of the analyticity of logic requires to divide logical judgments into several categories and to consider them one by one.

The first class includes those principles (mainly belonging to the theory of concepts and the theory of judgments), which in Kant's days belonged to logic with full right but have been excluded from the discipline during the development of modern symbolic logic. Examples of judgments of this class are "A concept is a universal representation" (Kant 1992, §1: 589) or "Propositions whose certainty rests on *identity* of concepts (of the predicate with the notion of the subject) are called *analytic* propositions" (Kant 1992, §36: 606).

Are this kind of judgments analytic? *Some* of them are analytic for sure. Consider the first of the examples above. In the passage of the first *Critique* commonly known as the "Stufenleiter" ((CPR) A320/B377), Kant divides the concept "representation" by following the traditional theory of the logical division of concepts (see Section 2), that is, by obtaining more specific concepts through the gradual addition of *differentiae specificae* to the higher ones. Through this method, a concept is said to be an objective representation with consciousness, whose relation to the object is mediate by means of a mark which can be common to several things. As a result of this investigation, the judgment "A concept is a universal representation" is analytic, because "representation" is the *genus* and "universal" is a *differentia* of the subject "concept".

But is it possible to conclude that *every* judgment of this kind is analytic? Probably not. For sure none of them is synthetic. Nevertheless, some of them

[35] See e.g. (CPR) A9/B13 and A155/B194.

might still be non-analytic. First of all, because Kant's theory of analyticity is spelled out in terms of containment and containment is the basic notion of the theory of concepts. This means that asking whether the theory of concepts is analytic might raise problems of autoreferentiality, such as in the second of the examples above, since the theory of concepts seems to be a meta-theory, rather than an object theory. Second, in many cases it is not clear at all whether the way in which judgments are formulated is essential to their meaning. In the *Jäsche Logic*, many statements are obviously not categorical (and, as such, cannot be analytic at all), although they can be easily turned into the "S is P" form.

To conclude, it is probably necessary to consider judgments of the first class one by one to determine which of them are analytic and which are not. But with high probability the result of this procedure won't be worth the effort.

Consider now the second class of statements that, contrary to the former, includes judgments that are clearly part of modern symbolic logic, while being not "logical" according to Kant's notion of the term. All of those validities turning essentially on relations belong to this class. In Kant's time, categorical judgments were still considered to be the most fundamental judgments of logic, which was intrinsically monadic in character and not equipped with dedicated instruments for handling relations. As a result, truths turning essentially on relations are not analytic:[36] this is not because they are not logical, but rather because they cannot be properly reduced to categorical propositions. Nevertheless, it is still possible for them to be synthetic, in so far as they were excluded from the domain of the logical, which is, as argued above, the domain of the non-synthetic.

The third and last class of statements to examine is given by those that are logical both according to Kant's traditional conception of the discipline and for modern symbolic logic. This class includes not only propositional inferences, such as *modus ponendo tollens* and *modus tollendo ponens*, but also the hypothetical judgments that can be obtained by these inferences by considering the conjunction of their premises as the antecedent and the conclusion as the consequent. Both of them cannot be reduced to categorical judgements, because of their form: their soundness and validity rely on the relations between judgments independently of the concepts involved. As a result, the statements of this class are neither analytic nor synthetic or, equivalently, the analytic-synthetic distinction cannot apply to them.

One might think at this point that all but non-categorical and non-autoreferential judgments that for Kant belonged to logic are analytic according to his definitions. But this is not the case. For example, also categorical and identical truths would not count, according to Kant, as analytic, because they are not illuminating as the containment-clarification criterion prescribes. Thus, it seems that the class of the logical judgments that are neither analytic nor synthetic according to Kant's definitions of the terms might be wider than what was usually taken to be.

7. Conclusion

Against the prevailing view according to which Kant maintains that logic is analytic, this paper has shown that Kant does not apply his analytic-synthetic distinction to logic at all and that the grounds for this reticence about the status of logic

[36] This point has been variously acknowledged. See e.g. Anderson 2015: 99ff. This is also one of the main reasons that led scholars to reject the very fact that Kant's analytic-synthetic distinction is restricted to categorical judgments. See e.g. Hanna 2001: 145ff.

has to be searched in Kant's peculiar conception of the discipline. Even attempting an analysis that Kant did not think worth pursuing, it has been maintained that no logical judgment is synthetic *a priori* and that at least some logical judgments are not analytic. In other words, following Kant's definition of the analytic-synthetic distinction, it turns out that many logical judgments are neither analytic nor synthetic.

This result is not a mere question of terminology or classification. It bears relevant consequences, in the first place, on Kant's overall theoretical philosophy. To make one example, it calls into question the most common interpretation of Kant's invention of transcendental logic as an instrument devised in order to deal with his new category of the synthetic *a priori*, as opposed to the alleged analytic domain of pure general logic.[37] But it shows its importance also for adjusting the reading of an essential turning point in the history of logic. The idea that analytic judgments are either logical truths or can be turned into logical truths by replacing synonyms for synonyms, together with its consequence that logic is the example *par excellence* of an analytic discipline,[38] belongs to a much later view on the discipline of logic. The bond between logic and analyticity was not so tight at the beginning of the story.[39]

References

Allison, H. 2004, *Kant's Transcendental Idealism: An Interpretation and Defense*, *Revised and Enlarged Edition*, New Haven: Yale University Press.

Anderson, R.L. 2015, *The Poverty of Conceptual Truth. Kant's Analytic/Synthetic Distinction and the Limits of Metaphysics*, Oxford: Oxford University Press.

Bocheński, I.M. 1961, *A History of Formal Logic*, trans. by I. Tomas, Notre Dame: University of Notre Dame Press.

Bolzano, B. 2014, *Theory of Science* (1837), P. Rusnock and R. George (eds.), Oxford: Oxford University Press.

Burge, T. 2005, *Truth, Thought, Reason: Essays on Frege*, Oxford: Oxford University Press.

Capozzi, M. 2002, *Kant e la logica*, Napoli: Bibliopolis.

Carnap, R., Hahn, H. and Neurath, O. 1973, "The Scientific Conception of the World: The Vienna Circle" (1929), in O. Neurath, *Empiricism and Sociology*, M. Neurath and R.S. Cohen (eds.), Dordrecht: Reidel.

De Jong, W.R. 1995, "Kant's Analytic Judgments and the Traditional Theory of Concepts", *Journal of the History of Philosophy*, 33, 631-41.

De Jong, W.R. 2010, "The Analytic-Synthetic Distinction and the Classical Model of Science: Kant, Bolzano and Frege", *Synthese*, 174, 2, 237-61.

[37] See the reconstruction in Tolley 2012: 419ff.

[38] See Carnap, Hahn and Neurath 1973.

[39] I am deeply grateful to Michael Beaney, Marcello D'Agostino and Massimo Mugnai for their illuminating remarks on previous drafts and to Enrico Moriconi, Mario Piazza and Giuseppe Primiero for discussing related issues. I wish to thank also the two anonymous referees, whose suggestions and criticisms contributed to a strong improvement of this paper.

Frege, G. 1960, *The Foundations of Arithmetic* (1884), trans. by J.L. Austin, Evanston: Northwestern University Press.

Hanna, R. 2001, *Kant and the Foundations of Analytic Philosophy*, Oxford: Clarendon Press.

Hazen, A.P. 1999, "Logic and Analyticity", in *The Nature of Logic*, A.C. Varzi (ed.), Stanford: Center for the Study of Language and Information, 79-110.

Hintikka, J. 1973, *Logic, Language-Games and Information*, Oxford: Clarendon Press.

Kant, I. 1902, *Kants gesammelte Schriften. Bde. I-XXIX* (1902-1983), Berlin: De Gruyter, Reimer.

Kant, I. 1992, *Lectures on Logic*, trans. and ed. by J.M. Young, Cambridge: Cambridge University Press.

Kant, I. 1997, *Prolegomena to Any Future Metaphysics that May Be Able to Come forward as a Science* (1783), trans. and ed. by G. Hartfield, Cambridge: Cambridge University Press.

Kant, I. 1998, *Critique of Pure Reason* (1787), trans. and ed. by P. Guyer and A. Wood, Cambridge: Cambridge University Press.

Kneale, W.C. and Kneale, M. 1962, *The Development of Logic*, Oxford: Clarendon Press.

Lapointe, S. 2012, "Is Logic Formal? Bolzano, Kant and the Kantian Logicians", *Grazer Philosophische Studien*, 85, 1, 11-32.

Lu-Adler, H. 2018, *Kant and the Science of Logic: A Historical and Philosophical Reconstruction*, Oxford: Oxford University Press.

MacFarlane, J. 2002, "Frege, Kant and the Logic in Logicism", *The Philosophical Review*, 111, 1, 25-65.

Mosser, K. 2008, "The Grammatical Background of Kant's General Logic", *Kantian Review*, 13, 1, 116-40.

Proops, I. 2005, "Kant's Conception of Analytic Judgments", *Philosophy and Phenomenological Research*, 70, 3, 588-612.

Quine, W.V.O. 1951, "Two Dogmas of Empiricism", *The Philosophical Review*, 60, 20-43.

Tiles, M. 2004, "Kant: from General to Transcendental Logic", in *Handbook of the History of Logic*, D.M. Gabbay and J. Woods (eds.), Amsterdam: Elsevier, vol. III, 85-130.

Tolley, C. 2012, "The Generality of Kant's Transcendental Logic", *Journal of the History of Philosophy*, 50, 3, 417-46.

Young, J.M. 1992, "Translator's Introduction", in I. Kant, *Lectures on Logic*, trans. and ed. by J.M. Young, Cambridge: Cambridge University Press, xv-xxxii.

The Paradox of Infallibility

Daniel Rönnedal

Stockholm University

Abstract

This paper discusses a new paradox, the paradox of infallibility. Let us define infallibility in the following way: (Def I) *t* is infallible if and only if (iff) everything *t* believes is true, where *t* is any term. (Def I) entails the following proposition: (I) It is necessary that for every individual *x*, *x* is infallible iff every proposition *x* believes is true. However, (I) seems to be inconsistent with the following proposition (P): It is possible that there is some individual who believes exactly one proposition, namely that she is not infallible. So, it seems to be the case that either (I) or (P) must be false. Yet, (I) is simply a consequence of (Def I) and (P) clearly seems to be true. This is the puzzle. I discuss five possible solutions to the problem and mention some arguments for and against these solutions.

Keywords: Paradoxes, Infallibility, Epistemic paradoxes, Dialetheism, Propositional quantifiers.

1. Introduction

In this paper, I introduce a new paradox, the paradox of infallibility. Intuitively, this puzzle can be formulated in the following way. Assume that someone is infallible if and only if (iff) everything she believes is true and that there is an individual who believes exactly one proposition, namely the proposition that she is not infallible. Suppose that this individual is infallible. Then everything she believes is true. Hence, she is not infallible, since she believes that she is not infallible. So, if she is infallible, she is not infallible. Suppose that she is not infallible. Then everything she believes is true, since the only proposition she believes is the proposition that she is not infallible. Accordingly, she is infallible. Consequently, if she is not infallible, she is infallible. It follows that she is infallible iff she is not infallible. But this is clearly a contradiction. Hence, it cannot be the case that someone is infallible iff everything she believes is true and that it is possible that there is an individual who believes exactly one proposition, namely the proposition that she is not infallible.

I will now describe the paradox more carefully. I will use the following definition of infallibility:

(Def I) $It =_{df} \forall A(BtA \supset A)$, where *t* is any term. For every *t*, *t* is infallible iff for every (proposition) A, if *t* believes that A, then A (is true).

Argumenta 8, 1 (2022): 189-197
ISSN 2465-2334

DOI 10.14275/2465-2334/20210.ron

(Def I) is a metalinguistic definition. This means that '$\forall$A(BtA $\supset$ A)' can be replaced by '*It*' in any context whatsoever, and vice versa; '*It*' is simply an abbreviation of '$\forall$A(BtA $\supset$ A)'. In (Def I), '*I*' is a predicate, 't' is a term, '$\forall$' is a propositional quantifier, 'B' is a doxastic operator, and 'A' is a propositional variable. All these symbols are used in a standard way. From (Def I) we can derive the following proposition:

(I) $\Box\forall x(Ix \equiv \forall A(BxA \supset A))$. It is necessary that for every (individual) x: x is infallible iff for every (proposition) A, if x believes that A then A.

Note that the first quantifier in (I) varies over individuals while the second varies over propositions or sentences. '$\Box$' is the standard (absolute) necessity operator. Hence, '$\Box$A' is true in a possible world iff 'A' is true in every possible world.[1] It should be obvious that (I) follows from (Def I), since if we replace '*Ix*' by '$\forall$A(BxA $\supset$ A)' in (I) we obtain '$\Box\forall x(\forall A(BxA \supset A) \equiv \forall A(BxA \supset A))$', which obviously is valid. The inconsistency argument (see below) shows (or seems to show) that (I) is incompatible with the following proposition:

(P) $\Diamond\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$. It is possible that there is some individual who believes exactly one proposition, namely that she is not infallible. More precisely, the informal reading of (P) is as follows: It is possible that there is some individual x such that x believes that it is not the case that x is infallible and for every proposition A, if x believes that A, then it is necessary that A iff it is not the case that x is infallible.

Again, note that the first quantifier in (P) varies over individuals while the second varies over propositions or sentences. '$\exists$' is a standard propositional quantifier and '$\Diamond$' is the standard (absolute) possibility operator. Accordingly, '$\Diamond$A' is true in a possible world iff 'A' is true in some possible world. '$\forall$A(BxA $\supset$ $\Box$(A $\equiv$ ~Ix))' does not say that x has only one belief, but it says that if x believes A then A is necessarily equivalent with (and so identical to) the proposition that x is not infallible. Hence, '$\Diamond\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$' is a reasonable symbolisation of the proposition that it is possible that there is some individual who believes exactly one proposition, namely that she is not infallible.

Furthermore, note that (P) only says that it is *possible* that there is an individual of a certain kind. It does not claim that there (*actually*) is an individual of this type. Probably, it is not the case that there is some (actual) individual who believes exactly one proposition, namely that she is not infallible. Still, this does not entail that (P) is false. In other words, (P) is compatible with the following formula: '$\sim\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$'.

I will now show that {(I), (P)} seems to be inconsistent. To establish this, I will assume that (I) and (P) are true in some possible world w_0 and derive a contradiction. I will call this derivation 'the inconsistency argument'. '$\Diamond$', '$\exists$', '$\Box$', '$\forall$' and '$\sim\forall$' in the deduction below are standard derivation rules. 'PL' means that the step follows by ordinary propositional reasoning. Intuitively, 'A, w' says that 'A' is true in the possible world w. Here is the derivation:

[1] For more on modal logic, see, for example, Blackburn, De Rijke, Venema 2001, Chellas 1980 and Hughes and Cresswell 1968.

The Inconsistency Argument

(1) $\Box\forall x(Ix \equiv \forall A(BxA \supset A))$, w_0	[Assumption]
(2) $\Diamond\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$, w_0	[Assumption]
(3) $\exists x(Bx{\sim}Ix \,\&\, \forall A(BxA \supset \Box(A \equiv {\sim}Ix)))$, w_1	[2, $\Diamond$]
(4) $Bc{\sim}Ic \,\&\, \forall A(BcA \supset \Box(A \equiv {\sim}Ic))$, w_1	[3, $\exists$]
(5) $Bc{\sim}Ic$, w_1	[4, PL]
(6) $\forall A(BcA \supset \Box(A \equiv {\sim}Ic))$, w_1	[4, PL]
(7) $\forall x(Ix \equiv \forall A(BxA \supset A))$, w_1	[1, $\Box$]
(8) $Ic \equiv \forall A(BcA \supset A)$, w_1	[7, $\forall$]
(9) Ic, w_1	[Assumption]
(10) $\forall A(BcA \supset A)$, w_1	[8, 9, PL]
(11) $Bc{\sim}Ic \supset {\sim}Ic$, w_1	[10, $\forall$]
(12) ${\sim}Ic$, w_1	[5, 11, PL]
(13) $Ic \,\&\, {\sim}Ic$, w_1	[9, 12, PL]
(14) ${\sim}Ic$, w_1	[Assumption]
(15) ${\sim}\forall A(BcA \supset A)$, w_1	[8, 14, PL]
(16) ${\sim}(BcX \supset X)$, w_1	[15, ~$\forall$]
(17) BcX, w_1	[16, PL]
(18) ${\sim}X$, w_1	[16, PL]
(19) $BcX \supset \Box(X \equiv {\sim}Ic)$, w_1	[6, $\forall$]
(20) $\Box(X \equiv {\sim}Ic)$, w_1	[17, 19, PL]
(21) $X \equiv {\sim}Ic$, w_1	[20, $\Box$]
(22) ${\sim}{\sim}Ic$, w_1	[18, 21, PL]
(23) $Ic \,\&\, {\sim}{\sim}Ic$, w_1	[14, 22, PL]

The individual *c* is either infallible or not in w_1. At step (9), we assume that *c* is infallible in w_1. This leads to a contradiction at step (13). At step (14), we assume that *c* is not infallible in w_1. This also leads to a contradiction at step (23). Accordingly, both assumptions lead to a contradiction. Hence, (1) and (2) cannot both be true in w_0. Since w_0 was arbitrary, we conclude that {(I), (P)} is inconsistent. The inconsistency argument clearly seems to be valid. So, either (1) = (I) or (2) = (P) (or both) must be false. Yet, both (I) and (P) appear to be true. (I) follows from (Def I) and (P) is intuitively plausible. Furthermore, the following argument supports (P). It is conceivable that there is some individual who believes exactly one proposition, namely that she is not infallible. Hence, it is (at least prima facie) reasonable to assume that it is possible that there is some individual of the required kind. This is the paradox of infallibility.[2]

[2] Two anonymous reviewers have challenged this claim. According to the first, it is not so obvious that it is conceivable that there is some individual who believes exactly one proposition, namely that she is not infallible. Such beliefs could not count as rational, according to the reviewer. Similarly, someone could assert 'this sentence is false', but couldn't be warranted to assert it. According to the second, the inconsistency argument shows that the existence of the 'modest believer' (i.e. a subject who believes just one thing, that is, that she is not infallible) is logically impossible. I do not assume that conceivability entails possibility. So, I would still say that the scenario is conceivable, but that this fact does not entail that it is possible. Even if the existence of the 'modest believer' should turn out to be impossible such a believer might be conceivable. If we assume that conceivability entails possibility, we should instead say that the scenario seems to be

Before I turn to the discussion of the possible solutions, I would like to briefly address one possible objection to the way the paradox of infallibility is formulated in this paper (this objection was raised by an anonymous reviewer of the paper). According to this objection, the definition of infallibility plays very little role in generating the puzzle. (P) could be formulated without reference to infallibility, as 'it is possible that there is some individual who believes just one proposition: that something she believes is not true', and still it would have paradoxical consequences. Let us call this sentence (P′). The real problem does not concern the concept of infallibility but the self-referential nature of (P′).

I am in general sympathetic to this kind of view and to the claim that the real problem does not concern the concept of infallibility. In fact, according to the solution that seems most plausible to me, solution 5 below, we can solve the paradox without changing our definition of infallibility. Furthermore, I agree that (P′) is problematic and that (P′) is similar to (P). Accordingly, it is possible that the puzzles generated by (P′) and (P) have similar 'solutions'. Therefore, (P′) (and its paradoxical nature) is interesting on its own. However, (P′) and (P) do not say exactly the same thing and (I) is an essential assumption in our inconsistency argument. Without this assumption we cannot derive a contradiction. Therefore, the paradox of infallibility, as it is formulated in this paper, is not the exact same paradox as the paradox generated by (P′). The paradox of infallibility should be interesting to anyone who philosophises about the concept of infallibility,[3] not only to anyone who philosophises about paradoxes. We want to know if and how we can solve various paradoxes, but we also want to know if the concept of infallibility is consistent or not. The paradox of infallibility is a potential threat to anyone who thinks that the concept of infallibility is consistent; (P′) is not, at least not in itself. The discussion of the paradox of infallibility, as formulated in this paper, should therefore not be replaced by a discussion of (P′) and its paradoxical nature.

Consider, for example, the debate between a classical theist and an atheist. The theist wants to claim that God is infallible. The atheist might respond that the paradox of infallibility shows that the concept of infallibility is inconsistent and that God therefore cannot be infallible. The theist might perhaps respond in the same way as the anonymous reviewer. She might claim that the concept of infallibility plays very little role in generating the puzzle and that the concept of infallibility is consistent. Or again, consider the discussion between an ideal observer theorist in metaethics and a critic. The ideal observer theorist might want to assert that an ideal observer is infallible. The critic might insist that the ideal observer theory is wrong since the paradox of infallibility shows that the concept of infallibility is inconsistent. The ideal observer theorist might perhaps respond in the same way as the anonymous reviewer and try to show that the paradox of infallibility does not establish that the concept of infallibility is inconsistent, etc.

conceivable (even though, in fact, it is not, because it is impossible). I agree that someone who believes in a contradiction cannot be perfectly rational. However, I am inclined to think that it is still possible for someone to believe in contradictions. So, this is not necessarily a problem for the conceivability argument.

[3] This might, for example, include some epistemologists, some (doxastic) logicians, some philosophers of religion and some moral philosophers.

If we reformulate the paradox and drop the concept of infallibility, we cannot understand these kinds of debates.

2. Possible Solutions

Is it possible to solve the paradox of infallibility? In this section, I will consider five conceivable solutions. Personally, I am inclined to believe that the last suggestion is the most promising. However, no proposal is without problems.

2.1 Solution 1

According to the first solution, we should accept dialetheism. According to this theory, there are sentences that are both true and false. If we accept this idea, we might also accept the proposition that it is possible to deduce a contradiction from {(I), (P)} even though both (I) and (P) are true. This might be perfectly reasonable if there are true contradictions.

Still, there are problems with this solution. Dialetheism is dubitable and even if the theory were true, it is not obvious that *every* contradiction is genuine (true). Consequently, even a dialetheist might think that the paradox of infallibility is problematic. Therefore, it seems unlikely that this solution should turn out to be the most plausible overall.[4]

2.2 Solution 2

According to the second solution, we should reject (P) because it is impossible that there is someone who believes that she is not infallible. '$\Diamond\exists x(\mathrm{B}x{\sim}Ix \,\&\, \forall\mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix)))$' entails '$\Diamond\exists x\mathrm{B}x{\sim}Ix$' (this is easy to see since '$\Diamond\exists x(\mathrm{A} \,\&\, \mathrm{B})$' entails '$\Diamond\exists x\mathrm{A}$'). So, if '$\sim\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true (valid), then '$\sim\Diamond\exists x(\mathrm{B}x{\sim}Ix \,\&\, \forall\mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv {\sim}Ix)))$' is true (valid) (again, the proof is easy and can be left to the reader). Accordingly, if we can establish that '$\sim\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true (valid), we may conclude that (P) is false (necessarily false).

According to standard doxastic logic, '$\mathrm{B}c\mathrm{A}$' is true in a possible world w iff 'A' is true in every possible world that is doxastically accessible from w for c. Furthermore, many doxastic logicians assume that for every individual c and for every possible world w there is a possible world w' such that w' is doxastically accessible from w for c, and that if a possible world w' is doxastically accessible from a possible world w for an individual c, then w' is doxastically accessible from w' for c.[5] Suppose that this is correct. Then we can show that '$\sim\Diamond\exists x\mathrm{B}x{\sim}Ix$' is valid in the following way. Assume that '$\sim\Diamond\exists x\mathrm{B}x{\sim}Ix$' is not true in some possible world w_0. Then '$\Diamond\exists x\mathrm{B}x{\sim}Ix$' is true in w_0. Hence, '$\exists x\mathrm{B}x{\sim}Ix$' is true in some possible world w_1. Accordingly, '$\mathrm{B}c{\sim}Ic$' is true in w_1 (where c is some arbitrary individual). By assumption, there is a possible world w_2 that is doxastically accessible from w_1 for c. Consequently, '$\sim Ic$' is true in w_2. By definition, '$\sim Ic$' is equivalent with '$\sim\forall\mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A})$'. Hence, '$\sim\forall\mathrm{A}(\mathrm{B}c\mathrm{A} \supset \mathrm{A})$' is true in w_2. It follows that '$\sim(\mathrm{B}c\mathrm{X} \supset \mathrm{X})$' is true in w_2 (for some arbitrary X). Therefore, '$\mathrm{B}c\mathrm{X}$' is true in w_2 and 'X' is false in w_2. By

[4] For more on dialetheism, see, for example, Priest, Berto and Weber 2018.
[5] For more on doxastic logic, see, for example, Fagin, Halpern, Moses and Vardi 1995 and Meyer and van der Hoek 1995.

assumption, w_2 is doxastically accessible from w_2 for *c*. Hence, 'X' is true in w_2. But this is absurd. It follows that our original hypothesis cannot be true. In other words, it is not possible that there is someone who believes that she is not infallible. It follows that (P) is false (and indeed necessarily false). This solves the paradox of infallibility.

The problem with this solution is that standard doxastic logic only seems to make sense if we assume that we are dealing with perfectly rational individuals. According to orthodox doxastic logic, it is necessary that every individual believes every logical truth. Furthermore, according to the assumptions above, it is necessary that no individual has any inconsistent beliefs and it is necessary that every individual believes that everything she believes is true. It seems very implausible to assume that this holds for *every* individual. So, if we assume that we are quantifying over *every* individual in (P) and not only over perfectly rational agents, it clearly seems to be possible that there is someone who believes that she is not infallible. In fact, there are probably many (*actual*) persons who believe this. And if there *is* someone who believes this, then certainly it is *possible* that there is some individual of this kind. Consequently, our second solution to the paradox of infallibility is quite problematic.[6]

2.3 Solution 3

According to the third solution, we should reject (P) because it is impossible that there is someone who believes only one proposition, namely the proposition that she is not infallible. '$\Diamond\exists x(\mathrm{B}x\sim Ix \,\&\, \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv \sim Ix)))$' entails '$\Diamond\exists x\forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv \sim Ix))$'. Therefore, if '$\sim\Diamond\exists x\forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv \sim Ix))$' is true (valid), then '$\sim\Diamond\exists x(\mathrm{B}x\sim Ix \,\&\, \forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv \sim Ix)))$' is true (valid). Hence, if we can show that '$\sim\Diamond\exists x\forall \mathrm{A}(\mathrm{B}x\mathrm{A} \supset \Box(\mathrm{A} \equiv \sim Ix))$' is true (valid), we may conclude that (P) is false (necessarily false). Why should we believe that it is impossible that there is someone who believes only one proposition, namely the proposition that she is not infallible? Well, according to this solution, we should believe this because it is impossible that there is someone who believes only one proposition, period. We can only have beliefs if we believe many things. To believe anything at all we need a whole web of beliefs. If this is true, we should reject (P). Hence, we can avoid the paradox of infallibility.

Is it true that it is impossible to believe only one proposition? This seems to depend on what we mean by 'impossible'. Perhaps it is historically (temporally) and naturally impossible. But the problem with this solution is that we are not (primarily) interested in these kinds of possibilities in this paper. (P) is supposed to be speaking about logical or metaphysical possibility. And it certainly seems to be logically or metaphysically possible that there is someone who believes on-

[6] An anonymous reviewer has suggested that solution 2 is clearly absurd and that the proper way to reject P is to claim that it is impossible that there is someone who believes that she is not infallible AND that this is her only belief. The falsity of this proposition can be argued on the basis of its self-referential structure, its similarity to the Liar (a belief that is true when false and vice versa). I tend to agree with the general sentiment of this view. If the inconsistency argument is sound (and we assume (Def I)), we must reject (P). The solution that seems most promising to me, solution 5, is similar to the solution suggested by the reviewer. However, solution 5 does not entail that (P) is false.

ly one proposition, even though it is perhaps not historically or naturally possible. If this is the case, we cannot use the third solution to solve the paradox of infallibility.

2.4 Solution 4

According to the fourth solution, we should reject the definition of infallibility (Def I) that we use to derive (I), and if (Def I) is not true (or correct), we have no reason to believe that (I) is true. Therefore, we can also reject (I). Hence, this solves the paradox.

The problem with this solution is that it is difficult to come up with some other definition of infallibility that is reasonable and that does not lead to similar problems. Let us consider one alternative attempt. Instead of (Def I) we should use the following definition of infallibility:

(Def I′) $It =_{df} \Box\forall A(B_t A \supset A)$, where t is any term. For every t, t is infallible iff it is *necessary* that for every (proposition) A, if t believes that A, then A (is true).

According to this definition, no one is infallible if it is *possible* that something she believes is false; it is not enough that everything she believes is true. (Def I′) does not entail (I), but it does entail something similar, namely (I′):

(I′) $\Box\forall x(Ix \equiv \Box\forall A(B_x A \supset A))$. It is necessary that for every (individual) x: x is infallible iff it is *necessary* that for every (proposition) A, if x believes that A then A.

However, if we try to replace (I) by (I′) in the inconsistency argument, it breaks down. So, we cannot use this deduction to show that {(I′), (P)} is inconsistent. Consequently, if we use (Def I′) instead of (Def I) to define the concept of infallibility, we can solve the paradox of infallibility. Intuitively, (Def I′) is even more plausible than (Def I). Hence, the fourth solution seems to be one of the more plausible. Nevertheless, it is not unproblematic, for we can show that (I′) is inconsistent with the following alternative to (P):

(P″) $\Diamond\exists x(B_x \sim Ix \;\&\; \Box\forall A(B_x A \supset \Box(A \equiv \sim Ix)))$. It is possible that there is some individual x such that x believes that it is not the case that x is infallible and it is necessary that for every proposition A, if x believes that A, then it is necessary that A iff it is not the case that x is infallible.

That is, we can prove that {(I′), (P″)} is inconsistent (the argument for this is similar to the inconsistency argument; see above). And (P″) seems to be true. So, even though we can use (I′) to avoid our original problem, we can derive a contradiction from {(I′), (P″)}. Therefore, it is doubtful that this is the best solution to the paradox.[7]

[7] An anonymous reviewer has a strong feeling that the paradox of infallibility has nothing specifically to do with the definition of infallibility. It has to do with truth, and thus indirectly with infallibility defined in terms of true beliefs (see the introduction). I am inclined to believe that this is true, or approximately true. According to solution 5, which seems most promising to me, the paradox of infallibility can be solved without changing the definition of the concept of infallibility in this paper. This solution has to do with the way we should understand propositional quantifiers (and therefore also with self-reference). However, I do not think we should take this for granted and assume that the concept of infallibility is consistent without any discussion.

2.5 Solution 5

According to the fifth and last solution, we should reject the inconsistency argument. It is not necessarily anything wrong with (Def I), (I) or (P), but the deduction is not valid. If the argument for the conclusion that {(I), (P)} is inconsistent fails, then of course we have solved the paradox of infallibility.

But what is wrong with the inconsistency argument? It clearly seems to be valid. The problematic step, according to this solution, is step (11). The universal quantifier cannot be instantiated with any sentence whatsoever. The quantifier in step (10) is a propositional quantifier and in step (11) we have instantiated A with '~*Ic*'. However, '~*Ic*' is simply an abbreviation of '~∀A(B*c*A ⊃ A)' and this sentence includes a propositional quantifier. It is a well-known fact that it is problematic to allow universally quantified sentences to be instantiated with universally quantified sentences when we use ∀-elimination for propositional quantifiers. If we allow such instances, several problematic consequences follow. Consider, for example, the following difficulty. Intuitively, '∀XA' says 'For all propositions X: A'. '∀XA' is true if and only if 'A[B/X]' for every proposition B, where 'A[B/X]' is the result of replacing all free occurrences of the propositional variable 'X' in A by B. Now, let A = ∀XX and assume that our substitution-instances can include any formula whatsoever. Then 'A[A/X] = A', for '∀XX[∀XX/X] = ∀X∀XX = ∀XX'. Hence, the truth-conditions for '∀XX' include '∀XX' itself. That is, to know the truth-value of '∀XX' we must first know the truth-value of '∀XX'. This clearly seems to be viciously circular. In a recursive definition of truth, the truth-conditions for a complex sentence should be defined in terms of simpler sentences. So, there are independently good reasons to suppose that we cannot replace 'X' by any formula whatsoever when we drop the quantifier in a sentence of the following form '∀XA'. We should not replace 'X' by a formula that includes a propositional quantifier.[8] If this is correct, step (11) does not follow from step (10). Hence, the inconsistency argument fails. {(I), (P)} is not inconsistent (or at least we have not seen any reason to believe that it is). Consequently, we can avoid the paradox of infallibility.

This solution seems to be the most promising to me. However, it is not entirely unproblematic. The solution entails that we treat '∀' as a 'substitutional' quantifier that varies over sentences and not as an 'objectual' quantifier that varies over propositions. The paradox of infallibility might still be a problem for everyone who wants to use 'objectual' propositional quantifiers that vary directly over propositions and for everyone who wants the elimination rule for ∀ to be unlimited.

I conclude that we should take the paradox of infallibility seriously.[9]

[8] Some systems of this kind are developed in Rönnedal 2019. For more on propositional quantifiers, see, for example, Lewis and Langford 1932: 178-98, Kripke 1959, Bull 1969 and Fine 1970.

[9] I would like to thank two anonymous reviewers for some interesting comments on an earlier version of this paper.

References

Blackburn, P., De Rijke, M. and Venema, Y. 2001, *Modal Logic*, Cambridge: Cambridge University Press.

Bull, R. 1969, "On Modal Logic with Propositional Quantifiers", *The Journal of Symbolic Logic*, 34, 2, 257-63.

Chellas, B.F. 1980, *Modal Logic: An Introduction*, Cambridge: Cambridge University Press.

Fagin, R., Halpern, J.Y., Moses, Y. and Vardi, M.Y. 1995, *Reasoning About Knowledge*, Cambridge, MA: The MIT Press.

Fine, K. 1970, "Propositional Quantifiers in Modal Logic", *Theoria*, 36, 336-46.

Hughes, G.E. and Cresswell, M.J. 1968, *An Introduction to Modal Logic*, London: Routledge (Reprinted 1990).

Kripke, S. 1959, "A Completeness Theorem in Modal Logic", *Journal of Symbolic Logic*, 24, 1, 1-14.

Lewis, C.I. and Langford, C.H. 1932, *Symbolic Logic*, New York: Dover Publications, Second edition 1959.

Meyer, J.-J. Ch. and van der Hoek, W. 1995, *Epistemic Logic for AI and Computer Science*, Cambridge: Cambridge University Press.

Priest, G., Berto, F. and Weber, Z. 2019, "Dialetheism", in Zalta, N.E. (ed.), *The Stanford Encyclopedia of Philosophy* (2019 Edition), https://plato.stanford.edu/entries/dialetheism/.

Rönnedal, D. 2019, "Semantic Tableau Versions of Some Normal Modal Systems with Propositional Quantifiers", *Organon F*, 26, 3, 505-36.

Book Reviews

Gilmore, Jonathan, *Apt Imaginings: Feelings for Fictions and Other Creatures of the Mind*.
New York and London: Oxford University Press, 2020, pp. x + 258.

Jonathan Gilmore's *Apt Imaginings: Feelings for Fictions and Other Creatures of the Mind* provides readers with insights in three distinct and venerable philosophical topics—fiction, art, and imagination—and addresses the delicate connections between them. It is both a helpful guide for scholars and students interested in the former, and an original perspective on the latter.

The text is extremely reader-friendly. It offers a clear introduction which functions as a helpful guide to navigate the text. In addition, every chapter consists in a careful overview of the debate to which the specific section refers to, so that most chapters can be considered as self-standing pieces. From a stylistic point of view, not only does the text flow very naturally, but it always manages to maintain a balance between the accuracy and richness of the linguistic choices and the accessibility of the content. Word from a non-native speaker of English.

The book explicitly revolves around a fundamental question which haunts both the philosophical and the psychological debate about fiction, that is whether and how our psychological reactions to fictional objects differ from those elicited by real-world experiences. Gilmore phrases this question in a clever fashion, that is by asking if there is *continuity* or *discontinuity* between our engagement with fiction and with our life. Zooming out, one easily sees that he is interested in an even broader and possibly more fundamental task, i.e. finding out if life and art can or cannot be reasonably placed along some sort of continuum.

Ambitious as they may seem, however, these questions—the one about fiction and reality and the one about art and life—are addressed in the book by exploiting the tools of analytic philosophy combined with psychological insights, so that this brave endeavor is immediately presented as a manageable one. Notably and uncommonly in the analytic debate, Gilmore is deeply aware that, in so doing, he runs the risk of depicting a "too dry, too rationalizing, or too abstract" portrait of aesthetically relevant issues. Yet, as he himself acknowledges by quoting Wordsworth: we cannot but murder if we need to dissect (14).

What is particularly convincing in Gilmore's argumentative strategy is its point of departure, which consists in the distinction too often overlooked in contemporary aesthetics between *the normative* and *the descriptive*. In short, while a normative perspective assesses the kinds of reasons that justify our cognitive, emotional, and conative responses to art and life, a descriptive perspective focuses on possible neurological, psychological, and phenomenal explanations of such responses.

Relevantly, it is argued, psychological descriptions of our responses to fiction (and, more broadly, art) do not always respond to those rational norms that allegedly govern their epistemic justification. In Gilmore's own words: "psychology doesn't always respect ontology" (14), i.e. what happens in our mind and brains do not necessarily go hand in hand with the way we want to justify our judgments about what's in the world. Thus, what he names the "continuity question" can be dealt with in different ways or, better, at two distinct levels.

Such a discrepancy allows the author to advocate normative discontinuity between fiction and reality, in spite of the descriptive continuity one might observe when offering psychological explanations of our responses in the two realms:

Argumenta 8, 1 (2022): 199-214
ISSN 2465-2334

DOI 10.14275/2465-2334/202215.boo

> I argue for normative discontinuity: make-believing is not epistemically rational in the same sense as believing, and the norms governing our desires, emotions, and moral evaluations vis-à-vis what is internal to a fiction can be inconsistent with those governing our responses to relevantly similar things in the real world (8).

According to the author, premising the distinction between the two levels of analysis allows the proposed account to avoid, on the one hand, the potential reductionist drifts of continuism, i.e. the idea that experiences of fictional works and real experiences are, all in all, the same thing; on the other hand, it dodges the difficulties encountered by discontinuism in accounting for what our emotional and cognitive responses to fiction and reality have in common.

The structure of the book proceeds by introducing all the ingredients one by one, so as to provide the reader with every necessary tool to understand the recipe. It presents a broadly conceived and relatively uncontroversial *cognitive theory of the imagination* according to which imagination amounts to a cognitive attitude paired with a propositional content. Such an attitude is descriptively continuous with beliefs, yet distinguishable from them on the basis of its functional role within our mental architecture. On this background, the kind of imagination required to engage with fiction is shown to demand for a more specific explanation than the one provided by mainstream cognitive theories. In particular, it is contended that factors which are *external* to our psychological engagement with fictions contribute to determine what and how should be imagined as holding in a fiction. The readers' identification of the story's genre, her awareness of its author's communicative intentions, and her ends as a consumer of fiction "shape which imaginings—of all those that a work might cause—would count as appropriate responses" to that work (30).

The following chapter is devoted to a rather syncretistic theory of emotions. The proposed view tries to combine insights from appraisal-based accounts of affects, with empirically grounded theories of emotions considered as subdoxastic reactions to stimuli. Descriptive continuity between the nature of emotions felt towards fiction and the real world is explained and defended. Yet, paralleling the structure of the previous chapter, the need for supplementing what is available on the emotion-theory market with a more focused account of emotional reactions to fiction (and artworks) is expressed (83).

Once the premises are displayed, the discussion comes into focus as to what normative continuity can vindicate over discontinuity. Chapter 4 is indeed dedicated to individuate the common assumptions of these two perspectives, namely—and unsurprisingly—the fact that they all commit to the role of emotions' *aptness*, that is, the fact that emotional reactions respond to some normative standard. The author acknowledges that normative continuity can account for our tendency to consider people's emotional responses to fiction as revealing of related real-world attitudes. Nonetheless, discontinuity is convincingly defended in Chapter 5 via the argument that aesthetic evaluations have the power to make certain emotions apt towards fictions that would be inapt in real-world situations.

Where the argument becomes subtle and more technical (but that doesn't make it any less accessible), is in Chapter 6, which leverages on the possibility to distinguish between the cognitive or representational component of emotions and their purely affective side. The question raised about emotional reactions in this chapter and for conative, epistemic and moral attitudes in the next two chapters is, roughly: are make-belief attitudes governed by the same rationality we usually

ascribe to belief-involving attitudes? Notably, when this question is applied to conative states, it prompts the challenging issue of mutually inconsistent desires. Instead, when it is reformulated having moral evaluations in mind, it triggers another debated aspect of our experiences of fiction, namely the possibility to produce judgments that are in explicit conflict with what would be considered good or desirable in the corresponding real-world situations.

In each chapter from 5 to 8, the question why the norms that govern our attitudes within practices taking place in real contexts differ from those that govern our responses within artistic (and fictional) contexts is convincingly answered by endorsing a functionalist view. In short: the kinds of reasons that justify our emotional, imaginative, and conative states would depend on the *functions* of the practices in which they are deployed.

Building on the rigorous architecture of the text, the author eventually manages to guide us to the very dense and conclusive Chapter 9. Here, the functionalist account the reader has familiarized with so far, finds its fulfilment in the delicate notion of artistic function. After admitting that artworks of any given kind can have a long and ever-changing series of functions—in light of which they can be evaluated as artworks (9.1)—Gilmore introduces *constitutive functions* as those essential to the artwork's identity (9.2). As such, constitutive functions normatively constrain our responses to art (and fiction), for they offer criteria against which our reactions can be considered apt or inadequate.

***

On the just sketched background, and among the many issues raised by the book, there are two aspects of Gilmore's admirable work that I would like to focus on. The first point is more general and concerns the pivotal distinction between *art and life* once one endorses normative discontinuity. The second, partially related point, has to do with the nature of what the author names *criterial qualities*.

As to the first point, there seems to be a sort of fluctuation throughout the book between the concept of art and that of fiction—and fictional narrative in particular. These two notions appear to play the same functional role in the argument's structure. This is not necessarily problematic as long as, whenever these terms are used, the author introduces examples and contextual elements that allow readers to grasp whether the explanatory weight is on the fictional or on the artistic side. What remains only partially addressed, however, is where exactly we shall draw the line of normative discontinuity between art and real life. As a matter of fact, artefacts belong to a wide variety of categories and so do fictional narratives, so that a sharp distinction might be hard to demarcate. I see three options, each of which is partly consistent with the view presented in the book, but still leads to unwanted consequences.

One first option would be to insist on the much debated distinction between fiction and non-fiction so as to individuate the two domains to which normative discontinuity applies. Take narratives for example. In this reading, there would be normative criteria that apply to fictional narratives and allow us to evaluate the aptness of our responses to such works, while different criteria would apply to non-fictional narratives such as historical reports or documentaries. This is not only reasonable, but also does justice to Gilmore's view that our deep imaginative engagement with fiction, i.e. *absorption*, is governed by specific criteria. However, as a result, we would have to conflate the category of fictional narratives and the category of artworks, as opposed to the category of real life encompassing non-

fictional reports. I am doubtful that Gilmore himself would be happy to take this path.

An alternative solution would be to apply normative discontinuity to the categorical discontinuity between aesthetically valuable and non-aesthetically valuable artifacts. Provided that the class of the aesthetic valuable narratives overlaps with the class of the artistic narratives, this reading would do justice to Gilmore's fundamental project of accounting for the distinction between art and life. Suppose we can take this overlapping for granted: we would have a helpful criterion to place the demarcation line between the norms that regulate the aptness of our reactions to art on the one hand, and those that govern our reactions to non-artistically relevant situations, on the other hand. Yet, this would imply accepting an expensive trade-off. For while granting the discontinuity between art and non-art, we would have to accept that there is normative continuity between our attitudes towards non-aesthetically valuable narratives and real life experiences. Again, this outcome is not unacceptable *per se*, but it does not mirror the difference we normally seize between life and (some of) the ways in which it is narrated. On the contrary, we (and Gilmore) may want to preserve the normative distinction between our reactions to real-life contexts and our reactions to narratives *qua* artifacts.

Finally, if we keep our focus on the case of narrative artifacts, normative discontinuity might apply to the categorical discontinuity between narrative and non-narrative objects. This reading is fully consistent with the claim that—unlike other experiences or ways that give us access to facts—narratives present us their objects from someone else's perspective, in a relatively *opaque* way, exploiting narrative tools in order to emphasize or diminish the importance of certain features, and thereby requiring some specific sort of imaginative engagement (42). But if normative discontinuity applied to narrative versus non-narrative objects, then we would have to treat all narratives as governed by those same norms that govern art—as opposed to real-life experiences. Nonetheless, narratives seem to trigger our responses in a specific way which differs from the way we engage with other kinds of artefacts—namely through absorption and the subsequent deployment of a certain perspective. This makes me think that the demarcation line between different normative domains cannot be drawn without consequences, thereby leaving the question open for further clarifications.

The second aspect that I found particularly challenging and worth discussing is Gilmore's notion of what he names *criterial qualities*, understood as those evaluative properties that emotions, desires, and moral evaluations present their objects as possessing (46). As we have seen, *Apt Imaginings* endorses a form of moderate cognitivism for affects, according to which an emotion should be responsive to the presence (or the absence) in its intentional object of the related criterial qualities. The same holds, with due specifications, for all evaluative attitudes. Moreover, following Peter Lamarque,[1] the argument takes narrative representations to be opaque, in that events, characters, and states of affairs are inherently constituted by the way in which they are described and narrated. Therefore, evaluative properties depend for their instantiation on the perspective from which they are introduced by creators and accessed by recipients: from within or from without the narrative; from a first or third person perspective; from the point of view of this or that character. In sum, they are inherently *perspectival features*.

[1] Lamarque, P. 2014, *The Opacity of Narrative*, Lanham: Rowman & Littlefield.

This amounts to saying that evaluative attitudes (emotions, desires, moral evaluations) are crucial for accessing and engaging with (especially fictional) narratives. Only if we feel, desire something for or about, or judge what is being narrated, we can actually grasp the perspectival features that constitute a work. However, this inherent perspectivalness seems to characterize evaluative properties of works independent of their being instantiated by fictional or non-fictional narratives. That is, all narratives are more or less opaque and present us with a perspective on facts, rather than with facts themselves. And all narratives require that we adopt the corresponding evaluative attitudes in order to get absorbed and to pick up their specific features.

The worry is thus whether this speaks in favor of normative *continuity* between—at least—fictional and non-fictional narratives. For this suggests that the same criterial qualities we grasp in evaluative attitudes not only systematically correlate with similar psychological mechanisms underpinning evaluative attitudes towards both fiction and non-fiction (in line with the descriptive continuity defended in the book), but that they are also the *criteria based on which we evaluate the aptness of* our evaluative responses to both fiction and non-fiction.

Should this concern be well-founded, then one may simply reply that, rather than with a radical discontinuity, we are dealing with a graded scale of opacity. Such a scale would ideally range from evaluative features instantiated by fictional narratives—independently of the corresponding features in the real world—to evaluative features presented within a more transparent perspective, typically instantiated by non-fiction.

Or maybe the reply can be found in the last chapter, titled *Artistic functions.* Although Gilmore declares that his "appeal to functions in this chapter is detachable from [his] arguments for discontinuity" (202), such an appeal to artistic functions as what actually—metaphysically—identifies artworks, their criterial qualities and therefore the way we are expected to respond to them, hints at a viable route. If we define artifacts in general and artworks in particular based on their intended purpose, then normative discontinuity can be grounded in the kind of artifact that is in front of us. This avoids the problems stemming from the definitions based on intentional attitudes. On this view, for instance, imaginings would be *apt* only as long as the metaphysical structure of a work prescribes them, the same holding for emotions, desires, and moral evaluations. Thus, despite the tangible effort made by the author to provide an extremely balanced account throughout the book, some readers can eventually be tempted to look for more radical solutions, whose seeds are already present in the text. And this is an extra merit of this work.

University of Eastern Piedmont MARTA BENENTI

Torbjörn Tännsjö, *Setting Health-Care Priorities: What Ethical Theories Tell Us*. New York: Oxford University Press, 2019, pp. xiv + 212.

The problem of setting priorities for the allocation of healthcare resources is one of the most pressing issues in today's discussion on bioethics and one on which different ethical theories have relevant contributions to offer. Normative ethical theories pursue a systematic evaluation of several relevant considerations to establish which allocative decisions are justified, when not all medical needs can be

satisfied. Tännsjö's book provides a careful discussion of the most important theories in the field of distributive justice and an examination of their consequences in the context of healthcare allocation. Specifically, the book concentrates on four main theories, namely utilitarianism, the Rawlsian theory, egalitarianism, and prioritarianism; its main conclusion is that, assuming levels of well-being as the common currency for comparing them, the practical consequences of these theories for healthcare allocation tend to converge. All theories, that is, strongly suggest redirecting healthcare resources from the present attempt to provide marginal life extension to elderly and terminally ill patients to catering for the needs of mentally ill patients. However, Tännsjö concludes by noting that such a practical convergence is likely to prove ineffective in real life; in fact, the power of human irrationality is such that the good reasons offered by philosophical theory will very likely be disregarded.

Tännsjö starts his discussion from utilitarianism, or "the view that we ought to maximize the sum total of happiness in the universe" (9). Happiness is traditionally conceived as the surplus of pleasures over pains, and Tännsjö has famously defended such a classic, hedonistic account.[1] Moreover, he has also defended the traditional act-centered version of utilitarianism, and this is his favored approach also in this contribution, which completely eschews the hypothesis of rule-utilitarianism. Tännsjö is well aware of the several problems that have been raised concerning utilitarianism's capacity to deal with distributive issues and particularly to take into account the 'separateness of persons'. Nonetheless, he believes that utilitarianism is overall the most defensible normative theory and that some of its consequences which have been strongly criticized are defensible: for example, its ageistic implications, according to which the duty to maximize the sum total of happiness leads to a general prioritization in favor of the younger, or the fact that we should not consider starting points when distributing happiness, but simply go for the highest increase in happiness, even if this will benefit those who are already better off.

The view that the worst off deserve special priority in the distribution has been much popularized by Rawls' theory of justice as fairness, which Tännsjö calls the 'maximin/leximin theory'. According to this theory, everyone should be granted fair equality of opportunity, and distributive inequalities are justified only to the extent that they favor the worst off. Tännsjö blames this view for three main faults. For one thing, the 'maximin/leximin view' is much more ageist than utilitarianism. Conceiving early death as the worst possible outcome, this view justifies prioritizing younger people over elderly ones in all situations, including those in which the elderly may benefit more from being treated. This, of course, would not be allowed by utilitarianism. Moreover, the Rawlsian theory does not take people's suffering seriously enough. In fact, in comparing the respective claims of different individuals, it concentrates on their entire lives: therefore, it tends to downplay the present severe suffering of someone who overall is comparatively well off, prioritizing the less urgent needs of some other patient who has had a less happy life. Finally, it does not consider the ability of different individuals to transform resources into happiness; therefore, it allows that some people who are among the worst off (and therefore deserve priority) and who perform badly in that transformation drain most of the available resources without achieving any considerable happiness. In other words, the maximin/leximin theory may lead to

[1] Tännsjö, T. 1998, *Hedonistic Utilitarianism*, Edinburgh: Edinburgh University Press.

the existence of individuals who may be called 'utility thieves', because they would cause a tragic waste of resources.

Another approach to justice that departs from the maximizing attitude of utilitarianism is egalitarianism. This view derives from the equal value of persons the paramount value of equality in distribution; since equality is *per se* valuable, egalitarians tend to accept that more equal distributions are better, even if they require a sort of leveling down, that is, an action that makes someone less well off without bettering the position of anyone else. The point of this theory is that anyone's condition cannot be defined in absolute terms but is always relative to the conditions of other people. Accepting the common currency of happiness, we can say that, for egalitarianism, the value of an outcome for every individual must be weighted with a factor indicating her position on a scale that evaluates all-life happiness. The fact that disadvantaging someone for the sake of equality is better in one respect does not commit the egalitarian to any actual leveling down; this can be avoided if the egalitarian also takes into consideration the value of the sum total of happiness and refrains from equalizing when such sum would be thereby reduced. According to Tännsjö, egalitarianism has the advantage, compared to the Rawlsian theory, to limit prioritization of the worst off to the extent that this realizes equality; however, it shares the drawback of disregarding the importance of suffering, since it will give priority to alleviating the mild suffering of someone who is comparatively worse off over sparing intense suffering to another one who is better off.

The fourth theory discussed by Tännsjö is prioritarianism. This is the approach according to which what matters in distributive decisions is the absolute level of happiness or well-being that each is experiencing. People who are particularly unfortunate have a privileged claim to resources to enhance their situation; however, according to this view the fact that the condition of some individual B is worsened in no way confers a benefit to another individual A who is less well-off unless the resources taken away from B are given to A. Prioritarianism does not accept an unconditional priority for the worst off, but suggests that there is a limit to the quantity of utility that can be sacrificed to ameliorate their predicament: it justifies choosing a modest increase of the happiness of the worst off over a slightly larger increase of the happiness of the better off, but it does not sanction the loss of a large amount of happiness for the well off for the sake of a small increase of happiness for the worst off. According to Tännsjö, prioritarianism is as much plausible as it does not put excessive weight on the increments of happiness of the worst off, that is, it is plausible inasmuch it does not sacrifice too much utility. Moreover, the best version of the theory is the one in which the assignments of weights is made by reference to the present conditions of happiness, and not taking into account entire lives. In fact, if we compare the levels of happiness of entire lives, prioritarianism becomes extensionally equivalent to egalitarianism, and shares the egalitarian (and Rawlsian) insensitiveness to suffering. Overall, Tännsjö's idea is that prioritarianism should be conceived of as an amendment to utilitarianism: the theory "urges us to maximize a weighted sum total of happiness, where utilitarianism urges to maximize a sum total of happiness as such" (50).

Tännsjö believes that all four theories are at least highly plausible; however, there are powerful objections that can be raised against the maximin/leximin theory and egalitarianism. While utilitarianism has its difficulties as well, it is nonetheless the most defensible view; prioritarian amendments to it have intuitive appeal, but in most cases they should be discarded. At all events, Tännsjö suggests that, from a practical point of view, all the accounts discussed suggest that we

cease to raise the number of resources invested in healthcare, and that, within the healthcare budget, we redirect resources from the attempt to provide marginal life extension to elderly and terminally ill patients to cater for the urgent needs of mentally ill patients. At the same time, he suspects that this theoretical convergence is likely to prove irrelevant in practice. For one thing, all sorts of compromise will be necessary for implementing the theories in real life, and the result may be a convergence on the actual practice; secondly, even when the reasons for diverging from the actual practice are clear and accepted by everyone, most people will not comply with the suggested policy: not because the theoretical requirements are practically impossible to implement, but because of the superior power of human irrationality. The conclusion is a bit desolate: "in real life, all attempts at radically changing the system of health-care rationing in direction of what is demanded by ideal theory, be it the maximin/leximin theory, egalitarianism or utilitarianism (with or without some added prioritarian concern), are in most cases bound to fail" (199).

This book is an important and timely contribution to an urgent debate in practical ethics. It offers an in-depth analysis of several normative theories and insights into their practical consequences. It also provides rich information on relevant facts concerning issues such as assisted reproduction, mental illness, orphan drugs, or research on cognitive enhancement. In what follows, I wish to put forward three critical considerations, concerning the characterization of prioritarianism, the prospects for convergence among the theories, and the relationship of moral philosophy to actual practices.

As for the first point, it can be suggested that the prioritarian approach should be considered as an amendment to egalitarianism, rather than to utilitarianism.[2] The basic idea of prioritarianism, as originally suggested by Nagel,[3] is not the maximization of a weighted sum total of happiness, but the urgency of providing help to people who are badly off. This idea incorporates a reading of the egalitarian account that differs from 'the equity view', according to which we should give to everyone equal chances of getting a certain benefit (e.g. some medical treatment), from those forms of 'telic egalitarianism' that consider equality as *per se* valuable, and also from the maximin/leximin unconditional prioritization of the worst off. The specificity of the solution offered by prioritarianism lies in the attempt to take into consideration both the urgency of protecting people against basic forms of suffering and unhappiness and the requirement to provide for effective use of the available resources.

According to prioritarianism, the fact that someone is comparatively worse off is a *pro tanto* reason to prioritize in her favor. This poses limits on utilitarian aggregation because it prevents conferring small benefits to a very large set of better off people rather than bestowing a significant benefit on one worse off individual; however, it also differs from other forms of egalitarianism because it does not conceive of inequalities as inherently bad, and therefore does not accept any leveling down, nor any unconditional prioritization of the worst off. The view requires that a reasonable balance be struck between providing relief to people who

[2] For this suggestion, see for example Fleurbaey, M., Tungodden, B., and Vallentyne, P. 2009, "On the Possibility of Nonaggregative Priority for the Worst Off", *Social Philosophy and Policy*, 26, 1, 258-85.

[3] Nagel, T. 1978, "The Justification of Equality", *Crítica: Revista Hispanoamericana de Filosofía*, 10, 28, 3-31.

are particularly in need and maximizing the benefits obtained by the available resources. Prioritarianism's concern for those who are most suffering makes it come close to egalitarianism and makes treating it as an amendment of utilitarianism doubtful; for the theory does not attribute value only to the quantitative increase in welfare, but also to the fact that the benefits fall on individuals in dire need. In other words, prioritarianism presupposes a pluralistic account of value, contrary to the monistic conception defended by utilitarianism.

As for the second point, it can be suggested that Tännsjö's confidence in the theoretical convergence of the theories is far from warranted. To give just an example, Tännsjö's main suggestion is that all theories would converge on redirecting our resources to the treatment of people with chronic illness or disabilities. However, utilitarianism's endorsement of QALYs may justify that, in a case where competition between two individuals exists for the use of a single medical resource, priority be given to an otherwise healthy patient over a patient who suffers from a chronic disease if the former will allow a larger increase in QALY; moreover, the theory will generally suggest investing in the treatment of acute diseases, that are amenable to complete recovery, rather than of chronic ones, in which the therapeutic output will always be less than optimal. On the contrary, both maximin/leximin and egalitarianism will favor treating the patient who is initially worse off and will suggest granting more resources to chronic diseases, even if this may not yield optimific results. And, if my suggested interpretation of the theory is right, prioritarianism would side with these egalitarian perspectives, at least in many cases. Similar differences in practical conclusions may emerge, I suggest, if we consider other topics, such as the non-therapeutic uses of assisted reproduction or biomedical research on cognitive enhancement.

Finally, it seems to me that Tännsjö's thesis concerning the ineffectiveness of moral philosophy in bringing about changes in actual practices may be questioned. Just think of the profound modifications that the stress posed by philosophers and lawyers on such topics as autonomy and informed consent has caused in the patient-physician relationship in the last decades. The efficacy of the bioethical discourse in granting a pivotal role to patient's autonomy and in introducing the very notion of fair distribution of resources into the medical domain is difficult to deny. If Tännsjö's prediction nonetheless seems to make sense, it is probable that the failure of attempts to redirect medical resources must not be ascribed to human irrationality, but to a different cause; namely, to the influence of other ideas opposing the conclusions of the theories we examined. A decisive influence may be exerted, for example, by the equal chances view (one that the author excludes from the set of the most plausible approaches), according to which everyone, irrespective of age and prognosis, has a right to receive all the healthcare resources that may obtain even a marginal life extension. If this is true, the reason why elderly people do not want to let go of their lives may not be that they fail to bring their intentions and motivations in line with their theoretical beliefs, but that they are not convinced that giving up on marginal life extension is right and fair, to begin with; and the reason for this may lie in philosophical ideas that have achieved an exaggerated (if probably implicit) success in contemporary medicine. Attempting to change such deep-seated ideas is perhaps one of the most essential contributions of philosophical reflection in the present situation.

Università Vita-Salute San Raffaele, Milano MASSIMO REICHLIN

Formato, Federica, *Gender, Discourse and Ideology in Italian*.
London, Palgrave Macmillan, 2019, pp. XX + 299.

1. A Summary

In *Gender, Discourse and Ideology in Italian*, Federica Formato provides an insightful examination of how Italian gendered expressions are used to convey and reinforce a specific gender ideology. In what follows, I briefly summarize the five chapters of the book.

In the first chapter, Formato argues why it is especially interesting to focus on Italy when studying gender. In particular, she argues that Italy is a *fruitful epistemological site*, namely "a physical space that can provide interesting and investigation-worth data" (1). Indeed, Italy is characterized by the cult of physical beauty and by a language with heavy gender marking that makes it easy to box people in fixed gendered categories, coupled with language ideology that represents a patriarchal cultural system and society.[1]

The second chapter is devoted to analyzing gender in Italian. Following Hellinger and Bußmann,[2] Formato identifies Italian as a language with grammatical gender and provides a taxonomy of Italian nouns according to their behavior with respect to gender. She identifies four categories: *lexical gender*, *morphological gender*, *syntactical gender*, also defined in Marcato and Thüne,[3] and *nouns with no gender*. Formato then briefly discusses the notions of androcentrism and sexism in language. She draws on Sara Mills' distinction between direct and indirect sexism, only the first being visible in linguistic forms.[4] According to Formato, Italian, being a grammatical gender language, is characterized by direct sexism. She then provides a further taxonomy, based on the previous, but adapted "to explain how the manipulation of grammatical forms can occur in practice" (54). This classification explores how the nouns, from all the categories presented above, are used with respect to social gender, namely "the extra-linguistic factors that speakers (and writers) generally associate with, and attribute to, language on the basis of gender, e.g. stereotypes, and, commonly held views of the societal and cultural arrangements of women and men" (50). Importantly, in this taxonomy, Formato distinguishes different usages of the generic masculine. In Italian, the masculine is not only used for men but also for mixed-gender groups, regardless of the proportion of members of each gender. For example, 'alunni' (pupils) in 'gli alunni sono 70% ragazze e 30% ragazzi' (pupils are 70% girls and 30% boys) is masculine. Formato labels these instances *versatile masculines* and she distinguishes them from *impersonal* and *personalized masculines*, where the masculine is used to indicate a general person and to express impersonal and personal experiences. Examples of these usages are 'non è venuto nessuno' (no one [masc] came), 'uno dovrebbe

[1] Formato adopts the following characterization of *patriarchy*: "the subtle but accepted and promoted way to institutionalise the inferiority of women and their subordination within a 'male as a norm' order (through several linguistics and non-linguistic practices)" (7).
[2] Hellinger, M. and Bußmann, H. 2001, "Gender across Languages: The Linguistic Representation of Women and Men", in Hellinger, M. and Bußmann, H. (eds.), *Gender across Languages*, Vol. 1, Amsterdam: John Benjamins Publishing Company.
[3] Marcato, G. and Thüne, E.M. 2002, "Gender and Female Visibility in Italian", in Hellinger, M. and Bußmann, H. (eds.), *Gender across Languages*, Vol. 2, Amsterdam: John Benjamins Publishing Company.
[4] Mills, S. 2008, *Language and Sexism*, Cambridge: Cambridge University Press.

leggere le notizie' (one [masc] should read the news), and 'un sindaco ama la sua città' (a mayor [masc] loves their city), respectively.[5] Moreover, it is common to find the masculine form of certain (usually prestigious) job titles, even when the person holding the role is a woman. Formato refers to these instances as *unmarked masculines*. By distinguishing these cases, Formato equips us with a powerful tool to understand how the masculine works in Italian and what sexist assumptions each usage relies on.

In the third chapter, Formato discusses the proposals of language reforms that have been put forward for Italian. She focuses in particular on the famous *Il sessismo nella lingua italiana* (1987), written by Alma Sabatini under the commission of the Italian Minister of Equal Opportunities, summarising Sabatini's recommendations to avoid sexist expressions.[6] She then surveys more recent publications on the topic and guidelines for gender-fair language use issued by various institutions. Formato also presents her own research concerning speakers' attitudes towards gendered terms in Italian. The research was conducted through an online survey and revealed a widespread misunderstanding of how gender marking works in Italian. Concerning the use of feminine job titles, respondents held diverse standpoints, heavily influenced by their personal experiences. Based on these results, Formato suggests that for a language reform in Italian to be successful it should be promoted not by women alone and has to be systematic.

The fourth chapter explores how gendered language is used to talk about women in public spaces, especially in politics. Here Formato presents a corpus analysis of the gendered forms of 'sindac-' (Mayor) to talk about three women running for Mayor in 2019. The corpus consisted of three high-selling Italian newspapers from three months before to six months after the elections. Semi-marked forms as 'la sindaco' (the [fem] mayor [masc]) and 'candidata sindaco' (candidate [fem] mayor [masc]) are especially frequent before the elections, while the feminine form 'sindaca' is increasingly used after the voting for the two elected candidates. The forms also vary depending on the newspaper considered and on the candidate. The latter might, according to Formato, indicate that journalists took into account each candidate's preferences when choosing which gendered form to use. Formato further presents an analysis of sexual terms and innuendo used to attack women in politics. According to Formato, sexual terms and innuendos "constrai[n] female politicians within a social domain of sexual activities, therefore, foregrounding their (factual or alleged) private life" (164). These include explicit and implicit reference to sexuality and sex work, as the use of "prostitute" (prostitutes) or "orgasm" to refer to female politicians and to the Italian Parliament with female MPs, respectively. These expressions, thus, are used to undermine women's right to have a role in politics by reducing them to their bodies and suggesting they belong to the private rather than to the public sphere. Formato closes the chapter with an investigation of how first-person plural forms are used to construct the group of women in parliamentary debates concerning equal opportunities. She concludes that women Members of Parliament, aware of the uncertainties concerning their role, used these forms to seek legitimization and visibility within the Parliament and the various political parties.

[5] Following a common linguistic convention, I use 'masc' for 'masculine' and 'fem' for 'feminine' in square brackets after a term's translation to indicate its grammatical gender in the source language.

[6] Sabatini, A. 1987, *Il sessismo nella lingua italiana*, Presidenza del Consiglio dei Ministri.

Finally, the fifth chapter concerns gendered violence and how it is addressed in the media. Formato discusses several positions according to which gendered violence is a device that works to keep the current gender order usually enforced by men that perceive a threat to their manhood towards women that defy traditional gender norms. Formato focuses in particular on gendered violence that results in murder, which in Italian is called 'femminicidio' (femicide), and on how it is addressed in Parliamentary acts and the media. She presents three studies on the word 'femminicidio' conducted on different corpora: parliamentary acts on gendered violence, newspapers coverage of femicides between 2013 and 2016, and newspaper articles concerning Sara Di Pietrantonio's femicide. For the first corpus, Formato discusses the *word list*, namely the most frequent words in the corpus; *keywords*, namely the words that in the target corpus are significantly more frequent than it would be expected by chance; and *multi-words* that could stand for 'femminicidio', namely the expressions consisting in multiple words that were used in place of 'femminicidio'. Based on these data, Formato concludes that femicide in Parliamentary acts is 'seen within a broader understanding of violence as an abstract phenomenon and its main focus is on women (as victims) rather than men (as perpetrators)' (219). She bases the analysis of the news corpus on *keywords*, *multi-words*, and *collocations*, namely words that are more likely to occur together with another word than otherwise. For example, the closest and most frequent word that collocates with 'gelosia' (jealousy) is 'per' (for), which, according to Formato, indicates that jealousy *caused* the killings. The results show a strong weight being placed on the *reasons* to kill, characterizing the action as sudden and uncontrollable, rather than on the killers' responsibility. Moreover, a systematic difference emerges in how men and women are referred to, the former through their professions, thus foregrounding their public position, while the latter mainly through their roles within relationships, hence restricting them to the private realm. Based on these observations, Formato concludes that women are treated asymmetrically even in the private sphere, where they have traditionally been relegated. The third and final corpus is analyzed with specific attention to how Sara Di Pietrantonio is referred to and to how much agency she is given in newspapers articles. The results are in line with findings concerning the news corpus. In particular, the victim was mainly referred to through her first name, a strategy which signals proximity, and described according to her young age, which, according to Ulrike Tabbert, could "emphasise [the person's] vulnerability and innocence" (Tabbert 2016: 76, cited in Formato 2019: 257).[7] Moreover, the sentences where Sara Di Pietrantonio was the doer described her breaking up with the killer or talking about the relationship they had. Hence, her agency was always depicted with respect to her relationship with the murderer and not as autonomous and independent of him.

To sum up, *Gender, Discourse, and Ideology in Italian* constitutes an original dissertation of how gender is realized in Italian and of the ideological stances its use is based upon. Formato used various research methods, from a theoretical analysis of gendered nouns to a survey to corpus linguistic studies, and she investigated different loci where gender unfolds, from noun usage to Parliamentary debates to newspaper articles. The topics studied are also diverse: they span from

[7] Tabbert, U. 2016, *Language and Crime: Constructing Offenders and Victims in Newspaper Reports*, London: Palgrave Macmillan.

grammar to how women are addressed in the public sphere, in particular in politics, and in the context of femicide. As any interesting work, Formato's book raises questions. I will consider those that came up to me in the next section, along with a couple of criticisms and suggestions.

2. A Few Observations

As mentioned above, I will devote this section to a few remarks and questions about Formato's work. I start with three minor comments concerning the taxonomies provided in chapter two. Within the category of syntactical gender, Formato distinguishes *epicene* from *semi-epicene* nouns, where the latter are invariable for gender in the singular only while the former in the plural as well. She writes that epicenes end with '-e' in the singular and with '-i' in the plural, and semi-epicenes end with '-a' in the singular and '-i' (for the masculine) and '-e' (for the feminine) in the plural.[8] However, most of these nouns have derivational suffixes and while some epicene nouns do end with '-e'/'-i', as 'preside' (headteacher) and 'giudice' (judge), most of them are derived from present participles and end with '-ente' or '-ante', as 'cliente' (client) and 'vigilante' (custodian). Similarly, most semi-epicenes end with the derivational suffixes '-iatra' and '-ista', like 'pediatra' (pediatrician) and 'fiorista' (florist), while only a handful terminates with '-a', like 'atleta' (athlete), 'profeta' (prophet), and 'stratega' (strategos). From a linguistic point of view, it is more accurate to indicate the whole derivational suffixes and not simply the last vowel.

My second remark concerns the label 'nouns with no gender', which Formato uses for those terms that do not reflect their referent's gender. I find this label deceiving as these nouns do have grammatical gender, as all Italian nouns do. For example, 'persona' (person) is grammatically feminine and 'individuo' (individual) is grammatically masculine. The peculiarity of these terms is that they can refer to individuals of any gender regardless of their grammatical gender. To me, 'nouns with no gender' incorrectly suggests that these terms are genderless, while "nouns with their own gender", "with stable gender" or "with insensitive gender" seem to better capture the distinctive trait of these terms without misleading the reader that they lack gender altogether.

The third minor point also regards a label belonging to Formato's second taxonomy: 'versatile masculines', which indicates masculine plurals referring to mixed-gender groups. For example, the masculine plural 'alunni' (pupils) is used for a group of pupils of various genders, regardless of the proportion: even if the boys are 30% and the girls 70%, they would be referred to by the masculine form. According to Formato, this convention shadows women as it does not reveal whether any woman belongs to the group. Indeed, the plural masculine is also used for all-male groups and the presence of one man in a mixed-gender group is sufficient to require the masculine form. However, even though Formato herself deems these instances problematic, the label 'versatile masculines' has a positive flavor as it expresses flexibility, which is usually appreciated. A different name, like 'overextended masculines', might highlight that using masculine forms for

[8] I here only talk about masculine and feminine not because I believe these are the only two genders for human beings, but because these are the only options in Italian. While other languages, such as German or Greek, have also a third gender, the neuter, Italian only has two grammatical genders: the feminine and the masculine.

mixed-gender groups is unwarranted and problematic and it would better express the author's own critical stance towards these forms.

I now turn to more substantial remarks. The first one concerns another category of the second taxonomy, that of *unmarked masculines*. These are masculine terms used for women, as 'Ministro' (Minister) in 'Lucia Azzolina fu Ministro dell'Istruzione' (Lucia Azzolina was the Minister of Education), where the masculine 'Ministro' refers to a woman, Lucia Azzolina. This usage is common for certain (usually prestigious) professions and it is usually considered sexist by feminist scholars and activists as it conveys that men usually occupy such roles and women who do are an exception. However, one might hold that these professional titles are actually *nouns "with no gender"*: their grammatical gender does not reflect the referent's gender. They are grammatically masculine but can refer to individuals of all genders, just like 'individuo' (individual). While I do believe this is not the case and professional titles like 'Ministro' (Minister) are not *nouns "with no gender"*, the author does not discuss this possibility. Consequently, she does not provide reasons why such professional titles should reflect the referent's gender and occur in the feminine when referring to a woman. To hold that unmarked masculines are sexists, Formato should rule out the hypothesis that they are in fact *nouns "with no gender"*. One possibility is to argue that professional titles like 'Ministro' (Minister) should reflect their referent's gender because that is usually the case with professional titles. However, there are a few exceptions to this generalization: the job titles 'guida' (guide), 'sentinella' (sentinel), and 'spia' (spy), for instance, are grammatically feminine regardless of their referent's gender. There is, though, a linguistic reason why these terms are always in the feminine: they derive from the expression 'fare la guida/sentinella/spia' (to act as a guide/sentinel/spy), where 'guida', 'sentinella', and 'spia' in Italian do not designate a person, but an action, which is in the feminine. This explanation, however, does not apply to 'recluta' (recruit), which can nonetheless be accounted for on linguistic grounds: it is a loan word from Spanish, and loan words are kept unaltered. Other *nouns "with no gender"* that are not job titles exist in one gender only for similar reasons. For instance, 'testa' (head), as in 'la testa dell'organizzazione' (the head of the organization), is grammatically feminine, regardless of its referent's gender, because it is a catachresis. 'Personaggio' (character), on the other hand, is always masculine for the same reason 'recluta' (recruit) is feminine: it is a loan word and it retains the grammatical gender it has in the source language, namely French. To summarize, the argument could be that professional titles like 'Ministro' should reflect the referent's gender as it is usually the case for Italian personal names and there are no linguistic reasons why these job titles should constitute an exception.

Formato further discusses *semi-marked forms*, namely job titles composed of one masculine and one feminine element, such as 'medico donna' (lady [fem] doctor [masc]) or 'signora medico' (Mrs. [fem] doctor [masc]). Formato writes that "semi-marked forms do not occur for men" (61). While I agree that no one would say 'medica uomo' (male [masc] doctor [fem]), it seems to me that *semi-marked forms* could be used for men with feminine job titles "with no gender" as 'guida' (guide) and 'spia' (spy): 'guida uomo' (male [masc] guide [fem]) does seem acceptable to me.

As mentioned in the book summary, one of Formato's most interesting analyses concerns the distinction between *versatile masculines* and *impersonal* and *personalised masculines*. That is, she distinguishes between masculine forms used to

refer to mixed-gender groups, as in 'gli alunni sono 30% ragazzi e 70% ragazze' (the pupils [masc] are 30% boys and 70% girls), and to indicate a generic or unknown person, as in 'uno dovrebbe leggere le notizie' (one [masc] should read the news) and 'qualcuno ha perso questo' (someone [masc] lost this). Formato, hence, provides a fine-grained analysis of various usages of the masculine in Italian beyond referring to men. However, it is not clear where certain usages fall into: if the masculine refers to a generic or unknown mixed-gender group, would it be a *versatile masculine* or an *impersonal* or *personalised masculine*? Take for instance 'dall'anno prossimo i genitori avranno un congedo più lungo' (starting next year, parents will have an extended parental leave), where the masculine 'i genitori' (the parents) refers to generic individuals of arguably different genders. It is not clear to me which of Formato's categories it should belong to. I see three possibilities: the current categories could be defined to make clear where cases like this belong to, or the taxonomy could be supplemented with a further category to account for these instances, or, again, the categories might not be mutually exclusive and these usages fall in their overlap.

In closing chapter two, Formato briefly discusses the difficulties of avoiding binarism in Italian. As she observes, certain strategies, like substituting gendered suffixes with an asterisk, are only possible in writing but cannot be used in speech as they lack a corresponding sound. She then writes that "Italian counts 5 vowels (*a*, *e*, *i*, *o* and *u*) and *-u* would be the only one that could be introduced as neutral" (73). Even if she does not state it explicitly, the focus on vowels depends on the fact that Italian terms end with vowels. Hence, an alternative suffix has to be constituted by vowels to maintain word structure and prosody. While Formato correctly observes that '-u' is the only non-gendered vowel in the Italian repertoire, she does not point out that two vowels would be needed: in Italian, the plural is not marked adding an '-s' to the singular form but with a distinct suffix, typically constituted by a different vowel. For instance, the plural of 'persona' (person) is 'persone', where the singular ends with '-a' and the plural with '-e'. Thus, gender-neutral word endings in Italian should be composed of one vowel for the singular and another one for the plural, but four out of five vowels are already employed for gendered suffixes. Besides Formato's concerns over speakers' resistance to such innovation, then, I would add that the question is complicated even from a mere linguistic point of view.

The most problematic point of Formato's book, to me, concerns the term 'femminicidio'. She writes that "[t]he term relates to women killed by their ongoing or former partners, husbands or boyfriends" (200). However, Formato quotes other sources that provide the more appropriate definition of femicide in terms of *motive for murder* rather than of killer-victim relationship. According to the Parliamentary Committee on Gendered Violence, cited on pages 206-207, "proper femicides" are murders of women "in which the motive for the crime is that of gender" (Impact Assessment Office 2018: 8).[9] Michela Murgia, cited on pages 249-250, writes that '*femminicidio*' "indicates the reason why they have been killed. [...] *Femminicidi* only concern women who are killed because they rejected the expectations of how women should behave set by both men and a patriarchal

[9] Impact Assessment Office 2018, *Femicide: The final report of the first Italian Joint Committee of Inquiry, Data and Statistics*, Senate of the Italian Republic, https://www.senato.it/application/xmanager/projects/leg18/UVI_Focus_Femicide_1.pdf [accessed 20 February 2022].

society".[10] Importantly, these characterizations are not equivalent to Formato's: if we take into account *motive*, a husband killing his wife to inherit her money does not qualify as femicide, while a man killing an unknown woman because she defies gender roles does. Having (had) an intimate relationship with the victim is neither sufficient nor necessary for a murder to be femicide. On pages 246-247, Formato reports as femicide a murder committed by a man that had never had a relationship with the victim: a janitor killed a teacher because she did not reciprocate feelings that he had not even expressed her, and with whom he had no relationship. It seems, then, that Formato fluctuates between the two understandings of femicide without realizing their difference.

Finally, I raise a concern on the analytic framework Formato employs to investigate the news corpus reporting femicides. The author categorized the newspaper headlines into three groups: *blame on the killer*, *a male perspective*, and *blame on the woman*. She includes in the last category headlines focusing on "[w]omen reporting the killers to police prior to the *femminicidio*" (242, table 5.13). However, it is not obvious to me that reporting the victim pressing charges against the killer necessarily puts the blame on the woman. On the contrary, it might show that the victim, far from being blameworthy, tried to prevent the murder. It can also be a way to expose institutional inaction. The details of each headline seem crucial in determining what that focus conveys.

To recap, I suggested a few different labels and I sketched an argument as to why professional titles should reflect their referent's gender unless contrasting linguistic reasons. I also pointed out that Formato's taxonomy may have overlapping or unclear categories or has to be supplemented and that introducing gender-neutral endings in Italian would require more vowels than the repertoire contains. Finally, I objected to Formato's characterization of femicide and raised a concern on her analytical framework to investigate news headlines. These remarks notwithstanding, *Gender, Discourse, and Ideology in Italian* is an insightful and rich book that provides an analysis of gender in Italian across different domains.

University of Genova MARTINA ROSOLA

[10] Murgia, M. 2016, "Dillo che sei mia. La trappola fatale dell'immaginario" in *Un altro genere di rispetto* (blog), https://unaltrogeneredirispettoblog.wordpress.com/2016/11/23/dillo-che-sei-mia-la-trappola-fatale-dellimmaginario-di-michela-murgia/ [accessed 20 February 2022].

Advisory Board