Polskie Towarzystwo Semiotyczne

STUDIA SEMIOTYCZNE

Tom XXXIV • nr 2

PÓŁROCZNIK

INFERRING TRUTH AND MEANING

Warszawa • 2020

Redakcja językowa:
Martin Hinton

Skład elektroniczny:
Dominik Dziedzic

Adres redakcji: Krakowskie Przedmieście 3, 00-047 Warszawa
e-mail: studiasemiotyczne@pts.edu.pl
http://studiasemiotyczne.pts.edu.pl/

ISSN 0137-6608; e-ISSN 2544-073X

CONTENT

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 5–8
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.01
From the Editors

MARTIN HINTON [*], PIOTR STALMASZCZYK [**]

PREFACE

The disciplines of general philosophy, philosophy of language, and linguistics have in common an interest in saying what it is that we can infer: what meaning, what truth; and how those inferences are to be justified. To do this, philosophers and linguists have endlessly discussed the concepts of truth and of meaning, and also the means of inference and its degrees of reasonableness and reliability. These debates do not narrow down to definitive answers, rather they broaden and spread their concerns into ever-widening fields of investigation. One of those areas thriving now as a result of the combination of insights from philosophical and linguistic research is the theory of argumentation; and it is a particular goal of the editors of this collection that the authors of those insights be brought together with researchers studying argumentative discourse for the mutual benefit of all. The papers collected in this special issue of *Studia Semiotyczne* all contribute further to these continuing discussions and to this aim: they exhibit a wide range of approaches and starting points, which may take readers to territories unfamiliar, and, we trust, stimulating; yet they are united by the desire to explore the connections between truth, meaning, and reasoning, by looking at language and all that it carries with it, unbeknownst to the humble conversationalist.

The authors whose work is gathered in the following pages were brought together at the 6th International Conference on Philosophy of Language and Linguistics (*PhiLang 2019*), held in Łódź, Poland, in May 2019, and organized by the Department of English and General Linguistics, University of Łódź. Incorporated within this meeting was a workshop dedicated to the Philosophy of Argu-

[*] University of Łódź, Department of English and General Linguistics. E-mail: martin.hinton@uni.lodz.pl. ORCID: 0000-0003-0374-8834.

[**] University of Łódź, Department of English and General Linguistics. E-mail: piotr.stalmaszczyk@uni.lodz.pl. ORCID: 0000-0002-1407-7610.

mentation (*PhilArg*), and the first three of the authors presented below were participants in that event. The conference has a long and rich history of publications (http://filologia.uni.lodz.pl/philang/archive) and we trust this special issue will prove a valuable and significant addition to that library of work.

The papers are organized within this issue in a way that sees a progression into increasingly abstract concerns: beginning with discussion of patterns of inference in practical reasoning, moving through studies of the nature of language and meaning, and finally into the consideration of the concept of truth. The first article, *Slippery Slopes Revisited* by Martin Hinton, contains a discussion of the reasoning pattern known as the slippery slope argument, generally considered to be fallacious. Hinton attempts to show where earlier characterizations of the argument form have gone wrong by trying to unify arguments which share only superficial features, and points out that there must be something distinct and unique about the reasoning employed in such arguments if slippery slopes are to be considered a type of argument and not simply a rhetorical device. This involves a strong criticism of Douglas Walton's account of slippery slopes, in particular. In the second part of the essay, Hinton finds the special nature of slippery slope arguments in their evocation of logical, rather than material, consequences, leading to an impossibility to prevent other, unwanted and unacceptable conclusions being made. The paper ends with a description of how this treatment of slippery slopes fits into his broader framework of argument assessment, instantiated in the Comprehensive Assessment Procedure for Natural Argumentation.

The second contribution to the issue also touches upon fallacy theory, leading to fundamental questions about the relationship between formal logical fallacies and the reasonableness of everyday human practices of inferring from evidence. Richard Davies writes persuasively *In Defence of a Fallacy*; the fallacy in question being the deductively indefensible error of affirming the consequent. Davies provides a detailed scholarly analysis of how the concept of fallacy develops in the work of Aristotle, and how discussion of affirming the consequent in modern accounts relates to that earlier foundation. Finally, the author analyses the examples of *epomenon* put forward by Aristotle in the Sophistical Refutations and finds them to be cases of abductive reasoning, similar to those which we employ continually in normal life, and consider quite respectable. This conclusion brings into stark relief the difference between logically sound inference and reasonable practical inference, bringing yet further doubt onto traditional conceptions of fallacy.

The third work which deals with argumentation, Cristina Corredor's *Speaking, Inferring, Arguing. On the Argumentative Character of Speech*, turns more explicitly towards the relationship of inference with language. She argues, contrary to some other approaches, that while speech is an inferential activity, language is not inherently argumentative. The main interest in the study is the degree to which meaning can be said to be dependent on argumentation if communication is based on inferring. This involves the careful examination of three major theories: Grice's account of communicated meaning, Brandom's normative pragmatics and Anscrombe and Ducrot's notion of radical argumentativity. The conclusions reached from this are that communication is an inferential activity due to its calculability, since meaning is reconstructed through inference; that arguing can be seen as the practice of evaluating reasons given to justify what has been communicated; and that the obligations assigned through speech acts are dialectical in character; but that this does not entail that language is itself argumentative.

The three remaining papers discuss various issues in the semantics/pragmatics interface, interpretation, use/meaning distinction, meaning ascriptions, truth and Kantian pragmatism.

The notions of speaker's reference and semantic reference were introduced by Kripke in order to counter the contentious consequences of Donnellan's distinction between the referential use and the attributive use of definite descriptions. Palle Leth argues in his paper that these notions do not have any application in the interpretive interaction between speaker and hearer. This is the case because hearers are solely concerned with speaker's reference: either, in cases of cooperation, as presented as such by the speaker, or, in cases of conflict, as perceived as such by the hearer. Any claim as to semantic reference is irrelevant for the purposes of communication and conversation. In conclusion, Leth observes that if the purpose of semantic theory is to account for linguistic communication, there is no reason to take definite descriptions to have semantic reference.

According to the quotational theory of meaning ascriptions, sentences like "'Bruder' (in German) means brother" are abbreviated synonymy claims, such as "'Bruder' (in German) means the same as 'brother'". Andrea Raimondi argues against the quotational theory of meaning ascriptions. He first discusses a problem with Harman's version of the quotational theory, next he presents an amended version defended by Hartry Field and addresses Field's responses to two arguments against the theory that revolve around translation and the understanding of foreign expressions. Finally, Raimondi formulates two original arguments against both Harman's and Field's versions of the theory. One of them targets the hyperintensionality of quotations, and the other raises a problem pertaining to variant spellings of words.

The last paper investigates the notion of truth and Kantian pragmatism. According to Jürgen Habermas, each class of statements raises a distinct validity claim (namely, that of truth, rightness or truthfulness). And each must be justified in a discourse, a special sort of dialogue, in which the validity claim is directly questioned and its justification is required; this validity claim and its relationship to Kantian pragmatism is an important topic in Habermas's theory of communicative action, explicitly discussed in *Truth and Justification*. Tomoo Ueda concentrates on Kantian pragmatism (as interpreted by Habermas) and the anti-deflationist account of truth. He observes that Habermas's notion of truth relies on the reliabilist conception of knowledge rather than the internalist conception that defines knowledge as a justified true belief. Ueda's interpretation is consistent with Habermas's project of weak naturalism and strongly suggests that Habermas's Kantian pragmatism counts as a pragmatist project. The author also draws some more general implications about the pragmatist notion of truth.

Taken together, we believe that this collection of papers provides a stimulating overview of some key current concerns in the fields of argumentation, linguistics, and philosophy of language, in particular the role of inferring in both reasoning and understanding. We wish to thank all the authors and the reviewers who have made this issue possible, as well as all those who attended *PhiLang 2019* and took part in the discussion around these papers and the issues they raise.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 9–24
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.02
Article

MARTIN HINTON [*]

SLIPPERY SLOPES REVISITED[1]

SUMMARY: The aims of this paper are to illustrate where previous attempts at the characterisation of slippery slope arguments (SSAs) have gone wrong, to provide an analysis which better captures their true nature, and to show the importance of achieving a clear definition which distinguishes this argument structure from other forms with which it may be confused. The first part describes the arguments of Douglas Walton (2015) and others, which are found wanting due to their failure to capture the essence of the slippery slope and their inability to distinguish SSAs from other consequentialist forms of argument. The second part of the paper puts forward a clear analysis of what is special about SSAs: it is argued that all SSAs, properly so-named, claim that reaching a certain conclusion, A, involves the negation of a thitherto accepted principle, P, and that that principle is necessary to argue against further conclusions (B, C, …, Z) which are considered unacceptable.

KEYWORDS: Slippery Slope, Douglas Walton, argument schemes, Periodic Table of Arguments, CAPNA.

1. Introduction

The Slippery Slope argument (SSA) is, in itself, a rather slippery customer. A very similar group of arguments is known under a variety of different names: “the thin end of the wedge” and “the camel’s nose” being the two best known; and, as if that were not problem enough, a great many different forms of argumentation have been considered as slippery slopes by various scholars. The only

[*] University of Łódź, Faculty of Philology. E-mail: martin.hinton@uni.lodz.pl. ORCID: 0000-0003-0374-8834.

[1] This article revises and refreshes arguments made in my (2018) and brings my thinking into line with broader views on argumentation expressed in my (2021).

thing which everyone seems to agree on is that in an SSA it is proposed that one relatively innocuous step will somehow lead to far worse consequences at some point down the line. Unfortunately, there are many ways in which one thing may lead to another, a fact which has prompted Govert den Hartogh to suggest that no uniform description of what is meant by SSA is possible: "If one tried to give a definition covering present usage, one would not come up with any distinctive argument form meriting a separate discussion" (1998, p. 280), a view which is backed by Lode (1999).

In this situation, the argumentation scholar is faced with a choice: either accept that SSAs form a nebulous concept, the further analysis of which is unlikely to lead to a single conception, or attempt to "clean up" the use of the name with a more precise definition and show why the uses of SSA outside the bounds of this definition are erroneous, since the arguments being referred to by that name would be better characterised differently. Douglas Walton, however, believes that he can find a third way: to identify a common strand to all apparent SSAs. His attempt to do this is described in some detail below, and the criticism offered reveals the impossibility of his task: the moment the definition process begins, certain forms of the SSA are pushed to the fringe and others taken as more paradigmatic, with no justification offered other than that those kept in the centre fit the new paradigm best.

Some scholars have pointed to two main categories of SSAs, the logical and the empirical. Anneli Jefferson describes the empirical version as the "most common variant" (2014, p. 671), though with no supporting evidence, and gives detailed discussion of the two types and instances of their use. She does not, however, give any particular reason why the two different argument structures she discusses should be grouped together as SSAs, other than that convention would have it so. Among those writers who have sought to make the term more precise and narrow the range of arguments accepted as cases of SSA, the common theme has been to require that slippery slopes must have logical, argumentational consequences. Thus, Rizzo and Whitman argue that: "first and foremost, slippery slopes are slopes of arguments […] They involve intellectual commitments that, as it were, take on a life of their own" (Rizzo & Whitman, 2003, p. 541). This element is key if SSAs are to be distinguished from other forms of argument based merely on the unpleasant material consequences of an action. If the SSA is to be distinct, interesting, and deserving of a categorisation of its own, it must be based on the idea that the first step on the slope commits the actor to accept the further steps, or at least prevents him from being able to rationally oppose them, not merely that by acting once he sets off a chain reaction of bad consequences. The scheme for SSAs which I offer in a later section of this paper (see also Hinton, 2018) makes clear the mechanism by which this commitment is made, establishes the logical SSA as the best candidate to be considered a truly distinct form and dismisses the so-called empirical SSA as an example of the argument from material consequences, with no claims to be treated differently.

2. Walton's Scheme

The principal difficulty for researchers in properly defining the SSA is not its varied use so much as the necessity to show that, despite appearing in all traditional fallacy lists, its use is not, in fact, always fallacious. If we accept that SSAs are always poor arguments, then we might simply argue that the fallacy lies in claiming that one thing follows from another without offering any evidence as to how or why that might happen, and relies on inspiring in others an irrational fear of the final catastrophic consequence at the end of the chain. In a chapter entitled *Fallacies and Unfair Discussion Methods*, Harrie de Swart dismisses the SSA claiming: "One makes a slippery slope argument when one takes several related ideas and inappropriately makes a generalization about them all" (2018, p. 494). This is, ironically, a somewhat inappropriate generalisation, and can be said to represent at best an old-fashioned view of SSAs and of fallacies in general, lacking in nuanced discrimination.

Walton, however, agrees with the trend towards regarding SSAs as potentially acceptable forms of argument, making it clear that in his paper:

> [I]t is argued that slippery slope arguments can be reasonable in some instances […]. But as one looks through the literature on slippery slope arguments, it is difficult or even impossible to find a single example of one that meets all the requirements for being a reasonable argument. (Walton, 2015, p. 284)

Walton offers a definition of the SSA and does tentatively propose an example of it in action which "appears to be […] reasonable" (2015, p. 285).[2] As mentioned above, Walton believes he has identified a common theme in all SSAs, namely the "gray area caused by indeterminacy, typically arising from vagueness, on a continuum in a contemplated sequence of actions [and the] loss of control combined with this indeterminacy" (2015, pp. 279–280). The similarities between SSAs, on this line of thinking, and the sorites paradox are obvious. A sorites paradox, also known as the paradox of the heap, occurs when we try to pin down vague terms in order to make distinctions. Since the removal of one grain of sand does not turn a heap of sand into a non-heap, it appears that, no matter how many times one removes a grain, the heap is still there, even when only one grain remains. This lack of a clear cut-off point is reflected in some SSAs concerned with abortion, for instance: since there is no distinct moment at which a fertilised egg becomes a human child, it might be argued that allowing the termination of a zygote commits us to allowing the termination of a zygote plus one day, then plus another day and so on until all unborn children are vulnerable; the moment of birth providing an obvious distinction. It is, however, far from clear that all SSAs, or even the most typical examples, actually involve the concept of vagueness. By basing his definition around the idea of the "gray area",

[2] Both the characterisation and the example are repeated in Walton (2017).

Walton is immediately putting at risk his project of providing a characterisation suitable for all types of SSA, as included in current usage.

In his earlier book on the subject, Walton (1992) identified four categories of slippery slope; one of which was those involving vagueness, another those which set a precedent, a third those based on causal mechanisms, and the fourth, the full slippery slope, in which all are combined and an element of changing public opinion added. He admits, however, that this work "fails to identify the core features common to all slippery slope arguments, and therefore does not provide a central definition that applies to all slippery slope arguments" (Walton, 2015, p. 274). This is the role, then, that the grey area is supposed to play. The lack of a core definition, however, was not the only criticism the book received. In his review, Wibren van der Burg notes that Walton allows chains of consequences to count as slippery slopes, citing an example where the pollution of a river leads to the death of much wildlife and a danger to humans. Van der Burg objects: "something is missing. I would suggest that it is essential for a slippery slope that it is not merely a sequence of events, but a sequence of actions […] This is merely a negative argument from long-term consequences" (1993, p. 224).

At first glance, Walton appears to have taken this criticism to heart, and he makes a point of stressing that his scheme allows us to distinguish between SSAs and arguments from consequences, noting that: "We usually think of slippery slope arguments as built around a connected sequence of actions and consequences starting from an initial action or policy and then proceeding through a sequence to an eventual outcome" (Walton, 2015, p. 282). There is a difficulty here, though, in the understanding of "action". Unless an action is the result of a rational process, it is hard to see what the difference is between actions and events in this context. This problem is brought out by the example of a supposedly reasonable SSA which Walton eventually gives.

In Walton's example, a father, Bob, is advising his daughter, Alice, not to experiment with narcotics. He points out to her that while such drug use is associated with gratification, it soon leads to dependency and then into the nightmare of full addiction, with all its terrible effects, from which it is very hard to escape. In this case, it is clear that the grey area refers to the point at which the body begins to become dependent on the drug, and it is that dependence which causes the loss of control, making it increasingly difficult to halt the slide. What is less clear is the degree to which the taking of more drugs by a person sliding into dependency qualifies as an "action". It isn't a rational decision—the person affected knows that he should stop, that it would be reasonable to stop, but carries on anyway under the influence of the chemicals in his brain. I would argue that, although it looks different because it features a human agent, the reasoning here is no different from the polluted river example. The drug taker does not make any intellectual commitments, he is not committed to arguing that taking more at each stage is the right or even a reasonable thing to do. From first try to full addiction is presented as a chain of consequences which would need to be supported by empirical evidence that, in fact, experimentation with drugs does frequently lead to addiction. The example is presented as good because it matches the definition, but it does not offer any support for that definition because it is obviously very different from some of the classic cases of SSAs being employed in debates on social policy and medical ethics.[3]

An unconvincing example does not, of course, invalidate the entire characterisation. Walton lists as many as ten basic characteristics of the SSA, several of which are worth questioning. The first few set the scene relatively uncontroversially, but number 4 runs: "There are factors that help to propel the argument and series of consequences along the sequence" (2015, p. 287). What these factors are is unclear, but an explanation is given later: "The factors referred to in characteristic 4 are called drivers. A *driver* is a catalyst" (2015, p. 288). So SSAs are propelled by "factors" which are "drivers" which are "catalysts". The exact nature of this force does not seem to be an important element in Walton's scheme, and yet it is vital to the way the argument proceeds: if an argument is driven by material cause and effect, it is a very different beast from one driven by force of logic, as I explain more thoroughly below. He does give some examples: "Drivers include such factors as precedent, social acceptance, vagueness and technological change" (Walton, 2017, p. 1518). The members of this group make very odd bedfellows: it is hard to see how a principle like precedent could operate in the same way as a fact like technological change.

A second major issue comes at points 5 & 6. Walton writes: "At the beginning of the sequence the agent retains control of whether to stop moving ahead. […] However during some interval along the sequence of actions, the agent loses control of the possibility of stopping from moving ahead" (2015, p. 287). This has been foreshadowed by earlier talk of the grey area, but here it is made explicit that the first part of the sequence occurs under the control of the agent, that is to say, that the first step is not necessarily slippery at all, and could be reversed, and the second step might not actually happen. This runs contrary to both how SSAs have usually been understood and how they are actually used: it also robs them of all their persuasive power. If the first step does not necessarily lead to the second step, then where is the force of the warning, which is explicitly based on the danger of the first step? It appears that Walton has twisted the meaning of the SSA to fit the theory rather than shaping the theory to the argument. All logical SSAs are, by the force of their logic, slippery at once—there simply is no area of uncertainty where control is lost—control has gone from the very first step. It would make no sense for logic to suddenly kick in half-way down the slope. This applies equally to arguments from precedents (which I do not consider SSAs, see below): once the precedent is set, there is immediately no control over how it will be followed. Walton acknowledges that any argument where the initial action appears to already belong in the grey zone apparently contravenes

[3] There is an abundance of such literature and an entire chapter in *Fallacies in Medicine and Health* by Louise Cummings, where she claims "Of all the informal fallacies used in medicine and health, none is more prominent than the slippery slope argument" (2020, p. 65).

the scheme, and offers the bizarre defence that “as the slippery slope argument is stereotypically used, it is not meant to advise the proponent not to take the initial step for the reason that even at this initial step she might already lose control” (2015, p. 303). This is an empirical claim for which no evidence is given and at once dismisses all logic based arguments as not slippery slopes. My intuition would be the exact opposite of Walton’s: SSAs are employed, for example by opponents of gay marriage or abortion, to argue against the first step, and it is the danger inherent in that first step which they stress, not some future loss of control.

This difficulty is re-affirmed in the last of the characteristics, where Walton writes: “The critic argues that the agent should not take the first step, because if she does, she will be led to unpredictably lose control, and then will be unable to avoid the catastrophic outcome” (2015, p. 288). This seems wrong for two reasons: firstly, warning someone that “she will be led to unpredictably lose control” seems a rather wishy-washy kind of an argument, and certainly doesn’t fit SSAs based on logical consistency or precedent setting; and, secondly, it is hard to imagine what kind of evidence might be offered in support of the argument. Clearly, there is no logical force linking step 1 to the catastrophic end point, since step 2 can still be avoided, so some empirical data would be required; but what kind of empirical data would show that you will unpredictably lose control at some point in the future? It would have to be the kind of data which supported the first step as a likely first link in a chain of causes leading to that outcome as a consequence. That would then be a simple argument from consequences.

Indeed, looking more closely at Walton’s characterisation, it seems to have been written specifically to fit the drug abuse example, and clearly doesn’t apply to a lot of arguments which scholars have wanted to include under the SSA umbrella. Drug abuse, however, is a very special case, where the agent falls into a trap and ceases to act reasonably. There are initial similarities with arguments against “designer babies” which state that small genetic changes now would eventually lead to greater ones and the coming of a generation of super-humans. The key difference is that the would-be baby designer needs to be convinced that those simple procedures would lead to something more, not that starting such procedures would render society incapable of rational decision making later on, which is the case with narcotics.

In his choice of example, Walton shows how he has conflated the idea of a metaphorical “slippery slope” in life, where things gradually get worse and worse: getting into debt, getting older, losing touch with loved ones, substance abuse; with slippery slope arguments. Simply saying “don’t do it—it’s a slippery slope” is not the same as employing an SSA.

The original 2015 paper did not feature a full list of critical questions (CQs) for SSAs, but one does appear in Walton’s (2017). There are five CQs in all (2017, p. 1524):

CQ1 What intervening links in the sequence of events A_1, A_2, …, A_i needed to drive the slope forward from A_0 to A_n are explicitly stated?

CQ2 What missing steps are required as links to fill in the sequence of events from A_0 to A_n, to make the transition forward from A_0 to A_n plausible?

CQ3 What are the weakest links in the sequence, where additional evidence needs to be given on whether one event will really lead to another?

CQ4 Is the sequence of argumentation meant to be deductive, so that if the first step is taken, it is claimed that the final outcome A_n must necessarily come about?

CQ5 Is the final outcome A_n shown to be catastrophic by the value-based reasoning needed to support this claim?

All of these questions may provide interesting information about a particular argument, however, with the exception of CQ5, they are clearly not critical questions in the usually accepted sense—the sense which Walton himself employs elsewhere. Critical questions are such that they must be answered correctly for an argument to be accepted, but CQs 1–4 do not lead to any kind of evaluation or assessment of the argument in question. As Yu and Zenker point out, CQs are "argument attacks or rebuttals" (2020, p. 16) which may target the data, the inference or the conclusion of an argument; yet CQs 1–4 above, while asking for more information about the argument, do not target anything at all: they are questions, but they are not critical.

The shortcomings of Walton's treatment are important because of the influence which his work, quite deservedly in general, has over the field of argumentation and beyond. His approach has been criticised by Hinton (2018) and Strait & Alberti who state that it "tends toward including arguments that are not really SSAs (e.g., the heap paradox), but also […] excludes arguments that should not be excluded" (2019, p. 1088), as well as being completely ignored by Philip Devine, who finds, quite independently, that "the argument has three forms—analogical, argumentative and prudential" (2019, p. 375), without any reference to Walton. Yet, for many authors it remains authoritative and definitive.

For instance, Louise Cummings (2020) considers SSAs to be of particular importance in questions of medical ethics, and she discusses them at length. She explains confidently that there are four logical features to them: avoidance of negative consequences, progression through interlinked actions, drivers propel series of actions, and that they are defeasible, presumptive arguments. She refers to all SSA-style arguments as logical in order to distinguish them from the metaphorical, and yet the fourth of her features would seem to rule out the possibility of what are usually known as logical SSAs. All of this is supported only by a passing reference to Walton (2017), although her section on evaluation makes no mention of his CQs.

Similarly, Liga & Palmirani set out to demonstrate how tree kernels can be used "to detect the famous 'Slippery slope' argument" (2019, p. 181). The SSA, it seems, is too famous to require any introduction, let alone explanation, and so they give it none. The single reference provided, but not discussed, is Walton (2015).

At the same time, a large-scale study by Blassnig et al. (2019) which looked at informal fallacies in populist rhetoric, including SSAs, considered only Walton's earlier (2008) definition in their brief description of the form. The authors were apparently happy to treat this definition as complete and uncontroversial, presumably unaware of the criticism it had received and the fact that Walton had later developed his own view considerably.

All of this, the problems with Walton's scheme and CQs, the prevalence of SSAs in important medical decisions, and the readiness of researchers to accept definitions without criticisms, highlights the need for a clearer understanding of what is actually going on in such arguments and a better appreciation of how they might be evaluated.

3. An Alternative Characterisation

The argument scheme offered by Walton, and based on the characteristics described above, is rather long and rather complex, consisting of six premises and a conclusion. All this, in spite of the fact that it only covers a limited set of arguments which are only debatably SSAs anyway. Before setting forth my own, simpler, scheme, there are a few points to deal with in terms of providing a better characterisation of the form of the argument.

The first is this: Walton is right to stress that any definition of the SSA must properly distinguish it from straightforward arguments from bad consequences. If the SSA cannot be so distinguished then it is no more than a vague category used popularly to describe certain situations or arguments with little fundamentally in common—a rhetorical device to produce fear and uncertainty in its audience. Walton does give a simple scheme for arguments from bad consequences—*A* will bring bad consequences, don't do *A*—but doesn't explain why that does not fit his drug example. In fact, this scheme will fit all SSAs, since SSAs are arguments from bad consequences. Those arguments, however, can be divided by the nature of their consequences—they may be logical or they may be material. Arguments from material consequences, no matter how long and twisting the chain of cause and effect, are essentially all the same. Every serious consequence of an action is a result of a chain of very small events, so whether the consequence is immediate or at the bottom of a slope is of no importance to the argument; thus, the so-called empirical SSA is no such thing, it is a simple argument from material consequence, with greater story-telling. Such argumentation must be supported with empirical facts, and, perhaps, probability statistics. Arguments from logical consequences, however, are something very different. The evidence for them is pure reasoning; they do not need to be backed up by science.

At least three distinct varieties of arguments from logical consequences can be identified, all of them relying at heart on the law of the excluded middle, that we cannot assert both p and not p without being guilty of inconsistency. The first is the simple argument from consistency: having made an assertion in one place, I cannot assert its negation somewhere else unless I am prepared to withdraw the original statement. If I insist on both my assertion and its negation, the consequence is inconsistency. Secondly, there are arguments from precedent: these are distinct from SSAs and a not a sub-type of them, since precedents are activated when a situation the same or very similar to the one in question appears, thus, the argument runs that if you act in this way now you are committed to acting in the same way again in future cases resembling this one. SSAs are different because they suggest a commitment to a logical step which has consequences very difficult to foresee. To return to the example of designer babies, it is not important whether or not we can imagine what types of genetic modification might be possible in the future, the thought is that whatever they are, we shall have no defence against them once we accept human modification in principle. The consequences of setting a precedent are clear—if the same situation occurs, we are bound, in fairness, to behave in the same way. The consequences of the slippery slope, however, are not so clear: surely, they are bad, but a certain mystery about just how far down the slope leads is part of their power.

It should also be remembered that in all of these cases of logical consequences, there may be no material consequences whatsoever. Being labelled inconsistent is unpleasant, but not painful; the situation for which one set a precedent may never arise again; the horrors at the bottom of the slope may never actually be realised, but our commitment to accepting them if they are is established at the first step. Arguments from logical consequences are clearly very different from arguments from material consequences, and the difference lies in the force driving the argument. This force in the case of so-called empirical SSAs is the physical concept of cause and effect; in the case of arguments from precedent and SSAs, as I have described them, it is the power of logical consistency. Two arguments with a different inferential force can never be united under one argument scheme, not least because the critical questions one would wish to ask are so very different.

4. An Alternative Scheme

My argument scheme for SSAs, first proposed in my (2018), takes as its starting point the idea that all such arguments involve the defence of a particular principle, which, if broken now, would be unavailable to us when arguing against other proposals which might arise later on. The scheme is laid out below, along with critical questions, and examples of how it is capable of filtering reasonable SSAs from unreasonable ones. Thus, an SSA is an argument which states that:

1. Accepting proposal (a) would mean breaking the hitherto accepted principle (p).
2. Upholding (p) is necessary/important to argue against proposals (b), (c), … (z).
3. Proposal (z) is clearly undesirable.

Therefore, (a) should be rejected.

This scheme can be altered to allow for the establishment of a new principle, hitherto not accepted, rather than the breaking of an old one, with no logical changes. There are three critical questions which this scheme implies:

CQ1 Does accepting (a) break (p)?

CQ2 Is (p) necessary/important in arguing against (b), (c) … (z)?

CQ3 Is (z) undesirable?

The first of these questions is more complicated than it looks since it is quite possible that an apparently broken principle is only being bent a little. For example, when I accept the killing of a man in self-defence, I have not abandoned the principle of not killing men, I have allowed it to be trumped by another principle, and I am not committed to a slippery slope of senseless murder. Also, for a principle to be broken it must be firmly established that it exists and is currently being relied upon. One of the examples below will show that upon deeper thought this may not always be the case.

The second question is where most apparent SSAs will break down. This question tests whether there is a true connection between the case in hand, the first step on the slope, and the other situations or arguments which may come later. It is important to stress that (b) is not a logical consequence of (a), only that objecting to (b), should somebody propose it, has become impossible as a logical consequence of accepting (a). It is to be assumed that when changes are made, they are made one step at a time, but, from a logical point of view, there is already no defence against (z) as soon as (a) is accepted. There is no stipulated length to the slope, and the number of stages to be gone through before the final disaster is reached is a topic on which the literature generally is silent. The question for anyone evaluating the argument is whether or not the removal of the principle at hand does, in fact, preclude any defence against more radical proposals.

Finally, as SSAs must lead downwards, it should be established that what the removal of the principle commits us to accept, or at least leaves us defenceless in arguing against, is, in fact, a catastrophic outcome. It may well be that those who oppose the first step also oppose its logical conclusions, while those who accept it are happy with where it leads. An example of this would be an argument against women's suffrage where a nineteenth century gentleman might have argued that allowing women to vote would commit us, in the long-term, to accept women in parliament and even a woman as Prime Minister! One man's disaster is another woman's progress.

To illustrate how the scheme and the questions work together to assess SSAs and sort the strong from the weak, they need to be applied to examples. Here, there arises a small problem: users of SSAs rarely set them out in full. In order to examine these arguments, then, the theorist must attempt a reconstruction of the thinking behind the argument. While that reconstruction should be done as generously as possible, there is a danger that ideas are being put into the heads of those who never had them. This should be borne in mind during the discussion of the fairly well-known positions described below.

One of the major roles of argumentation schemes and their accompanying critical questions is to allow us to show exactly where arguments which we instinctively feel are weak go wrong, and thus, to be able to properly refute them. An example of this is the somewhat absurd argument employed by certain American organisations, against the legalisation of same-sex marriage. These groups (see, for example, the TFP Student action website) claim that allowing such unions sparks a slippery slope leading to incestuous, paedophilic, or even inter-species marriages. Reconstructing the argument as generously as possible, it runs something like this:

1. Accepting same-sex marriage would mean breaking the hitherto accepted principle (p) that marriage is always and only between a man and a woman.
2. Upholding (p) is necessary/important to argue against incestuous, paedophilic and inter-species marriages.
3. These marriages are clearly undesirable.

Therefore, same-sex marriage should be rejected.

Now, applying the critical questions, we see that this argument does pass the first test: there has long been such a principle in existence in most of the world and it would be broken by allowing same-sex marriages. Most people, I suggest, would also agree that the third premise is correct and the types of union mentioned should not be accepted, so the third question is also satisfied. It is the second critical question, however, which reveals the error in the argument: the principles restricting marriages with children, with close relatives and any other objects, animate or otherwise, exist independently of the man plus woman tradition and are unaffected by its removal. A man is not free to marry his sister, a little girl or a cow, despite their all being female. Premise 2 is demonstrably false: principle (p) is not involved in the arguments against those forms of marriage at all. This example, then, illustrates how the scheme and questions are able to specify precisely where the weakness in the argument lies.

Other arguments appear more persuasive and, rather than exposing their absurdity, the scheme works to find points at which they may be questioned and thus helps the debate progress towards better conclusions. SSAs are often referred to in medical ethics, not least in euthanasia debates (Feltz, 2015; Lewis, 2007; Potter, 2019). Sometimes the term here is used to refer to the material consequences of legalisation, but a logical argument can also be made, that once doctors begin to use their skill to assist those suffering from great physical pain to die, there is little argument to prevent their using it to help those experiencing psychological pain from doing the same thing. Thus:

1. Legalising euthanasia would mean breaking the hitherto accepted principle (p) that doctors always try to preserve life.
2. Upholding (p) is necessary/important to argue against assisted suicide on demand.
3. Assisted suicide on demand is clearly undesirable.

Therefore, legalisation of euthanasia should be rejected.

In this case, all three of the premises are questionable, but none is obviously false. Firstly, although there is no doubt that allowing assisted suicide would break the principle of always preserving life, it is far from certain that that principle is currently in operation. It has become common practice to withhold certain treatments from patients in a terminal condition, as they would only unnecessarily prolong their suffering. This takes us into the distinction between acts and omission and its moral complications, but, without entering such debates, it can be noted that the first premise is, at least, open to question.

The second premise is also worthy of debate. It is hard to draw a clear distinction between what forms of pain, and in what degree, qualify one for a mercy killing and what forms do not. However, it does seem that physicians have an over-riding duty of care to their patients such that any form of assisted suicide would have to be the last resort and in cases such as teenage depression, substance abuse or grief, experience and training would suggest that other methods of alleviating the suffering are possible and should be tried first. Still, once it is legal for a doctor to take life, it becomes a question of an individual's (or perhaps a panel's) judgement as to whether or not the suffering is sufficiently severe and what other methods are worth trying.

Thirdly, although many people would be horrified to find that their local family doctor was assisting patients to kill themselves, there is clearly a libertarian case to be made in favour of freely available access to pain-free, easily administered, life-ending drugs. In short, the bottom of the slope may not seem so bad from a certain point of view.

In this case, then, the scheme helps to pick out which parts of the argument are controversial and require further debate or evidence. Argumentation schemes can be extremely useful in showing those employing certain forms of reasoning what they are actually claiming, giving them the chance to decide whether or not they really want to make such claims and whether or not they have reasonable evidence for them.

5. Slippery Slopes in a Wider Framework

In order to complete the account of the SSA, it is a necessary to state briefly how the structure and its evaluation fit into the fuller theory of argumentation set out in Hinton (2021). In the theoretical underpinning of the Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) introduced in that book, the identification of fallacies through some comparative analysis that finds similarity between an instance of an argument and a defined named-fallacy is discarded, and does not form any part of the evaluation procedure. It is important to reiterate, therefore, that the discussion of slippery slopes in this paper is a discussion of a form of arguing, not the definition of a fallacy, and that my preoccupation has been to show how it may be separated from other forms, in order to make the name meaningful and clear.

The CAPNA itself is a procedure with three main stages of evaluation: of the process, the reasoning, and the language of natural argument. The discussion in this paper, and the CQs given above pertain only to the stage of reasoning analysis, because the SSA is here being considered, in the abstract, as a form of reasoning. Any particular instance of such an argument form would, by necessity, take place within a process and be expressed in language, and would face procedural questions (PQs) on all three levels. Still, it is worth considering for a moment how the CQs for an SSA might differ from the PQs it would face within the reasoning evaluation of the CAPNA.

The reasoning stage is based upon an identified argument type in accordance with the Argument Type Identification Procedure (ATIP) set out by Wagemans (2020). This procedure involves a reconstruction of the argument, where necessary, and the identification of the relationship between the premise and the conclusion through consideration of the subjects and objects of the sentences expressing them. The nature of the statements, whether they be of fact, value or policy is also noted. The analysis is entirely systematic and procedural: at no point is an attempt made to compare arguments to traditionally named structures. The reasoning of every argument can be evaluated in terms of the acceptability of its data premise, and in terms of the strength of the warrant or "lever" which is necessary to reach the given conclusion.[4]

When this procedure is compared to the scheme above, it is clear that the first PQ (or set of PQs), concerning the data premise, is equivalent to CQ1, and that the second PQ (or set of PQs), concerning the lever, is equivalent to CQ2. Which leaves CQ3 apparently unaccounted for. There is, however, a good reason for this. At a more careful level of analysis, an SSA is actually two arguments: one is that accepting (a) will lead to accepting (z), and a second is that (z) is undesirable and any action that leads to it should be rejected. CQ3, then, is equivalent to the data premise PQ of the second stage of the argument.

This insight, derived from the systematic approach to identifying arguments employed by the ATIP and the CAPNA, goes a long way towards explaining the confusion over slippery slopes. The second stage of the argument is the same for all those wide-ranging examples cited in the varied literature on the topic—the consequence is bad so its cause should be avoided, a value statement leading to a policy statement—but the first stage of the argument is different. The PQs which would be asked of a so-called empirical SSA, where both premise and

[4] See Wagemans' (2019) for further theoretical background.

conclusion are factual, would be the same as for any other argument from material consequence. The PQs examining the lever of the SSA as I have described it, however, would be different, because the argument always involves statements of value, thus showing that form to be distinct, and deserving of separate consideration.

This leads to an important realisation: for a slippery slope argument to be distinct from other consequentialist arguments, it must deal with statements of value, of what is true or acceptable, rather than statements of fact. This can be illustrated with the example of drug addiction. The argument that Alice will get addicted to heroin because Alice smokes cannabis, no matter how many intervening steps are placed in between, is an empirical claim about cause and effect, no different from the claim that the water will boil because it is being heated. On the other hand, the claim that Alice's taking heroin becomes acceptable because Alice's smoking cannabis is acceptable, is a claim about values and the logical connection between holding one and holding the other: this, then, is a different form and can be safely referred to as an SSA.

6. Conclusion

In this paper, I have made four basic claims. Firstly, I have argued that current usage of the term slippery slope argument is inexact and covers a variety of forms of reasoning which cannot be treated as a common argument form. Secondly, I have shown how Douglas Walton's attempt to find a common strand amongst these disparate arguments has led him into error via the quite unnecessary positing of a "gray area" in which control is lost. Thirdly, I have suggested that if SSAs are to be examined at all, they must be differentiated from other forms of argument from consequences, necessitating the restriction of the term to those arguments whose consequences are of a logical, argumentational, rather than a natural or material nature. Lastly, I have proposed an improved and greatly simplified argument scheme, with critical questions, and illustrated with examples how it is capable of recognising the flaws and strengths of slippery slope arguments.

In making these claims, I realise that I may be accused of hijacking the term "slippery slope" and re-defining it to meet my own interpretation. To an extent, I acknowledge this to be the case; however, I would argue that in order for distinct argument forms to be described, it is essential to identify their distinguishing features, not only at the level of appearances, but in the workings of their inference. I have shown also how a systematic procedure of analysis can highlight the differences between forms and proves a much better judge of which arguments are alike and which are not than a simple comparative analysis.

REFERENCES

Blassnig, S., Büchel, F., Ernst, N., Engesser, S. (2019). Populism and Informal Fallacies: An Analysis of Right-Wing Populist Rhetoric in Election Campaigns. *Argumentation*, *33*(1), 107–136. doi:10.1007/s10503-018-9461-2

Cummings L. (2020). *Fallacies in Medicine and Health*. Cham: Palgrave Macmillan. doi:10.1007/978-3-030-28513-5_3

Devine, P. (2018). On Slippery Slopes. *Philosophy*, 93, 375–395.

den Hartogh, G. (1998). The Slippery Slope Argument. In H. Kuhse (Ed.), *Companion to Bioethics* (pp. 280–290). Oxford: Blackwell.

Feltz A. (2015). Everyday Attitudes About Euthanasia and the Slippery Slope Argument. In: M. Cholbi, J. Varelius (Eds.), *New Directions in the Ethics of Assisted Suicide and Euthanasia* (vol. 64, pp. 217–237). Cham: Springer. doi:10.1007/978-3-319-22050-5_13

Hinton, M. (2018). Slippery Slopes and Other Consequences. *Logic and Logical Philosophy*, *27*, 453–470.

Hinton, M. (2021). *Evaluating the Language of Argument*. Cham: Springer.

Jefferson, A. (2014). Slippery Slope Arguments. *Philosophy Compass*, *9*(10), 672–680.

Lewis, P. (2007). The Empirical Slippery Slope from Voluntary to Non-Voluntary Euthanasia. *The Journal of Law, Medicine and Ethics*, *35*(1), 197–210.

Liga D., Palmirani M. (2019). Detecting "Slippery Slope" and Other Argumentative Stances of Opposition Using Tree Kernels in Monologic Discourse. In: P. Fodor, M. Montali, D. Calvanese, D. Roman (Eds.), *Rules and Reasoning. RuleML+RR 2019. Lecture Notes in Computer Science* (vol. 11784, pp. 180–189). Cham: Springer. doi:10.1007/978-3-030-31095-0_13

Lode, E. (1999). Slippery Slope Arguments and Legal Reasoning. *California Law Review*, *87*(6), 1469–1544.

Potter, J. (2019). The Psychological Slippery Slope From Physician-Assisted Death to Active Euthanasia: A Paragon of Fallacious Reasoning. *Medicine, Health Care and Philosophy*, *22*, 239–244. doi:10.1007/s11019-018-9864-8

Rizzo M., Whitman, D. (2003). The Camel's Nose in the Tent: Rules, Theories and Slippery Slopes. *UCLA Law Review*, *51*, 539–592.

Strait, L. P., Alberti, L. (2019). The Role of Decision-Making Agency in Distinguishing Legitimate and Fallacious Slippery Slope Arguments. In B. Garssen, D. Godden, G. R. Mitchell, J. H. M. Wagemans (Eds.), *Proceedings of the Ninth Conference of the International Society for the Study of Argumentation* (pp. 1083–1092). Amsterdam: SicSat.

de Swart, H. (2018). *Philosophical and Mathematical Logic*. Cham: Springer.

TFP Student Action. (2015). 10 Reasons Why Homosexual "Marriage" is Harmful and Must be Opposed. Retrieved from: http://www.tfpstudentaction.org/politically-incorrect/homosexuality/10-reasons-why-homosexual-marriage-is-harmful-and-must-be-opposed.html

Van der Burg, W. (1991). The Slippery Slope Argument. *Ethics*, *102*, 42–65.

Wagemans, J. H. M. (2019). Four Basic Argument Forms. *Research in Language*, *17*(1), 57–69. doi:10.2478/rela-2019-0005

Wagemans, J. H. M. (2020). Argument Type Identification Procedure (ATIP)—Version 3. Retrieved from: www.periodic-table-of-arguments.org/argument-type-identification-procedure

Walton, D. (1992). *Slippery Slope Arguments*. Oxford: Oxford University Press.

Walton, D. (2008). *Informal Logic: A Pragmatic Approach* (2nd ed.). Cambridge, NY: Cambridge University Press.

Walton, D. (2015). The Basic Slippery Slope Argument. *Informal Logic*, *35*(3), 273–311.

Walton, D. (2017). The Slippery Slope Argument in the Ethical Debate on Genetic Engineering of Humans. *Science and Engineering Ethics*, *23*(6), 1507–1528.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 25–42
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.03
Article

RICHARD DAVIES [*]

IN DEFENCE OF A FALLACY

SUMMARY: In light of recent developments in argumentation theory, we begin by considering the account that Aristotle gives of what he calls sophistical refutations (*elenchoi sophistikoi*) and of the usefulness of being able to recognise various species of them. His diagnosis of one of his examples of the grouping that he labels *epomenon* is then compared with a very recent account of the matter, which, like Aristotle, calls on us to attribute a mistake or confusion to anyone who uses this kind of argument. From examination of three other examples that Aristotle himself supplies of *epomenon*, it appears that there are cases of inferences of this kind that we need not, and perhaps cannot, avoid making. The suggestion is made that this is because the whole family of what Peirce calls abductions have important characteristics in common with *epomenon*.

KEYWORDS: sophistical refutations, fallacies, affirming the consequent, abduction.

Deficiencies

There has been a significant trend over recent decades to broaden traditional characterisations of fallacies as, for instance, "arguments that seem valid but are not", so as to contemplate many sorts of "deficient moves in argumentative discourse" (van Eemeren, 2001, p. 135). Proponents of this broadening, at least indirectly inspired by Hamblin (1981), such as the argumentation-scheme approach associated with Douglas N. Walton (systematised in Walton et al., 2008) and the pragma-dialectical approach associated with Frans H. van Eemeren (sys-

[*] University of Bergamo, Department of Letters, Philosophy, Communications. E-mail: davies@unibg.it. ORCID: 0000-0002-3866-8757.

tematised in van Eemeren & Grootendoorst, 2003), tend to set out argumentation schemes or rules for the conduct of discussions and indicate that violations of the schemes or infringements of the rules give rise to the deficiencies in question.

This broadening of the definition of fallacy has the merit of capturing several of the moves in argumentative discourse that appear in Aristotle's listing of sophistical refutations in the fourth and fifth chapters of the book that carries that title (hereinafter *SEl.*, in Aristotle, 2016). Thus, *prosodia* (*SEl.*, v, 166b1–9) or *duo erotomata* (167b38–8a16), are not even arguments. Likewise, many instances of *diaeresis* and *synthesis* (*SEl.*, iv, 166a6–38) or *schema tes lexeos* (166b10–19), are not likely to seem valid to most people even if unpicking them calls for some nimbleness.

The broadening also has the merit of keeping front and centre the dialogical setting that is a key to understanding why some sophistical refutations that proceed by way of valid inferences should be counted as fallacies (Rapp & Wagner, 2013). Thus, *to en arche aiteisthai* (*SEl.,* v, 167a36–39, sometimes Latinised as *petitio principii* or Englished as "begging the question", though quite what is at issue would take us too far afield), of which Aristotle does not give even one example in the short chapter that unpacks it (*SEl.*, xxvii; but see *AnPr.*, II, xvi and *Top.*. VIII, xiii), and perhaps *elleipsis tou logou* (*SEl.,* v, 167a21–35, which is sometimes half-Latinised as *ignoratio elenchi*, but might be as well rendered as "missing the definition"), are often deficient moves in debate and hence fallacies, especially when, in the former case, premises are suppressed or covert (Iacona & Marconi, 2005, pp. 33–34).

These merits are not negligible in prising apart the notions of deficiency and invalidity, the latter being a feature of deductive arguments such that the premises can be true yet the conclusion false; but the notion of deficiency in the newer definition of fallacies may itself need a little more elucidation. For instance, it seems that not all debating moves that are deficient by violating the schemes or infringing the rules in the recent approaches will be deficient in the somewhat broader sense of not being expedient tactics in discussion. After all, if one knows that one's interlocutor will not notice that a certain argmentative manoeuvre is deficient by the rules, there may be nothing to advise against its use. If such a move may be effective in embarrassing, confusing or silencing the interlocutor, that may speak in favour of its deployment (Schopenhauer, 1830). This may apply especially in encounters where the interlocutor has already shown themselves unscrupulous: fair play is mandatory where it is reciprocal, but perhaps not otherwise. Even if one is exposed to the risk of being accused of violation or infringement, there may be circumstances in which it is worth running that risk. On the one hand, an explicit accusation, for instance, of infringing the "Validity rule" ("The reasoning in the argumentation must be logically valid or must be capable of being made valid by making explicit one or more unexpressed premises"; van Eemeren, 2002, p. 183) is a very unlikely thing indeed and exposes the accuser to the counter-accusation of pedantry. And the counter-accusation will reveal a deficiency that well-meaning and broad-minded Canadians and Dutchmen with their schemes and rules may be insensitive to. On the other, the accusation of infringing the Validity rule is defeasible because some good arguments and inferences are not logically valid, as we shall see.

Counterfeits

Though it is often thought of as the ninth book of, or as an appendix to, the *Topics* and appears in the traditional order of the Organon at the end of the sequence, it is not wild to suppose that much of the *Sophistical Refutations* dates to an early period of Aristotle's philosophical activity, in the first instance as an instructor in Plato's Academy. If we bear in mind that the text that we have may stand in some relation to lecture notes (whether Aristotle's or some collation of his students'), then we may think of the course or courses into which it feeds or from which it derives as responding to two demands, one theoretical and the other more practical.

The theoretical demand is that of putting some order into the medley of specious reasonings and wordplays presented in Plato's *Euthydemus*. Though Plato may not have had much of a theory about the differences between good and bad arguments, his exhibition of sophisms shows that he was aware that some differences can be discerned and that their perpetrators should be exposed as frauds, which is the objective of so many of the dialogues that target sophists. Aristotle sets himself to schematise such differences and explain their ruses. As the *Topics* promotes the orderly conduct of dialectical debates, so the *Sophistical Refutations*, following much the same scheme (see the correspondences listed in Aristotle, 2007, L–LI), indicates some kinds of argumentative ploys, especially in competitive encounters (*agonistikoi*; *SEl.*, ii, 165b11), that one should be forearmed against and that one should not oneself use, because they have already been exposed as fraudulent.

The practical end of the instruction that Aristotle was presumbly imparting is the cut-and-thrust of the assemblies and tribunals of the Athens of his day (Ryle, 1965a; 1965b). The students in the Academy would later be called on to take part in the public life of the city and would need to know their way about good and bad arguments, so that having some theory for the former and at least some labels for and some practice at recognising the latter would stand them in good stead.

Against this background, it is perhaps not altogether surprising that the *Sophistical Refutations* does not offer more than a generic characterisation of its programme. The first line of the text we have may well be an addition from a later stage in Aristotle's career to bring it into continuity with the bulk of the *Topics* (likewise by common consent a relatively early work), but what it says is that sophistical refutations are those that "seem to be refutations while they are in fact paralogisms and not refutations" (*SEl.*, i, 164a20). *Paralogismos* is a word that Aristotle uses fourteen times in this work and it is very tempting to render it with "fallacy" with the connotations of the broader sense of that word to which we have already adverted: "deficient move in argumentation" rather than the narrower "argument that seems valid but is not".

In addition to the reasons that have been adduced for saying that invalidity is not a necessary condition for being a paralogism or a sophistical refutation (e.g. Hansen, 2002, p. 143–145), we may note that, at the time of composing the *Sophistical Refutations*, Aristotle did not have at his disposal a perspicuous or well-defined notion of invalidity as we (may) have come to understand it. Indeed, it is only a slight exaggeration to say that, even in the later operation of building the theory of the syllogism in the *Prior Analytics*, Aristotle did not have at his disposal a perspicuous or well-defined notion of validity.

He reproposes the definition of a syllogism—literally a putting together of words—that we find in the *Topics* (I, i, 100a25–7) at *Prior Analytics* I, I, 24b19–21 saying that it is a reasoning (*logos*) in which, given certain things, something other than them follows from them of necessity; but the notion of "following" (*symbainein*) is not much worked out. In building his theory, Aristotle distinguishes between a figure, later known as Barbara, that is "perfect" or "complete" (*telaios*: the latter English rendering in Aristotle, 1989) and the figures that need to be perfected or completed by or reduced to one that is; those that do stand in that need (whatever it may be: see Striker, 1991), such as Barocco, are nevertheless valid by our lights and in the terms of the given definition of a syllogism. When Aristotle wants to say that, from some combination of things given, a certain other thing does not follow of necessity, he says simply that there is no syllogism (*ouk syllogismos*; for instance, *AnPr*., I, iv, 26a8, 11–12, 32, and 37, 26b3, 10–11, and 17–18; v, 27a19, 27b3, 13, 23, and 36–7, vi, 28a32, 28b3–4, 22–3, 32, and 36–7, 29a9), and in the whole of the *Prior Analytics*, he uses *paralogismos* just once (II, xvi, 64b13).

Nevertheless, for the purposes of the *Sophistical Refutations*, it remains reasonable to say that, if a proposed refutation is an argument and is not a syllogism, then it is a paralogism. If, that is, the other thing, which is the contradictory of the thesis being defended, does not follow of necessity from the given things, then the refutation is sophistical. Or, in more modern garb: even if it is not necessary (and, so, not part of the definition of paralogism), invalidity is sufficient for fallacy.

As to the "seems" element in the traditional definition of fallacy, we have already heard Aristotle saying that sophistical refutations seem to be refutations but are, in reality, paralogisms, and terms germane to *phainomenon* (*SEl*., i, 164a20) recur insistently in the opening moves of the book (164a24, a26, b20 and b26). To get a grip on the respect in which this seeming is also a deceiving, we may briefly relay the four (or five) analogies that Aristotle offers for the relation between sophistical refutations and refutations in good order.

The first analogy (at i, 164a26–b21) is with the difference between beauty and the appearance of beauty. An argument that is in good order is genuinely beautiful, while a paralogism is a make-up effect. As cosmetics do not aim at the health of the subject, but only at the fleeting and deceptive pleasure of the beholder (a Platonic theme: *Grg*., 462c3–d10), so paralogisms do not aim at rational persuasion, but only at seducing the opponent. A paralogism is a painted meretrix.

The second (at i, 164b21–4) is between the glittery outward appearance of a metal, such as tin or iron pyrites, and the inner constitution and valuable properties, such as low chemical reactivity and malleability, of what it might be mistaken for, such as silver or gold respectively. An argument in good order is precious, but only the gullible would take a paralogism for a syllogism.

The third analogy (at i, 164b26–5a1) seems to involve both lack of expertise and inability to look more closely. This is what the rehearsals of the *Topics* are meant to cure. If one is not already alert to where tricks might be pulled, one might be unready to fend off the paralogisms that are deployed in discussion.

And the fourth (at i, 165a5–15) appears to offer two contrasts between mental arithmetic and the use of an abacus. The simpler would be that, if we try to do sums in our heads, we are more likely to overlook mistakes (our own or others') than if we set things out explicitly and keep tabs. The more complex depends on the fact that any given word has to stand for (*semainein*: 165a13) many (indeed infinite) things. Thus, as the position of a bead on an abacus changes its value according to its position, so also the words in an argument may change their meaning according to context. A sly use of paralogism is a trick in which the moves look obvious but are not and will deceive the unwary.

I have deliberately—perhaps even illicitly—presented Aristotle's analogies as evoking low-life traffic to bring out a sense in which, for him and for much of the later tradition, the use of ploys similar to the list of sophistical refutations in *SEl.*, iv–v is dirty play. In line with the broader definition of fallacy, it is a violation of the schemes or an infringement of the rules that pretends not to be.

Conversions

We have already noted that, in Aristotle's listing of thirteen sophistical refutations, the relations between the labels adopted and the examples furnished are not entirely unproblematic. Of *to en arche aiteisthai* not even one example is given and those that appear under *diaeresis* and *synthesis* are so heterogeneous as to have inspired poor William of Heytesbury to distinguish eight different phenomena here (Guglielmus, 1494); likewise with *para to pe* (*SEl.*, v, 166b37–7a20), which seems hardly more "in a given respect" (traditionally *secundum quid*) than "part for whole" (cf. the case of the Indian at 167a8–9). In the case of the eleventh label—*epomenon* (*SEl.*, v, 167b1–20)—we have five examples that all seem to conform to a single logical structure, even if they are further subdivided, for instance, by Peter of Spain in his re-casting of them in a syllogistic format (Petrus, c. 1230, VII, §§ 150–163).

The first example that Aristotle gives of *epomenon*, at 167b5–6, may be reconstructed as follows:

(Y) If this is honey, then it is yellow
This is yellow
(therefore) This is honey

Though Aristotle formulates the case in terms of neuter adjectives (likewise at vi, 168b30), our indexicals and anaphoric "it" play much the same logical role for the purposes of exposition.

The reason why Aristotle includes a piece of reasoning like (Y) in his listing of sophistical refutations is that the premises may be true, but the conclusion false As some recent commentators on the text say, lumping (Y) together with the other examples that we shall come to, (Y) is a "logical error" (Aristotle, 1995, p. 296) or a "paralogism" (Aristotle, 2007, p. 122). That is to say, (Y) is an invalid argument because, holding firm the premises, we might replace the conclusion with some other sentence that is true and that does not say that this is honey, indeed that says that this is not honey. The replacement that Aristotle hints at is: "this is bile". If this is bile, then it is yellow but is not honey, which may be a play on an opposition of honey as sweet and bile as bitter. Likewise, if this is a lemon, then it is yellow but is not honey, which also plays on an oppostion between the sweet and the sour. As well we all know, many things, including bile and lemons, other than honey are yellow.

In Scholastic parlance, doubtfully attributable to Aquinas (ps-Thomas, 1998, Chap. 3), the invalidity of (Y) is its *causa defectus*: what makes it a fallacy (cf. Ebbesen, 1987); but perhaps more interesting is what makes us fall into it, its *causa apparentiæ*. Aristotle offers a diagnosis of this in two phases.

The first phase is to say that there is the supposition (*to oiesthai*; 167b1) that the terms of the conditional in the first premise ("if this is honey, it is yellow") convert (*antistrephein*; 167b1–2) meaning that there is the supposition "if this is yellow, it is honey". It is worth noting how impersonal is this bit of supposing: it is not explicitly associated either with the attacker of the thesis (that this is honey) or with its defender. And, if it were pinned on anyone, it would in any case be rather forced. On the one hand, there are not many people who think of terms as converting or otherwise; such arcana of logical jargon are perfectly alien to the overwhelming majority, who do not suppose so because they have never thought of it or do not know the verb "to convert" in this sense. On the other, it is hard to think that anyone at all supposes that, if something is yellow, it is honey; to suppose so would be to suppose something that everybody knows is false. And it is better not to attribute to people supposings they know not of or that they flatly reject because they know something of bile or lemons and other yellow things.

Yet, Aristotle himself in *SEl.*, viii (169b30–7) seems to suggest that, in every sophistical refutation, there is some suppressed premise that the opponent of the thesis defended smuggles in to blindside the defender; and there are highly sophisticated modern accounts of how the thesis of the "false validating premiss" can make sense of how paralogisms can take in the unwary (Fait, 2012). So: what sort of supposing must be in play here?

The root verb, *oiomai*, from which Aristotle's gerundialisation derives can be naturally rendered in many contexts as thinking or believing or even, with respect to goods, hoping and, with respect to harms, fearing. But, in the context we are considering, it must be a fainter thing, closer to the concessive or intercalary uses that turn up when one wishes to admit to possible ignorance instead of certainty or to soften an assertion with an "it seems to me" (the "methinks" of yesteryear). Indeed, fainter still: the depersonalised supposition (*to oiesthai*) may, indeed, be a letting-it-pass or a not-noticing-that-not. That is, someone who uses an argument structurally similar to (Y) is not presenting himself as believing that terms that take the places of "honey" and "yellow" in (Y) convert, nor that only honey is yellow; rather, he is seeing whether he can get away with it.

Can he reasonably hope to get away with it? If the second phase of Aristotle's diagnosis of the *causa apparentiæ* of (Y) is correct, then the answer will be "often enough". As we shall see a little further on, the sorts of inference of which (Y) can be taken as an unnourishing example are our daily fare because they are about beliefs deriving from the senses (*peri ten doxan ek tes aistheseos*: 167b4). But from these, Aristotle says, arise tricks or deceptions (*apatai*; 167b4). Which is a pretty shocking thing for him to say. While we might expect sceptics to harp on about sticks looking bent in water, towers that look round and square, and so on, it is quite unexpected to find Aristotle being so harsh on beliefs deriving from the senses; but the gist of this passage seems to have to be that, when dealing with beliefs deriving from the senses, we are apt to fall into tricks or deceptions of which (Y) is an example.

There are perhaps two things to note here. One is that hardly anyone would persevere with an argument that depended essentially on (Y). It is, so to say, a counter-instance to the logical structure of which it is an example. We shall come in a moment to consider some ways of illustrating the structure in question with the use of variables. But (Y) is an instance that does have to do with beliefs deriving from the senses, and that shows that the structure might be a source of trickery and deception because the premises may be true but the conclusion false. The other thing to note is that it is hard to imagine that talk about honey and yellow will figure in the sort of debates that the cycle of lessons from which *SEl.* presumably derives is supposed to be preparing its students for. To this end, it may be that school exercises were set that took as their ostensible subject-matter also metaphysical arguments like that attributed to Melissus at 167b12–17; but these do not have anything to do with beliefs deriving from the senses. Indeed, we shall have no more to say about Melissus' argument precisely because such trickery and deception as it involves has to do with the very abstract notions of the ungenerated and the infinite.

Aristotle's overall diagnosis, and his reason for including *epomenon* in his list of sophistical refutations, then, is that it involves a supposition that it is hard to suppose anyone making in any full-blooded sense (whether about conversion or about yellow things), and that it arises from the tricks and deceptions of the senses. These two elements correspond in some degree to the Scholastic categories of the *causa defectus*, which is a matter of what is amiss with the argument, and of the *causa apparentiæ*, which is what makes us fall into the trap.

Mistakes and Confusions

In 2019, Robert Arp, Steven Barbone and Michael Bruce edited a book with the unequivocal title *Bad Arguments* (2019), in which they collected short essays on "100 of the Most Important Fallacies in Western Philosophy", of which the second is called "Affirming the Consequent". And its prominent position in so lengthy a listing gives us reason to think that instances of affirming the consequent are among the most important of the most important fallacies in Western Philosophy: not only bad arguments, but conspicuously bad arguments.

The author of the essay on affirming the consequent, Brett Gaul, begins by giving a rather abstract account of this "fallacious form of reasoning in formal logic" (Gaul, 2019a, p. 42). If, by this, he means no more than that arguments that exhibit a certain structure may have true premises and false conclusion, then we need not worry. Indeed, it is the sort of thing that most readers of this journal will have enountered in the first month or so of an elementary logic course. The trouble is that the editors' *Introduction* to *Bad Arguments* proposes to recognise only two basic types of reasoning, deductive and inductive (Arp et al., 2019, p. 13), and proceeds as if an argument that is proposed as deductive but that may have true premises and false conclusion is a fallacy and "a fallacy is a bad thing [and] should be avoided at all costs" (p. 19). This is less than convincing, but it may not represent the view of Gaul himself.

Gaul expounds his abstract account of affirming the consequent in terms of major and minor premises of a propositional syllogism and this expository choice puts anything that answers to the account pretty firmly in the class of arguments that are to be judged by the standards of deductive reasoning. As an anonymous commentator for this journal aptly expresses it, an affirmation of the consequent "shoots at" validity. And misses.

The major premise is described as both "general" and a "conditional", which "expresses a link between the antecedent […] and the consequent" (Gaul, 2019a, p. 42). Then, like Aristotle with his *antistrephein*, Gaul says that affirming the consequent is "the mistake of assuming that the converse of an 'if-then' statement is true" (2019a, p. 42). But, as we suggested of Aristotle, it is hard to think that, on most occasions that a consequent is affirmed, the affirmer is assuming anything of the sort, except in the very faintest way indicated earlier. We do not really have here a full-blooded assumption, if only (i) because the notion of a converse is at the disposal of those who have studied some formal logic (however little); and (ii) because anyone who was asked whether "if *p*, then *q*" is equivalent to "if *q*, then *p*" would very likely deny it, at least once what is meant by talking about *p* and *q* is explained. Here, though, we have the elements of *causa defectus* of affirming the consequent: it is defective because, taken as a deductive argument, it may have true premises and false conclusion.

A further suggestion that Gaul makes is that those who affirm the consequent sometimes do so because “it is mistaken for” (2019a, p. 42) the valid argument form *modus ponendo ponens*, and he then sets the two out side by side in skeletal format using propositional variables (p. 42):

	If p, then q		If p, then q
	p		q
(therefore)	q	(therefore)	p

While affirming the consequent is what Aristotle would call a paralogism and sophistical because it seems like or resembles a syllogism, Gaul thus suggests which syllogism it is that it resembles, which is a *causa apparentiæ*.

Descending from this level of abstractness, Gaul offers a comparison between a *modus ponendo ponens* with the two premises (1) “If Sophia is in the Twin Cities, then she is in Minnesota” and (2) “Sophia is in the Twin Cities” to arrive at the conclusion (3) “Sophia is in Minnesota”, and the affirming of the consequent with the same first/major/conditional premise (1) plus (4) “Sophia is in Minnesota” which fail to “guarantee” (Gaul, 2019a, p. 43) the conclusion (2) “Sophia is in the Twin Cities”.

The example is well-chosen for anyone who believes that the Twin Cities fall within but are not coextensive with Minnesota (though I gather that two counties of this conurbation are in fact in the territory of Wisconsin). One reason why the choice is good is that the spatial relations between being in the Twin Cities and being in Minnesota can be very intutively rendered with Euler/Venn set diagrams. The lack of the guarantee in the passage from (1) and (4) to (2) can be seen from there being areas of Minnesota that are not in the Twin Cities, such as Marshall in Lyon County, where Brett Gaul teaches. While being in the Twin Cities is a sufficient condition for Sophia’s being in Minnesota, her being in Minnesota is a merely necessary condition for her being in the Twin Cities, and Gaul suggests that affirming the consequent arises from a “confusion” of these (2019a, p. 44). Yet it may just be the case that Sophia, supposing her to be an obdurate city-dweller, would never go anywhere in Minnesota outside the Twin Cities, but this is not guaranteed by the conjunction of (1) and (4).

Gaul is by no means alone in characterising passages from premises to conclusion that lack deductive guarantee as “mistakes” and “confusions”, but I elected to look at his essay because it is recent, short and very clear indeed in the position it takes on the need to “avoid committing this fallacy” (2019a, p. 45), where the choice of the verb “to commit” would indicate a sin or a crime: not merely a bad argument, but a bad thing, to be “avoided at all costs” (Arp et al., 2019, p. 19, cited above). It may be salutary, therefore, to look a little more closely at some classic cases of this infamy and to seek some understanding of why we are so inclined to “commit” it.

Three Examples

After the example of honey and yellow, Aristotle offers four examples of *epomenon*, and we have already said we shall not further consider the case of Melissus because it does not fit the diagnosis of arising out of the tricks or deceptions of the senses. Which leaves three. Because I do not possess a fully worked-out algorithm for getting from the textual traces to formal presentation of arguments, it is with due hesitation that I offer the following suggestions.

At 167b6–8, we seem to have an argument that would, when more fully spelt out, look like this:

(W) The ground is wet
When it has rained, the ground is wet
(therefore) It has rained

The inversion of the order as between what Gaul calls the major and minor premises is of no logical significance, although it seems to possess a greater naturalness. In (Y) and Gaul's Sophia examples, the dominant functor in the major premise is "if", but in (W) it is "when". This seems a close enough relative of "if" (think German *wenn*) to have many of the same logical powers, and, indeed, might be appealed to to supply the link that Gaul refers to between antecedent and consequent in a conditional—in (W), a temporal and causal sequence. But, as Aristotle says, it is "not necessary" (*ouk anankaion*; line 8), which will remind us of "not a syllogism".

The next example (167b10–11) is introduced by Aristotle's saying that something of the sort might be used in the rhetorical elaboration of an accusation, presumably for immorality. Unlike the honey and yellow example, this brings us closer to the wider world outside the Academy. Suppose the accused is Coriscus; thus, the argument would run:

(A) Coriscus is abroad at night smartly turned out
Those who are having an affair are abroad at night smartly turned out
(therefore) Coriscus is having an affair

Again, the minor premise is placed first. The major is not explicitly quantified and the conditional functor is likewise rather implicit in the generalisation, which could be spelt out rather too explicitly for plausibility as "If someone is having an affair, he/she is abroad at night smartly turned out".

After his exertions, Coriscus may return to fill out Aristotle's last example of *epomenon* (167b18–20):

(F) Coriscus is running a temperature
A fever makes you run a temperature
(therefore) Coriscus has a fever

Naturally, the major could be expressed as an explicit conditional: “if you have a fever, you run a temperature”, and even impersonally with “one” or “anyone” and an anaphoric “he/she”; as it is, the “makes” indicates the link between antecedent and consequent.

One thing to observe about (W), (A) and (F) is that they all exhibit the same logical structure as (Y), Gaul’s skeletal uses of variables and his Sophia example. If being a fallacy were purely a question of logical form, then we might expect that any argument that shared a logical form with a fallacy would be a fallacy. There are general reasons for doubting this (Davies, 2012), but, on one plausible assessment, (W), (A) and (F) are all arguments that we might find ourselves proposing.

Another thing worth observing is that, from consideration of (W), (A) and (F) set out in full, we can see pretty much straight off that Aristotle’s list of thirteen sophistical refutations is incomplete. For he nowhere takes account of the structure specular to *epomenon*, which a later tradition labels *negatio antecedentis* and Gaul’s following essay in *Bad Arguments* calls “denying the antecedent” (Gaul, 2019b); Gaul gives both the skeletal structure with propositional variables “If *p*, then *q*, but not-*p*; therefore not-*q*” and fits a Sophia-and-Twin-Cities case into this logical form. Without over-regimenting (suppressing the major in each case), arguments specular to (W), (A) and (F) may be set out in a quasi-dialogic form:

(W*) It hasn’t rained? The ground will be dry and I needn’t put my boots on.

(A*) Coriscus isn’t having an affair? He’ll be at home this evening, so I’ll pay him a call.

(F*) Coriscus doesn’t have a fever? He won’t be running a temperature.

If Aristotle had noticed something similar to them, then *epomenon*’s mirror image might have lengthened his list to fourteen, still nowhere near Arp & Co.’s 100, Calemi and Paolini Paoletti’s “exactly 150” (2014, p. 7) or the vulgarity of Bennett’s “Over 300” (2015), to notice but a few of the recent counts. It would be another story to explain why such counts are, in the nature of the case, spurious.

An Abductive Defence

The apparently provocative “defence of a fallacy” promised in my title amounts to little more than the observations: (i) that (W), (A) and (F) are examples of the sorts of inferences we make much of the time; and (ii) without such inferences, we would be quite at a loss to go about our everyday business. But I permit myself to flesh them out a little.

Let us look again at (W) and (W*). One or other is among the first inferences I make pretty much every morning, even before drinking a coffee. If I look out of the window on the street and see that there is water on the ground, and I look out of the window on the courtyard and see that there is water on the ground, I infer that it has rained: (W). For sure, the street may have been cleaned by the public services and the courtyard may have been sprayed by my neighbour washing his car. Thus, my premises (what I see out of two of my windows) would be true but my conclusion (that it has rained) could be false. But nothing induces me to suppose such wayward coincidences, which do not, in any case, exclude its having rained. Viceversa, if the street and the courtyard are both dry, I infer that it has not rained and choose my shoes accordingly: (W*). Not to make such inferences would not be loyalty to formal logic, but early-morning doziness.

As the day goes on, I continue to make inferences that conform in one way or another to (W), (A) and (F) in affirming the consequent or to (W*), (A*) and (F*) in negating the antecedent. When I arrive at the bus stop, if I see that there is no-one waiting, I infer that the bus has just passed because, when the bus has just passed, there will be no-one waiting: those who had been waiting have got on and are no longer waiting. In more challenging environments, such as the workplace, the inferences I make become ever more adventurous, interesting and risky: more likely for the premises to be true though the conclusion may turn out to be false (especially when they involve attributing specific mental states to my colleagues). But still I make them, and I take it that this is not a merely autobiographical confession. Rather, it reflects what we all do most of the time.

I am very slow to allow that we are all making mistakes and confusions all day long or that we are supposing that terms that plainly do not convert convert. If I did allow such a thing, then we would all have to attribute to everybody massive amounts of logical ineptitude in making such mistakes and confusions. Yet, there is strong evidence to show that, even when we are challenged in circumstances that put us on our mettle, the drive to affirm the consequent is strong and constant.

Traces of this evidence can be found in the robustness of the results of a celebrated test first made explicit by Peter Wason (1968) and variously reproduced (Evans, 1982; Manktelow, 1999). The so-called selection task induces about 90% of the general population (and about 75% of mathematics majors) to affirm the consequent when asked to verify a conditional in the rather artificial conditions of a psychology experiment, though this figure is significantly lower in "thicker" social settings (e.g., Cosmides, Tooby, 1992). There has been considerable debate about just what is going on here (Motterlini, 2008, pp. 20–28, 263–265), but one fixed point seems to be that the Wason effect is an obstacle to people's arriving at the "correct" or "optimal" (Oaksford & Chater, 1994; Zenker, 2017, pp. 449–452) solution to the task proposed.

It is indeed true that when the task is one of verifying a conditional such as "If there is an 'A' showing on one side of a card, then there is a '2' on the other", turning over a card with a “2” showing is perfectly irrelevant. But, of a morning, what I am doing is not seeking to verify the quasi-conditional “When it has rained, the ground is wet”. At most, I am reassuring myself that nothing too untoward has been going on in the night. That is, the water on the ground both in the street and in the courtyard is explained most simply by its having rained. And the simplicity here can be put numerically. If the water in the street could have been due to the public cleaning service and that in the courtyard to my zealous neighbour—two causes for two effects—the rain’s causing both dousings can be regarded as one cause for perhaps just one effect, and gives rise to no need to puzzle over a temporal coincidence. Rain removes any cause for surprise at widespread water: if there has been rain in the night, water on the ground is a matter of course.

For someone living in Manchester, water on the ground is not at all surprising; but for someone living in the Atacama desert, it is. In the one, an inference like (W) is a matter of course and alternates with (W*) in a ratio of about 7:12 over the course of the average year; in the other, there may be years when (W) doesn’t occur even once to the inhabitants, who are condemned to repeating (W*) and never needing to put on waterproof shoes. If we are loyal to formal logic, we may say that Mancunians switch between affirming the consequent and denying the antecendent, while a certain number of Chileans do nothing but deny the antecedent.

Inferences (A) and (F) and their counterpart denials of the antecedent, call for slightly different criteria of evaluation. Just how good (A) and (A*) are as inferences depends on what Coriscus is like. Though we know of a historical Coriscus, a friend of Aristotle’s in the time he spent at Skepsis, his name appears as a dummy for this or that man more than 60 times in the *Corpus*. In our ignorance of what he is like, we may entertain two hypotheses about him (as we could about Gaul’s Sophia and her potential refusal to visit Minnesota outside the Twin Cities). If Coriscus is generally a stop-at-home sloven, then his being out and about and well turned out is a change in his behaviour that is rather surprising and that calls for some explanation. If he has recently been heard talking excitedly about someone in particular, then we have a clue in favour of (A). But if he has long been a party-goer with a sharp dress sense, then it is hard to be sure whether he is taken up with anyone in particular just at the moment, so that (A) is rather under-motivated. Conversely, (A*) is boosted if he has just broken up and is that bit depressed so that maybe an evening visit will cheer him up.

By contrast, the diagnosis in (F) derives from medical facts. If Coriscus has a temperature of 38.5°, then the conclusion that he has a fever will account for this symptom. In certain circumstances, such as those of the time of writing, the minor premise might lead us to suppose that he has been infected with Covid-19 and, at least in line with a principle of precaution, to treat him as such and quarantine him; but, at other times, a more generic and less alarming conclusion might be all that we feel entitled to.

In (W*) and (A*), the practical consequence drawn in each case, regarding footwear and an evening visit respectively, comes without undue strain, and while (F) might induce us to prescribe at least an antipyretic, in (F*), the absence of fever doesn't indicate any particular course of action. This is because health, being a normal state, does not call for treatment, while disease does. The asymmetry here is a material matter rather than a formal one, and if we allow ourselves to be guided only by formal logic, we might miss it. For formal logic has nothing much to say about what is or is not out of the normal.

Not only do the notions of being normal or a matter of course and being surprising or alarming resist formalisation, the relations among them are problematic. As Peirce says, they are "very little hampered by logical rules" (*CP*, 5.188). Nevertheless, the following scheme of inference looks as if it captures what makes inferences like (W), (A) and (F) attractive, morning, noon and night:

(P) The surprising fact, C, is observed;
But if A were true, C would be a matter of course,
Hence, there is reason to suspect that A is true (*CP*, 5.189)

If we apply Gaul's abstract description of affirming the consequent, we have the major (second) premise, which is general or a conditional, of which the consequent is affirmed in the minor (first) premise (as in (W), (A) and (F)), and the conclusion is the antecedent of the major.

Thus, at least formally, (P) is a case of mistake or confusion. But there are at least three points to be considered to make clearer how much or how little our "defence of a fallacy" amounts to. One is how to regard the presence of "surprising" and "a matter of course" in (P). A second is whether the use of the subjunctive in the major premise of (P) undermines the assimilation we are suggesting. And a third is what we are to make of the relation between the premises and the conclusion, given that there is no clear sense in which the conclusion "follows of necessity" (*ex anankes symbainei*, *Top.*, I, i, 100a26, and *AnPr.*, I, i, 24b20, cited above) from them.

If we can illustrate, albeit in a preliminary way, that plausible responses to these points do not go against what we are suggesting, then inferences like (W), (A) and (F) and (W*), (A*) and (F*) may not be in such a bad condition as those who call them mistakes or confusions would have us believe, because they are what Peirce calls abductions. Though the assessment of individual abductions is not formalisable, the thesis that none are in good logical shape is so paradoxical as to be a betrayal of mere ignorance on the part of anyone who suggests it. To insist that only deductively valid inferences are in good logical shape (van Eemeren's Validity rule) is to fall into the trap of Maslow's hammer: because we have accounts of some good deductions, such as those that are formed with "if… then—" sentences, it is easy to suppose that every inference in good logical shape will conform to that pattern (and so will be a nail for the hammer we happen to have). This easy supposition is itself an affirmation of the consequent, and should shame those who think that this sort of inference is to be "avoided at all costs".

As to the first point, we have already adverted to some differences between the surprise of water on the ground in the Atacama, the spectacle of idle Coriscus suddenly out and about, and the alarm caused by a temperature of 38.5° as against the usual damp in Manchester, snappy Coriscus strutting his stuff and there being no need even to take out the thermometer if there is no perceived deviation from normal temperature. But these differences do not make a difference to whether the inferences are attractive or not, nor to whether particular actions are called for. The presence or absence of water on the ground is something observed: its presence is not a surprise in Manchester but very much so in the Atacama; but the supposition of rain makes this a matter of course in both places and makes sense of this or that choice of footwear. And likewise in the other cases. If what we have reason to suspect explains or makes sense of what we have observed, then the inference is, as Peirce says, "the only logical operation that introduces any new idea" (*CP*, 5.171–2).

Second, the use of the subjunctive in Peirce's formulation does not seem to be essential to understanding what an abduction is, and (P) could as well be re-written as:

(P*) The surprising fact, C, is observed;
But if A is true, C is a matter of course,
Hence, there is reason to suspect that A is true

In Gaul's terminology, the major (second) premise is, in both (P) and (P*), a conditional and general. One might even say that, with the subjunctive formulation, the invocation of the link that Gaul refers to is stronger, but this is to stray into the tormented field of the analysis of conditionals. Even if the second premise of (P*) does not express any link that is lawlike in any strong sense, the idea is that terms that take the place of "A" should be in one way or another explanatory of those that take the place of "C", as we seem to have in (A), where the idea of a lawlike generalisation would surely be out of place. Episodic personal behaviours do not lend themselves to lawlikeness. Coriscus' having an affair makes his dapperness understandable because he is trying to make a good impression on the person he is courting. Perhaps this is a second level of explanation, but surprises sometimes need to be multiply contextualised. Nevertheless, each of (W), (A) and (F) could equally well be re-written with a subjunctive in the major premise: "if Corisucus were having an affair, then his being out and about at night would be a matter of course", and so on.

And third, as to the question of "following", if this depends on logical form or structure, conformity to (P) or (P*) will not, as Gaul puts it, "guarantee" the conclusion. After all, the point of departure with (Y) was that this could be yellow, and yet not honey but bile or a lemon. In this respect, having a false conclusion is a reason for deprecating a given abduction; but having a conclusion that might be false even though the premises are true is neither here nor there in assessing such an inference. In the cases of (W), (A) and (F), and despite Aristotle's deployment of them to illustrate *epomenon*, the commentators' collusion in thinking they are logical errors or paralogisms and Gaul's allegation of mistakes and confusions, we have inferences in which, indeed, the conclusion does not follow (*symbainein*) from the premises, but absent which we would be in the dark about the water on the ground, Coriscus' nocturnal behaviour and the temperature he is running.

Peirce himself characterises the status of the conclusions of abductions in terms of the cyclical nature of investgations, where the upshot of an abduction provides a "hypothesis" (*CP*, 2.619–44, 5.599–600, 6.466–70) or a "conjecture" (*CP*, 2.755, 5.189, 6.469, 8.209) that calls for further testing. For this reason, he sometimes says that it is "in the interrogative mood" (*CP*, 2.758, 6.469) or even a "guess" (*CP*, 2.121, 2.753, 6.491, 7.219).

This is not the place to go further into the roles that abductions play both in day-to-day reasoning and in the more formal business of testing conjectures by seeking refutations of them, but we have already seen some surprising facts to do with how often and how stoutly everybody commits affirming the consequent and denying the antecedent. If the ancients and moderns who tell us that arguments that can have true premises and false conclusion are bad things to be avoided at all costs were telling the whole story, then we would have to attribute massive logical ineptitude to everybody. If, in some cases, such inferences are more or less decent abductions, the facts in the case would be a matter of course and we would not have to attribute massive logical ineptitude to everybody. Hence, there is reason to suspect that they are more or less decent abductions. This is the guess, hypothesis or conjecture that my defence of affirming the consequent invites the reader to interrogate. And I arrive at it by affirming the consequent or, as I prefer to say, by an abductive inference.

REFERENCES

Aristotle. (1989). *Prior Analytics* (tr. R. Smith). Indianapolis: Hackett.

Aristotle. (1995). *Le confutazioni sofistiche* (text with facing Italian translation and commentary, ed. M. Zanatta). Milan: Rizzoli.

Aristotle. (2007). *Le confutazioni sofistiche* (text with facing Italian translation and commentary, ed. P. Fait). Rome-Bari: Laterza.

Aristotle. (2016). *Organon* (text with facing Italian translation, ed. M. Migliori). Milan: Bompiani.

Arp, R., Barbone, S. Bruce, M. (2019). *Bad Arguments*. Hoboken, NJ, Chichester: Wiley Blackwell.

Bennett, B. (2015). *Logically Fallacious: The Ultimate Collection of Over 300 Logical Fallacies*. Sudbury, MA: Archieboy.

Calemi, F., Paolini Paoletti, M. (2014). *Cattive argomentazioni: come riconoscerle*. Rome: Carocci.

Cosmides, L., Tooby, J. (1992). Cognitive Adaptations to Social Exchange. In J. Barkow et al. (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163–228). New York: Oxford University Press.

Davies, R. (2012) Can We Have a Theory of Fallacy? *Rivista Italiana di Filosofia del Linguaggio*, *6*(3), 25–40.

Ebbesen, S. (1987). The Way Fallacies Were Treated in Scholastic Logic. *Cahiers de l'Institut du Moyen Âge Grec et Latin*, *55*, 107–134.

Evans, J. (1982). *The Psychology of Deductive Reasoning*. London: Routledge.

Fait, P. (2012). The "False Validating Premiss" in Aristotle's Doctrine of Fallacies. An Interpretation of *Sophistical Refutations* 8. *History of Philosophy and Logical Analysis*, *15*(1), 238–266.

Gaul, B. (2019a). Affirming the Consequent. In R. Arp et al. (Eds), *Bad Arguments* (pp. 42–45). Hoboken, NJ, Chichester: Wiley Blackwell.

Gaul, B. (2019b). Denying the Antecedent. In R. Arp et al. (Eds.), *Bad Arguments* (pp. 46–47). Hoboken, NJ, Chichester: Wiley Blackwell.

Guglielmus Hentisberi. (1494). *De sensu composito et diviso*. Venice: Bonetus Locatellus.

Hamblin, C. (1981). *Fallacies*. London: Methuen.

Hansen, H. V. (2002). The Straw Thing of Fallacy Theory: The Standard Definition of "Fallacy". *Argumentation*, *16*, 133–155.

Iacona, A., Marconi, D. (2005). *Petitio principii*: What's Wrong? *Facta Philosophica*, *7*, 19–34.

Manktelow, K. (1999). *Reasoning and Thinking*. Hove: The Psychology Press.

Motterlini, M. (2008). *Trappole mentali*. Milan: Rizzoli.

Oaksford, M., Chater, N. (1994). A Rational Analysis of the Selection Task as Optimal Data Selection, *Psychological Review*, *101*, *4*, 608–631.

Peirce, C. S. (*CP*). *Collected Papers* (8 vols, Eds. C. Hartshorne, P. Weiss, A. W. Burks). Cambridge, Mass: Harvard University Press.

Petrus Hispanus. (c.1230). *Summule logicales* (ed. L. M. De Rijk). Assen: Van Gorcum.

Rapp, C., Wagner, T. (2013). On Some Aristotelian Sources of Modern Argumentation Theory, *Argumentation*, *27*, 7–30.

Ryle, G. (1965a). The Academy and Dialectic [originally: Dialectic in the Academy]. In G. Ryle, *Collected Papers* (vol. I, pp. 89–115). Bristol: Thoemmes.

Ryle, G. (1965b). Dialectic in the Academy [originally: Aristotelian Dialectic). In G. Ryle, *Collected Papers* (vol. I, pp. 116–125). Bristol: Thoemmes.

Schopenhauer, A. (1830). *Dialektik*. In A. Hübscher (Ed.), *Der handschriftliche Nachlaß* (vol. III, pp. 666–695). Munich: Deutscher Taschenbuch Verlag.

Striker, G. (1991). Perfection and Reduction in Aristotle's *Prior Analytics*. In M. Frede, G. Striker (Eds.), *Rationality in Greek Thought* (pp. 203–219). Oxford: Oxford University Press.

pseudo-Thomas de Aquino. (1998). De Fallaciis ad Quodam Nobiles Artistas. Retrieved from: https://www.documentacatholicaomnia.eu/03d/1225-1274,_Thomas_Aquinas,_De_Fallaciis._(Dubiae_Authenticitatis),_LT.pdf van Eemeren, F. H. (2001). Fallacies. In F. H. van Eemeren (Ed.), *Crucial Concepts in Argumentation Theory* (pp. 135–164). Amsterdam: Amsterdam University Press.

van Eemeren, F. H., Grootendoorst, R., Snoeck Henekmans, A. F. (2002). *Argumentation: Analysis, Evaluation, Presentation*. Mahwah, NJ: Lawrence Erlbaum.

van Eemeren, F. H., Grootendoorst, R. (2003). *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge: Cambridge University Press.

Walton, D. N., Reed, C., Macagno, F. (2008). *Argumentation Schemes*. Cambridge: Cambridge University Press.

Wason, P. (1968). Reasoning About a Rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.

Zenker, F. (2018). Logic, Reasoning and Argumentation: Insights from the Wild. *Logic and Logical Philosophy*, *27*, 421–451.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 43–64
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.04

Article

CRISTINA CORREDOR [*]

SPEAKING, INFERRING, ARGUING. ON THE ARGUMENTATIVE CHARACTER OF SPEECH[1]

SUMMARY: Within the Gricean framework in pragmatics, communication is understood as an inferential activity. Other approaches to the study of linguistic communication have contended that language is argumentative in some essential sense. My aim is to study the question of whether and how the practices of inferring and arguing can be taken to contribute to meaning in linguistic communication. I shall suggest a two-fold hypothesis. First, what makes of communication an inferential activity is given with its calculability, i.e. with the possibility to rationally recover the assigned meaning by means of an explicit inference. Secondly, the normative positions that we recognize and assign each other with our speech acts comprise obligations and rights of a dialectical character; but this fact does not entail nor presuppose an argumentative nature in language or speech. Both inferring and arguing are needed, however, in the activity of justifying and assessing our speech acts.

KEYWORDS: arguing, inferring, argumentative value, inferential meaning, illocutionary, normativity of speech, Austin, Grice.

Introduction

Some philosophical and linguistic approaches to the study of the pragmatics of language, following Grice (1989), have defended the idea that linguistic communication is an inferential activity. The inferential nature of speech is con-

[*] University of Valladolid, Department of Philosophy. E-mail: corredor@fyl.uva.es. ORCID: 0000-0002-1317-1728.

[1] This work has been supported by the Spanish Ministry of Science and Innovation, research projects PGC2018-095941-B and PID2019-10748GB-100.

nected to a notion of meaning *qua* speaker's meaning, where the speaker's communicative intentions have to be inferred by the hearer. Notwithstanding the reference to psychological attitudes in his definition of speaker's meaning, Grice's view was primarily semantic and philosophical. He aimed at clarifying notions such as those of sentence meaning and word meaning. In the domain of pragmatics, the derivation of implicatures was not intended by him as a psychologically real process, but as a rational reconstruction of how this implicaturated meaning might be obtained. Some recent neo-Gricean theories (prominently, relevance theory and contextualism) orientate their approach in a different direction by aiming at a psychologically real, empirically testable theory.

Other approaches to the study of linguistic communication have contended that language is argumentative in some essential sense. Inferentialism, in the form given to it by Brandom (1994; 2000), presupposes in the speakers a pretheoretical capacity to participate in the practice of giving and asking for reasons. Semantic meaning results from the contribution that expressions make to the inferential relations of the sentences in which they occur. From a different, linguistic approach, the theory of argumentation within language (Anscrombe, Ducrot, 1976; 1988) contends that the semantic meaning of words determines the dynamics of discourse, and this in a form that is argumentatively orientated. This theory aims to show how a fact can be differently understood and communicated depending on the linguistic formulation chosen, and this election is taken to determine which other linguistic and argumentative moves are available.

My interest lies in the question of how inferring and arguing can be taken to contribute to meaning in communication. In particular, I hope to clarify how meaning can be said to depend on, or to be essentially related to argumentation. At this point, the formulation of the question must remain broad, since it is intended to comprehend different theories that endorse dissimilar views of this contribution and do so by focusing on different dimensions of meaning and communication. Knowingly, recent views in neo-Gricean pragmatics have developed a view of communication that understands it as an inferential activity, where the hearer's inferential work plays an indispensable role in grasping the speaker's communicative intentions and thus in capturing what can be conceptualized as pragmatic, communicated meaning. Also, the theory of argumentation within language has defended a view according to which semantic meaning in use cannot be dissociated from its argumentative value. And Brandom's normative pragmatics contends that the practice of giving reasons (and evaluating those reasons) is constitutive of meaning at the semantic level. Therefore, the contribution of inference and argumentation to meaning has been taken to impinge on both semantic and pragmatic levels. My aim is to consider in turn both theoretical possibilities by means of studying the influential theories mentioned above, namely, Grice's account of communicated meaning, Brandom's normative pragmatics and Anscrombe and Ducrot's notion of radical argumentativity.

Although there have been other theories dealing with this issue,[2] here I shall focus my attention only on the above mentioned ones. I take them to be highly representative of the topic at hand and I expect that discussing their main ideas will help me to give plausibility to my own views. In what follows, my aim is to give support to the following two hypotheses. Firstly, the idea that linguistic communication puts in place the interlocutors' inferential capabilities (together with other competences) is uncommitted and seems to me to be correct. But this fact should not be taken to give support to a stronger thesis, which would make of meaning an intentional entity and would explain linguistic communication solely in psychological terms. Following Grice (1975), I contend that what makes of communication an inferential activity is given by its calculability, i.e. by the possibility to recover an utterance's meaning by means of a rational reconstruction. This normative requirement, already present in Grice's views, is what I have tried to capture by means of a first hypothesis. Secondly, in my view, the way in which some expressions seem to codify certain inferential relations should not be seen as the product of an argumentative nature inherent in language. Argumentation is a special form of communication and interaction, where an arguer gives support to a claim by adducing reasons. This is not pre-codified in language, but an activity performed by giving reasons and assessing those reasons.

In what follows, my aim is to examine the above mentioned relevant theories, focusing on the way in which they have related inference and communication (or communicated meaning), on the one hand, and on the other, argumentation and semantic meaning. I hope this will allow me to clarify the concepts involved and give support to my views.

1. Inferring and Arguing

In order to approach the issue of the relation between meaning, inference and argumentation it is advisable to begin by considering the conceptual distinction between inferring and arguing. In a pre-theorical, intuitive approach, inferring is making the step from a belief to another (in thought or speech). We can be said to infer when we come to believe something on the basis of another previously entertained thought. Nevertheless, this tentative approach is lacking. It makes room for cases in which no reasoning links the first and last beliefs, and it does not distinguish personal, consciously endorsed inferences from other processes in which some belief causes, in a fortuitous or merely associative way, another belief. The idea that there must be a chain of reasoning articulating the step from a belief to another allows for this distinction, but it introduces another concept in

[2] Notably, Habermas's theory of communicative action subsumed a formal pragmatics in which understanding a speech act amounted to knowing the reasons that might justify it, should this justification be required by other interactants. This interesting theory cannot be addressed here (Habermas, 1981).

need of clarification, namely, that of reasoning. In its turn, reasoning may be said to be drawing inferences, which would be obviously circular.

A possible way out of this conceptual difficulty is offered by Frege's views. He writes, "To make a judgment because we are cognizant of other truths as providing a justification for it is known as inferring" (Frege, 1979, p. 4).[3] The burden of this definition lies on the high-level notion of justification on which it relies. In Frege's theoretical framework, however, we may safely assume that he is implicitly considering the availability and application of formal rules of a deductive kind. Notwithstanding this, his normative requirement in order for a transition from one judgement to another to qualify as inferring is that truth be transferred. Although Frege's definition seems in principle only related to theoretical reasoning (due to its presupposed connection between justification and truth), a similar view with a broader scope is due to Grice. In a preliminary approach, he says, "reasoning consists in the entertainment (and often acceptance) in thought or in speech of a set of initial ideas (propositions), together with a sequence of ideas each of which is derivable by an acceptable principle of inference from its predecessors in the set" (Grice, 2001, p. 5). Notwithstanding Grice's appeal to rules of inference, his notion of reasoning is broader than Frege's in that it is not limited to deductive rules and comprises practical reasoning as well. For Grice, inferential rules should be seen as transitions of acceptance which guarantee the transmission of some value of satisfactoriness, truth being but a particular case.

The appeal to rules of inference may seem unduly restrictive, if also our informal, ordinary reasoning has to be accounted for. Grice himself suggests that inferential rules can be seen as directives and their observance as a desideratum, but he carefully avoids conjecturing about their nature. It may be useful at this point to take into account the distinction put forward by some recent theories between two distinct processing modes or types of reasoning. The first one is characterized as automatic, fast, and non-conscious; it also is described as associative, heuristic or intuitive. The second one is controlled, conscious and slow; it is also taken to be rule-based, analytic or reflexive (Kahneman, 2011; Frankish, 2010). Intuitively, it seems that only the second mode of reasoning could be related to both Frege's and Grice's notion of inference and their appeal (tacit or explicit) to rules. Yet this conclusion would be hasty, in view of what I take to be an essential aspect in both of their views. It concerns the normative role played by inference rules, in that they should guarantee the transmission of some value among judgements. From this perspective, what is at stake in both Frege's and Grice's accounts does not need to be a psychologically real process. Instead, inferring is seen from a normative viewpoint, as a process in which a transition is effected from a judgement to another (in thought or language), and such that this

[3] Quoted by Boghossian (2014, p. 4). Boghossian's own proposal is to understand inferring as a matter of following a rule of inference in one's thought. Although this is an interesting view, it cannot be discussed here.

process can be assessed according to a normative requirement of transmission of correctness (truth for Frege, in theoretical reasoning; a value of satisfactoriness for Grice, in both theoretical and practical reasoning). From now on, this is the notion of inferring that I shall be considering here.

It can thus be said that we infer when we make the step from one judgement to another, in thought or language, in such a way that the transition we perform is in principle subject to assessment as to its preserving correctness (truth or otherwise practical correctness). Inferences are the product of acts of inferring, and reasoning is drawing inferences. From a logical point of view, inferences in reasoning can be represented by means of entailments between propositions (propositions being representational units with complete truth conditions, hence a theoretical object), and those entailments can in turn be reconstructed as carried out by virtue of certain rules. On the pragmatic level, however, real acts of inferring are not necessarily guided by formal rules. Our ordinary reasoning can and seems largely to be carried out through material inferences, which rely on conceptual, non-formal relations. Acknowledging this fact does not amount to pointing out the difference between abstract inferences, conducted within a formal system, and psychologically real ones. Rather, it endeavours to highlight the relation between actual inferences, seen as the product of the activity of inferring, and their susceptibility to assessment according to independent criteria of correctness.

Even if the above approach to acts of inferring is broad and remains intuitive, it should help us to realize the difference between inferring and arguing. Here, I am going to consider argumentation as a communicative activity that fulfils an essentially epistemic function. Argumentation consists in adducing reasons in order to justify a claim and in assessing those reasons. My approach is not intended as a formal definition, but as a very general and intuitive characterization. Moreover, here I endorse the widely held view according to which argumentation articulates three dimensions, namely, logical, dialectical, and rhetorical, respectively related to its product, its procedure, and its process. Although this is not a universal view,[4] the distinction stands as a useful one in characterizing theoretical proposals.

My own approach is pragmatic in that I am considering argumentation as a type of communicative and interactional activity. I take it that adducing reasons can be seen as a speech act of the assertive family,[5] internally related to the speech act of concluding a claim (another speech act of the same family). What makes of these acts an act of arguing is the internal connection between them. From a logical point of view, this internal relationship is what Toulmin (2003)

[4] Wenzel (1992) is usually credited with having put forward the idea that there are three perspectives in the study of argumentation, namely, logical, dialectical, and rhetorical. More recently, there is a wide consensus among scholars that these perspectives should be better seen as three dimensions of a single practice or form of activity. For an alternative view, centred on arguments, see, e.g., Blair's (2012).

[5] The idea that a group of different types of speech act belong to the assertive family is due to Green (2018).

termed *warrant*. In his view, warrants are inference-licenses or canons of argument, able to be made explicit in the form of hypothetical statements, and such that "authorise the sort of step to which our particular argument commits us" (Toulmin, 2003, p. 91). Warrants can be made explicit but will usually remain tacit or implicit. Although making a warrant explicit usually entails the adoption of an analytic stance on acts of arguing, this element (*qua* inference-license) must be seen as an essential component in them. When it is lacking, the resulting speech acts would be two assertions not argumentatively related to each other. Whenever data *D* is adduced as reasons in support of claim *C*, we can take it that the internal relation between *D* and *C* is to be captured, in an analytic form, by means of a warrant. Toulmin sets forward a tentative, general formulation in the following terms, "Data such as *D* entitle one to draw conclusions, or make claims, such as *C*", or alternatively, "Given data *D*, one may take it that *C*". Even if making this form explicit (whenever it is left tacit or implicit) presupposes the adoption of an analytic point of view, no piece of speech or discourse can be seen as argumentative unless this component relationship is part of the performed act.

It is worth stressing that the notion of warrant, as introduced by Toulmin, belongs to the logical dimension of argumentation. To that extent, it can be seen as an abstract, theoretical notion that tries to capture what should have a pragmatic realization. There have been different suggestions that address this issue. To mention but a few that are, perhaps, closer to my outlined position, warrants have been understood as general practical statements (Hitchcock, 1985), as correlated to implicit assertions (Bermejo-Luque, 2011), and as Gricean conversational implicatures (Labinaz & Sbisà, 2018). In my view, a pragmatic account in speech-act theoretical terms should try to identify the speech act or acts, if any, whose role can be captured on the logical (semantic) level by means of Toulmin's notion of warrant. Up to this point, I am not in a position to give a complete and satisfactory account. My intuition is that these "warranting" acts are not full-fledged speech acts, and that whenever made explicit, they acquire also the character of verdictive speech acts.

Moreover, taking into account not only the act of adducing reasons, but also that of assessing those reasons makes of my approach a dialectical one.[6] This assessment can be carried out by means of questioning and criticizing the adduced reasons, by questioning the support that the adduced reasons lend to the claim (the relationship captured by means of the notion of warrant), and also by

[6] The dialectical perspective on argumentation sees it as a special form of communicative interaction, where certain regulated procedures guide and allow the participants to produce and assess their acts of arguing (cf. Wenzel, 1992; also, Eemeren and Grootendorst, 2004). My focus here is on argumentation as a communicative practice that consists of putting forward acts of arguing; and I take it that these, in their turn, answer to certain felicity conditions. To the extent that these felicity conditions can be understood to be regulating the practice, they allow the participants to adopt a normative stance and assess other participants' (and their own) acts and arguments.

It is in this sense that I consider my approach to be pre-eminently dialectical.

presenting conditions of rebuttal.[7] By virtue of this process of adducing reasons, and of criticizing those reasons (in themselves, and in their internal relationship with the corresponding claim), acts of arguing have epistemic value. Notwithstanding the different goals that argumentation can fulfil,[8] it allows us to give support and thus to justify our claims. This is, in my view, what can be taken to be essential in argumentation.[9] When we argue, we put forward the reasons that, according to us, give support to the claim we purport to be true or otherwise correct. This support amounts to argumentative justification and provides an epistemic basis for the rational acceptance of the claim at issue.

Now, it should be easier to see why inferring and arguing are not one and the same concept. From a logical point of view, as pointed out before, the steps we perform in reasoning can be represented by means of implications between propositions, of a form that is evaluable as to their preserving correctness. Arguing is an epistemic activity, conducted communicatively, in which we adduce reasons in order to give support to and thus justify a claim. Following Toulmin, it can be said that in acts of arguing the transition from reasons to claim becomes legitimate by virtue of an inference-licence that authorizes it. While in reasoning the inferential steps we make (in thought or speech) do not need to invoke such legitimated or authorized character, the fact that arguing is a communicative activity makes of this requirement an essential component of a correct performance. In arguing, we interact with others and their assessment or appraisal has an effect on our own performance. The activity of arguing cannot be detached from the activity of adducing reasons to justify a claim, which entails a commitment by the arguer (possibly tacit) to the inference-licence that authorizes the step. Acts of arguing cannot be understood unless an interpersonal or social context is given, where the adduced reasons and their relationship to the raised claim can be assessed.

Still, it is possible to doubt whether both inferring and arguing are on a par in that both can be assessed as to correctness or incorrectness.[10] The difference here, to my mind, lies in the fact that an act of arguing is an action, is an act performed according to certain conventional felicity conditions, in the same sense in which any speech act is so. This is not the case of an inference. From an analytical point of view, both inferences and arguments can be approached as

[7] In this, I am following the classic characterization due to Toulmin, Rieke and Janik (1984), for whom argumentation is "the whole activity of making claims, challenging them, backing them up by producing reasons, criticizing those reasons, rebutting those criticisms, and so on" (p. 14).

[8] See Mohammed (2016) for a review and critical discussion of the many goals that theoreticians have considered central to argumentation.

[9] Bermejo-Luque (2011) has convincingly contended that the constitutive goal of argumentation is to show a target-claim to be correct. Although I do not share all the details of her proposal, I am indebted to her for the discussions we have maintained in relation to this topic.

[10] I am grateful to an anonymous reviewer for raising this doubt.

products and the analysis can be focused on their logical proprieties. But acts of arguing have conditions of felicity or, as I have also put it, of pragmatic correctness, to which acts of inferring are not subjected. In particular, an act of arguing requires, together with the acts of adducing a reason and of drawing a conclusion (or of raising a claim), that these acts be connected through a further act, namely, a warranting act that is performed by the speaker and, if recognized by the interlocutors, legitimizes the step from reason to claim. This is not to say that the interlocutors take the resulting argument as strong enough or convincing. But the speech act can be recognized as an act of arguing. If the warranting act is lacking, we would not say that the speaker is arguing. She would be presenting two speech acts without an argumentative connection.

Whenever speaker and interlocutors recognize that an act of arguing has taken place, this speech act can be assessed on different levels. In many cases, a very relevant dimension of assessment corresponds to the fulfilment of certain objective conditions. These objective conditions can be determinant in establishing that an act of arguing is good, cogent, solid, etc. For example, if the speaker says: “It is raining, you should take your umbrella”, an objective condition for the speech act to be good is that it is actually raining. But these conditions have a different character from the pragmatic conditions of correct performance. In the latter, together with the conditions for verdictive speech acts (the acts of adducing reasons, and of concluding or raising a claim), the corresponding conditions for the pragmatically correct performance of a warranting act have to be fulfilled, in order for the speech act to be possibly recognized as an act of arguing.

In the case of acts of inferring, from a logical point of view it is in principle possible to assess if the step from premises to conclusion has been made in accordance with some rule. Alternatively, it is also possible to assess whether correctness is transferred from reason to claim. This assessment can be accomplished without attributing any further act to the agent. My intuition is that what we have here is a process, not an action performed, where there are no conditions of pragmatically correct performance that should be taken into account, and on which it would depend that the act of inferring is such an act.

2. Speech as an Inferential Activity

The idea that inferring is an essential mechanism in the interpretation of utterances is an explanatory hypothesis widely held in neo-Gricean pragmatics. Communication is also a rational activity, in that it requires from the speaker to choose the most promising means for her to convey her communicative intentions to a hearer. As is well known, Grice defined the notion of speaker’s meaning as a complex, reflexive intention, in which the speaker has the intention to induce an attitude in their audience, together with the intention that her first intention be recognized by the audience, and the further intention that the recognition of the first intention be in part the reason that the audience has to adopt the purported attitude.

When Grice states the third clause in his definition of speaker's meaning, he writes, "*U* intended the fulfillment of the intention mentioned in (2) to be at least in part *A*'s reason for fulfilling the intention mentioned in (1)" (Grice, 1969, p. 153). The speaker, *U* had the intention that the recognition of her first intention, namely, to induce an attitude in the audience *A* were, "at least in part", the reason *A* has to have the attitude. Grice does not clarify the notion of reason, or of "having a reason" that he is assuming in the quoted essay. As it stands, the notion seems to require from the audience an explicit awareness of the speaker's intentions for her utterance to be successfully communicated. And it fulfils a clear normative role, namely, that of making of the audience's induced attitude a rational, justified one.

Grice's emphasis on seeing communication as a rational activity also becomes manifest in his theory of implicatures. The capacity to carry out inferences plays, as is well known, an indispensable role in the particular case of conversational implicatures. In Grice's model of communication, the meaning of what is said (a semantic level of meaning) is supplemented with an additional level of implicaturated meaning. Implicatures get communicated by virtue of an inferential process in which inferences are guided by the cooperative principle and its maxims. Although the type of inferential processes that allow hearers to grasp implicaturated meaning do not need to be conscious, Grice claimed that conversational implicatures must be calculable, i.e. that it should be possible, at least in principle, to carry out an explicit reconstruction of the inferential process that covers the steps from the conventional meaning of the words used, together with the cooperative principle and any available information (linguistic and non-linguistic) to the communicated meaning. This reconstruction was not aimed at describing a real, psychological process. Grice's idea is that

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless the intuition is replaceable by an argument, the implicature (if present at all) will not count as a CONVERSATIONAL implicature. (Grice, 1975, p. 50)

Here, the concept of argument that Grice takes into account is in line with his concept of inference (as seen in the preceding section). In his posthumous (2001), Grice considers the difficulty of connecting ordinary reasonings to patterns of complete argument which are valid by canonical standards, by which he means that a systematization by formal logic could be expected. He then distinguishes two concepts of rationality, those of "flat" and "variable" reason. The first one is manifested through a (non-degree-bearing) capacity of applying inferential rules that are transitions of acceptance in which transmissions of satisfactoriness are to be expected (including non-deductive cases). Variable reason is of the kind we can find exemplified in real life. Flat reason is not only manifested in variable reason, but provides an inferential base for determining the nature of variable reason itself.

It seems safe to interpret Grice as seeing flat reason as an abstract, unconditioned capacity and the source of the inferential rules that play the normative role of directives in our ordinary reasoning. And, since this flat reason manifests itself in variable reason, the latter is the kind of rationality that can be granted to our ordinary reasonings. If this interpretation is correct, then the requirement that conversational implicatures must be capable of being worked out in the form of an explicit argument is two-fold. Firstly, this methodological procedure can guarantee that the candidate implicaturated meaning satisfies the directives of rationality in communication. It does so by showing how the steps from data to implicature meet forms of transition that are acceptable principles of inference, as assessed according to the requirements of flat reason. Secondly, Grice's point of view seems not merely that of the speaker, whose communicative intentions can be expected to be known for her, nor the point of view of a theoretician formulating an empirical hypothesis about the speaker's intentions. The possibility of an explicit reconstruction guarantees the rational availability of the intended implicature for the audience. It is thus the audience's point of view, together with the assumption that speaker and audience share a common rationality, what makes the communication of implicatures possible. Flat reason, and variable reason understood as a manifestation of the former, guarantee that the same standards are available for speaker and audience.

Grice sets the requirement of explicit calculability only for conversational implicatures. It is worth remembering that he considered linguistic meaning to be a standardization or conventionalization of communicative intentions, and took the linguistic meaning of a sentence to express a complete proposition with complete truth conditions. In contrast to some recent neo-Gricean views in contemporary pragmatics, he did not endorse the view that has been stated as the thesis of underdeterminacy of linguistic meaning. According to this thesis, the linguistically encoded meaning of an utterance inevitably underdetermines its explicitly communicated propositional content (see, e.g., Sperber & Wilson, 1986; Bezuidenhout, 1997; Carston, 2002; Recanati, 2004). For Grice, and from an abstracted point of view, the semantic meaning of what is said by uttering a sentence was to be equated with the truth conditions of the sentence, and these truth conditions were also supposed to be linguistically codified. Any additional non-truth-functional meaning would be communicated meaning and should thus be obtained in the form of implicatures. The level of pragmatic, implicaturated meaning includes not only conversational, but also conventional implicatures. In this latter case, there is a conventionalization of meaning as linked to certain expressions, but this meaning does not contribute to the truth conditions of the utterance and is seen by Grice as pragmatic. All this should allow us here a generalization: the requirement according to which it must be, in principle, possible to recover the meaning of an utterance by means of a rational reconstruction (by working out an argument, in the sense above) must be applicable to the complete meaning of the uttered sentence, both to its semantic and pragmatic levels. Semantic meaning and logical form are guided by linguistic codification (and are thus so susceptible of reconstruction); pragmatic, communicated meaning (implicaturated meaning) is susceptible of being worked out, in the form of an explicit inference.[11]

Other approaches in neo-Gricean pragmatics have suspended this requirement in what concerns the level of semantic meaning, termed what is said or explicature. A common idea in these theories is that the recovery of the content of an utterance in context involves additional pragmatic processes of conceptual enrichment, which allow the interpreter to obtain a complete representation of the logical form of the speaker's utterance. Success in communication depends on the interpreter's being able to sufficiently grasp by these means the speaker's communicative intentions. It is commonly held that certain processes are associative, heuristic and non-inferential, even if there is disagreement as to what components are recovered in this form. Yet, in order for these different proposals to give plausibility to their hypotheses, an important methodological resource is that of explicitly reconstructing in theoretical terminology a plausible inferential path from the explicit information available (linguistic and non-linguistic) to the purported complete meaning of the utterance.

Even if there are empirical research and solid arguments giving support to the view of neo-Gricean pragmatics as an empirical, psychological theory of utterance interpretation, my point here is that Grice's approach was not empirical and psychological, but analytic and philosophical.[12] He aimed at providing a rational representation of meaning in communication, under the assumption that communication must be seen as a rational activity and also as a cooperative one, inasmuch as it is orientated to goals. In my view, Grice's tenet that communication is an inferential activity cannot be detached from his core assumption that commu-

[11] In my view, Grice considered it safe to assume that the meaning of what is said is given with the linguistic codification of the uttered words (i.e. certain linguistic conventions), together with a determination of the referents of referential expressions, and the time and place of the utterance (cf. Neale, 1992, p. 520). Neale also explains that when the sentence uttered is in the indicative mood, "what is said will be straightforwardly truth-conditional". And, where the sentence is in the imperative or interrogative mood, what is said "will be systematically related to the truth conditions of what U would have said, in the same context, by uttering the indicative counterpart" (Neale 1992, p. 521). This does not entail that there cannot also be linguistically conventionalized meaning that is pragmatic, as is the case of conventional implicatures. Conventional implicatures do not contribute to the truth-functional meaning of what is said, and thus belong to the pragmatic level of meaning.

[12] According to Carston (2005), three different general tendencies can be distinguished in contemporary pragmatics. Those following Grice see it as a philosophical project; other views concentrate on its interaction with grammar; finally, cognitive pragmatics focuses on an empirical psychological theory of utterance interpretation (she refers to them as the Gricean, neo-Gricean and relevance-theoric). Here I am focusing on the first and third projects, since, as far as I know, it is here that an appeal to inferential processes plays a main role.

nication is a manifestation of reason and hence, communicated meaning must be capable of being explicitly represented by means of assessable inferences.

The discussion so far suggests a hypothesis that aims to relate utterances in context, communicatively used, with the inferential character of pragmatic processing. A tentative formulation would be the following.

(H1) *Hypothesis 1*. Communication in speech is an inferential activity to the extent that it is calculable, i.e., to the extent that it is, in principle, possible to recover the pragmatic meaning of an utterance in context by means of a series of explicit inferences—and eventually, by means of an argument justifying that the corresponding meaning be ascribed to the utterance.

But notice that this explicit, rational reconstruction does not need to have psychological realization in the interactants' minds. It fulfils a normative role, that of justifying the assignment of a certain pragmatic meaning to the utterance. Moreover, it allows us to see the interlocutors as rational and as competent in deploying this rationality in their speech and action. I have suggested that this perspective is not merely that of an individual speaker who intends to convey their communicative intentions, but that of an audience which interprets the speaker's utterance with the help of a common rational capability. The explicitation of the pragmatic meaning of an utterance in context, its explicit recovering by means of a reconstructed inferential process is not a representation of the speaker's cognitive context, or that of the audience. This methodological requirement situates the recovering of the speaker's meaning in the interpersonal and social context of what can be linguistically explicitated and normatively assessed by means of explicit reasoning and argumentation.

3. The Argumentative Nature of Language and of Discourse[13]

Hypothesis 1 would seem to be questioned, in a straightforward way, by other theoretic models dealing with pragmatic meaning and linguistic communication. Notably, inferentialism contends that a sentence's meaning can be accounted for by considering its inferential relations with other sentences. Another relevant theoretical view that can be seen as against H1 is the theory of argumentation within language.[14] My aim in this section is to critically consider some of the

[13] A terminological precision is needed here. In linguistic pragmatics, discourse is the process of meaning-creation and interaction, either in writing or speech (cf. McCarthy, 2001, p. 96). It is thus a notion belonging to the pragmatics of language. Although my discussion takes discourse, particularly in speech, as its target, the consideration that the semantic level of language is argumentative in its very nature has import on my own views and I also address this perspective.

[14] The reason why the two theoretical views here considered can seem in conflict with hypothesis 1 is that it does not accord a constitutive role to the inferences that lead to the ascription of a particular meaning to an utterance. As already stated, these inferences are

main ideas in each model, in order to suggest in the next section an alternative view of pragmatic meaning and linguistic communication that does not need to endorse the idea that speech and language are argumentative in nature.

In the version of inferentialism due to Brandom (1994; 2000), a sentence's inferential relations are bestowed by the agents' normative attitudes or commitments (and entitlement to those commitments) in the practice of giving and asking for reasons, i.e. of making assertions and challenging or evaluating the assertions of others. Assertions are the minimal units of language for which we can take responsibility within this practice. The inferential relations that result can thus be seen as conferred by the very practice of giving and asking for reasons. Moreover, the semantic content of a sentence is itself the product of its inferential relations. Propositions are what can serve as premises and conclusions of inferences, which means that they stand in need of reasons. Brandom contends that it is by virtue of their use within this practice that sentences acquire their semantic contents, as resulting from the inferential relations in which those sentences stand.

Brandom's normative pragmatics gives to social practices, notably to the discursive practice of giving and asking for reasons a constitutive value in the institution of semantic content. The representational properties of semantic content are explained as consequences of the practice of inferring, which is seen as essentially social. In this sense, the traditional representational vocabulary has an expressive role, namely, that of making inferential relations explicit in virtue of the way in which it figures in *de re* ascriptions of propositional attitudes. Brandom aims to so account for the objectivity of concepts, inasmuch as the representational vocabulary (words like "of", "about", "represent") specifies the particular inferential structure that the practice of giving and asking for reasons must have in order for this practice to institute norms of application that answer to the facts.

This form of inferentialism thus equates semantic content with inferential import, which in turn must be seen as instituted by the social practices of arguing and inferring. It represents a powerful proposal in setting a notion of semantic meaning that results from those practices. Notwithstanding this, there are, I think, two points that raise doubts as to Brandom's theoretical success. The first one concerns the perspectival character of asserting. The second is related to the pre-eminent role played by assertions with respect to other types of speech acts.

Regarding the first point, Brandom claims that the game of giving and asking for reasons has a perspectival nature in a double sense. On the one hand, the "score" of commitments and entitlements corresponding to each interlocutor is socially kept and, given that everyone can have non-inferentially acquired commitments and entitlements corresponding to different observable situations, no two interlocutors will have exactly the same beliefs or acknowledge exactly the same commitments, and thus the same score cannot be attributed to each of

seen as rational reconstructions that, as such, legitimize or justify the corresponding ascription, but do not constitute meaning as such.

them. On the other hand, scores are also kept by each interlocutor, so that part of the activity of giving and asking for reasons consists in keeping track of the commitments and entitlements of other interlocutors. Brandom writes, "What *C* is committed to according to *A* may be quite different, not only from what *D* is committed to according to *A*, but also from what *C* is committed to according to *B*" (Brandom, 1994, p. 185). As a result, a sentence's inferential relations are also ultimately relative to each interlocutor's perspective. This perspectival character of the practice of giving and asking for reasons raises doubts as to its epistemic efficiency. Even if a common structure is accorded to the practice, in Brandom's account there seems to be no normative requirement which is independent of the interlocutors' perspectives and with which these must comply. Nor is it apparent how an argumentative exchange should help the interlocutors to agree on a common conclusion, given the irreducibly perspectival character of their respective ascriptions of commitments. Since the propositional content of a claim or commitment can be specified only "from some point of view" (Brandom, 1994, p. 197), and it would be different for different interlocutors occupying different perspectives, its epistemic import is at stake.

The second point that raises doubts concerns the pre-eminent role assigned to assertion and the subordination to it of other possible types of speech acts. Brandom's normative pragmatics accounts for different speech acts in terms of how the corresponding performances affect the commitments (and entitlements to those commitments) acknowledged or otherwise acquired by those who perform the speech acts. But, at the same time, he writes, "Performances count as propositionally contentful in virtue of their relation to a core class of speech acts that have the pragmatic significance of claims or assertions" (1994, p. 629). In my view, this form of subordination, which is entirely coherent with the inferential role semantics that Brandom has put forward, cannot do justice to the concept of speech act *qua* illocution that originates from Austin (1962). Within this latter framework, the felicity conditions for the correct performance of illocutionary acts must be kept apart from the semantic dimension of analysis of those acts. Although acts of asserting can bring about certain obligations and rights related to their justification and assessment, and thus to perform them can give rise to entering the game of giving and asking for reasons, this possibility also affects other types of speech act. And conversely, the correct performance of an illocution different from an assertion does not necessarily seem to be in a constitutive dependency with the assertive speech acts with which it could be related; this performance necessarily depends on the set of (pragmatic) correctness conditions which make of the speech act the illocution it is, as these conditions are socially known or interpersonally acknowledged. In the next section, I suggest the idea that speech acts bring about certain obligations and rights which have a dialectical character; but I think that this fact cannot give enough support to the thesis that Brandom defends.

Another theoretical view seemingly in conflict with Hypothesis 1 is the theory of argumentation in language set forth by Anscombre and Ducrot (1976; 1988). According to these authors, sentences (and not merely the utterance of those sentences) have argumentative connections with each other that cannot be seen as inferred (in a formal-logical way) from their informative contents. They contend that such argumentative relations have to be seen as a "brute fact" within language (*langue*), not derived from its use. The semantic value of a sentence consists in the sentence's putting forward and imposing certain argumentative viewpoints. This thesis finds support by showing how the meaning of words constrains the dynamics of discourse and how a fact can be understood in different ways depending on the linguistic formulation chosen to communicate it. In a more detailed way, it is alleged that the semantic value of a sentence is distributed in asserted value and presupposed value, which means that an assertion of the sentence conveys pieces of information that can be either asserted or presupposed. According to Anscombre and Ducrot, both values are argumentative, in that they introduce certain argumentative constrictions by allowing or prohibiting certain types of conclusion.

It is worth considering some examples in order to have a clear idea of the theoretical tenets in play. In the sentence

1. Je pars demain, puisque/car tu dois tout savoir [I am leaving tomorrow, since you need to know everything] (Anscombre, Ducrot 1976, p. 7)

the connective *puisque* (alternatively, *car*) imposes a point of view according to which the second part of the sentence, "tu dois tout savoir", must be seen as informing of the reason that explains the first part of the sentence. Given that it is not possible to make sense of the explicitly asserted sentence, "Je pars demain", as being the fact that is explained by means of "tu dois tout savoir", it must be inferred that there is a presupposed content, namely, a semantic representation of the act that the speaker is performing, "Je t'annonce que" (I announce to you that), which the second part of the sentence explains:

1' (Je t'annonce que) Je pars demain, puisque/car tu dois tout savoir [(I announce to you that) I am leaving tomorrow, since you need to know everything].

A second example concerns the comparative expressions *aussi… que* [as… as] and *le*/*la même* [the same]. Let's consider

2. Pierre est aussie grand que Marie [Pierre is as tall as Marie].
3. Pierre est de la même taille que Marie [Pierre is the same height as Marie] (Anscombre, Ducrot 1976, p. 10).

Here, the authors say that the two sentences are *quasi*-synonyms. But their negations do not have the same behaviour. Compare:

2’ Pierre n’est pas aussie grand que Marie [Pierre is not as tall as Marie] (meaning: Pierre is shorter than Marie);

3’ Pierre n´est pas de la même taille que Marie [Pierre is not the same height as Marie] (meaning: Pierre is either taller or shorter than Marie).

The semantic difference between both expressions, *aussi… que* and *le/la même*, affects the informative content in the negative construction, but not in the affirmative one. This difference determines the conclusions that are logically pertinent in each case.

In general, the authors claim that the discursive articulation between an argument-sentence and a conclusion-sentence is always made effective by virtue of general principles that they term *topoi*, which cannot be seen as formal, deductive principles of inference. They clarify this last point by explaining that, if from a sentence *A* another sentence *B* follows, it is not because *A* points out to a fact *F*, *B* to a fact *G*, and the existence of *F* makes *G* unavoidable. Rather, it is because sentence *A* presents fact *F* in such a way as to make legitimate the application of a *topos* (or of a chain of *topoi*) leading to a sentence *B* in which a linguistic casing for fact *G* can be discerned. The general thesis states then that the meaning of a sentence is the set of *topoi* whose application is authorized by the sentence in the very moment of its utterance. Whenever a speaker chooses to utter a sentence (rather than another), she is choosing the exploitation of certain *topoi* (and not others). In this sense, the semantic value of a sentence consists in its imposition of certain argumentative points of view before the facts (cf. Anscombre, Ducrot, 1994, p. 207; 1988, Chap. v, Sec. 4).

It seems to me that, from a more overarching perspective, some of the “brute facts” of language that the theory of argumentation within language is studying could be analysed in alternative theoretic terms, e.g. as conventional implicatures (in the terminology of Grice, 1975) or even implicitures (see Bach, 1999). Some others, provided that the corresponding expressions contribute to the truth conditions of the utterance, would be taken to be part of the meaning of what is said or explicature (in neo-Gricean pragmatics). My interest here is not to proceed to such a discussion, but to take at its face value Anscombre and Ducrot’s idea that their theory captures an argumentative value which is already present in language. Contrary to Brandom’s normative pragmatics, here the origins of those values cannot be traced back to the use of language, but are located at the semantic level of meaning and have to be seen as primitive data. In Brandom’s theory, the inferential contribution of certain expressions (including the logical connectives) is a consequence of the material inferences[15] that are socially attributed or

[15] The notion of material inference, as developed by Brandom, stems from Sellars (1953). In opposition to formal inferences, which are a function of the syntactic structure of language, material inferences do not depend only on syntactic structure, but are based on internal conceptual relations. A well-known example is the inference from “It is raining”

otherwise acknowledged in the practice of giving and asking for reasons. As we have seen, this idea generalizes to a notion of meaning as the content that results from the contribution made by expressions to the inferential relations of the sentences in which they occur. The theory of argumentation within language proceeds in the opposite direction. Here, the argumentative relations that an utterance of a sentence may have with others are constrained by the argumentative value of the sentence used and of the words that compose that sentence.

In the case of Brandom's inferentialism, I have suggested that the theory unduly extends certain conditions characterizing the speech act of assertion to other speech acts. In Anscombre and Ducrot's theory, what seems to underlie their proposal is a reluctance to see the use of language as conferring meaning, together with an assignment of meaningfulness to the term "argumentative" that places the notion at the semantic level. The authors refuse to use the term "inferential" because they take it to refer to formal-deductive inferences. Yet it seems to me that, taking into account the wider notion of inference we have considered above, what Anscombre and Ducrot are aiming at is a notion of inferential import that is codified in language and can thus be seen as part of the conventional meaning of words and sentences. But I think we should resist the idea that this conventional meaning is argumentative in a strict sense.

If argumentation is seen as a communicative activity, as I have been endorsing here, then only in discourse, either in speech or written form, can we find acts of arguing. For only in the activity of using language do we adduce reasons in support of a claim, draw a conclusion, or otherwise object, criticize and oppose an argument, etc. Moreover, from the perspective introduced by Hypothesis 1, any consideration whatsoever about the inferential or argumentative character of our sentences, assertions and speech acts is a consideration on whether and how the corresponding relationships should be reconstructed. In my view, this type of reconstruction is guided, in its turn, by an effort to understand and justify or assess our speech actions.

My suggestion is that both Hypothesis 1 and the above considerations can find articulation and support in an approach to discourse that takes into account its normative dimension. In the next section, my aim is to make explicit the main features of such an approach. In so doing, I shall be assuming that a piece of written discourse can also be analysed in the terminology of speech acts, and, therefore, that the same theses can be applied to it.

4. The Normative Dimension of Speech

By referring to the normative dimension of speech, I am pointing to the way in which our illocutions bring about certain obligations and commitments, entailments and rights, and similar normative stances. In this concern, I am endors-

to "The streets will be wet". It is the web of material inferences in which a word or expression is involved in that determines its meaning.

ing the Austinian approach to speech act theory that has been put forward by Sbisà (2002; 2006; 2009). According to this view, speech acts can be characterized by saying how they change the social and interpersonal context of the interlocutors. These changes affect the interlocutors' normative positions by modifying certain obligations, responsibilities and commitments; as well as rights, authorizations and licenses, as these are socially recognized and/or mutually ascribed.[16] Sbisà contends that these changes in the interlocutors' normative positions can only be effected if there is interpersonal or social recognition of the fact that they have been produced. In this sense, the effects can be seen as conventional. She suggests that in this way, Austin's (1962) original idea that there are conventional procedures explanatory of the illocutionary force of speech acts and of their conventional effects can be generalized to ordinary, non-institutional speech.

The Austinian framework outlined above[17] can be applied to the case of assertion in those cases in which asserting is an illocution (pre-eminently, a verdictive speech act, in Austin's terminology).[18] This is in general the case of making a claim, and also in particular that of adducing reasons. In illocutionary acts of asserting, the speaker presents herself as cognitively competent and incurs the obligation to give the reasons that could support her claim, if and when this is required by her interlocutors. Correspondingly, her interlocutors acquire the right to ask for justification, express doubts and objections, or otherwise accept the assertion. Whenever they recognize and accept the speaker's assertion, they become entitled to make other assertions (and possibly other speech acts as well) that are based on or supported by the former. What I would like to highlight here is that acts of asserting introduce certain obligations and rights (and other similar normative positions) that have a dialectical character. By this I mean that these obligations and rights are fulfilled and exercised as new moves in the argumentative dialogue. They comprise the obligation to justify, the right to critically question the assertion, and also the authorization to other assertions that are supported by it.

Assertion is not the only illocutionary act that brings about dialectical obligations and rights. Illocutions in general can be described by saying how they change the normative stances of the interlocutors, and some of these are, in my view, dialectical rights and duties. For example, acts of advocacy (which belong to the group of exercitives) presuppose some form of authority or authorization on the part of the speaker and assign to the interlocutors the right to accept or otherwise question this presupposition, as well as to accept or question the rea-

[16] Cf. Witek's (2015), for an accurate presentation and defence of this approach. Witek puts forward an original view which emphasizes the interactional effects of speech, contending that the force of an illocution depends on what counts as its interactional effect (see also Witek, 2019).

[17] I have also tried to present and develop this framework in former works, by applying it to presumptions, the dynamics of discourse and speech acts in deliberation (Corredor, 2017; 2019; 2020).

[18] My precaution here is related to the possibility of using some speech acts of the assertive family to perform a different act from that of a verdictive, e.g., in narrative fiction.

sons given in support of the advocated case (a person, organization, idea, etc.). Here, certain dialectical rights are in force. But there are other cases of exercitives where the effected changes do not need to have a dialectical character. For example, in cases of institutional acts such as a judicial sentence or an arbitral decision, provided the speaker's authority is granted, the conventional effect of the illocution is related to assigning (or cancelling) rights or obligations to other interactants. But this effect does not need to be seen as dialectical, as allowing or requiring a new argumentative move. In commissive acts, such as a promise, the Austinian approach takes it that they presuppose the recognized capacity to perform the act on the part of the speaker; moreover, they bring about the speaker's commitment or obligation to comply with her promise, and assign to the interlocutors the right to a legitimate expectation that the promise will be fulfilled. Here again, the obligations and rights brought about by the performance of the illocution need not be seen as dialectical.

Notwithstanding this, to the extent that our illocutions are recognized as introducing changes in the normative positions of the interlocutors, it is possible for those interlocutors to assess how the obligations and rights so introduced are fulfilled. Moreover, it becomes legitimate to ask the speaker for justification, before granting their recognition. In this way, the normative dimension of speech makes possible a dialectical practice of justification and assessment of our illocutionary acts. In my view, this does not entail that speech has an argumentative nature. But it seems to me right to say that illocutions are performed in virtue of the recognition, social or interpersonal, of certain duties and rights, some of them of a dialectical character.

The above considerations give support to a second hypothesis, which would complement the first one (H1). It could be formulated as follows.

(H2) *Hypothesis 2*. The normative positions that we recognize and assign each other with our speech acts comprise obligations and rights of a dialectical character. They also make possible a dialectical practice of justification and assessment of our speech acts.
This normativity of speech does not bring with it, however, that the semantic contents or pragmatic meaning of our utterances have an inferential or argumentative nature.

If H2 is correct, then we should resist the idea that it is the inferential or argumentative potential of a sentence what yields its semantic meaning.[19]

In the approach to speech acts endorsed here, the idea that discourse, in writing or speech, is essentially argumentative can be clarified by taking into account the conventional effects and conditions of correct performance that make of an illocution the illocution it is. In the particular case of acts of asserting, the Austinian approach makes explicit the justificatory obligation undertaken by the

[19] I am grateful to an anonymous referee for suggesting this clarification.

speaker and the corresponding dialectical rights acquired by her interlocutors. Other forms of illocution also comprise rights and duties of a dialectical character, as pointed out above. Moreover, the fact that our speech acts are subject to conditions of correctness and in need of recognition allows for their justification and critical assessment. But from that it does not follow that argumentation is the basis of meaning either at the semantic or pragmatic level.

5. Conclusion

I have examined some relevant proposals in contemporary pragmatics and in the semantics of language in order to consider two theses that relate language and communication to inference and argumentation. According to the Gricean framework, communication is an inferential activity. I have tried to clarify the notion of inference that can be originally attributed to Grice, and explored its possible applicability to the communication of meaning. I have also taken into account the constitutive role that acts of giving and asking for reasons play in normative pragmatics. Finally, I have studied the main thesis put forward by the theory of argumentation within language, according to which the semantic import of words and sentences is in part an argumentative value. In my discussion, I have argued for a twofold hypothesis. Firstly, what makes of communication an inferential activity is given with its calculability, i.e. with the possibility to recover the pragmatic meaning of utterances by reconstructing a series of inferences or an explicit reasoning. In this light, arguing is a practice of adducing and evaluating the reasons that justify (or could justify) what is communicated. Secondly, the normative stances that we recognise and assign to each other with our speech acts comprise obligations and rights of a dialectical character. However, I have suggested that this fact does not presuppose or entail an inferential or argumentative nature of speech.

REFERENCES

Anscombre, J.-C., Ducrot, O. (1976). L'Argumentation dans la Langue. *Langages*, *10*(42), 5–27.

Anscombre, J.-C., Ducrot, O. (1988). *L'argumentation dans la Lange* (2nd Ed.). Liège: Pierre Mardaga Editeur.

Anscombre, J.-C., Ducrot, O. (1994). *La argumentación en la Lengua*. Madrid: Gredos.

Austin, J. (1962). *How to Do Things with Words*. Oxford: Clarendon Press.

Bach, K. (1999). The Myth of Conventional Implicature. *Linguistics and Philosophy*, *22*(4), 327–366.

Bermejo-Luque, L. (2011). *Giving Reasons: A Linguistic-Pragmatic Approach to Argumentation Theory*. Dordrecht: Springer.

Bezuidenhout, A. (1997). Pragmatics, Semantic Underdetermination and the Referential-Attributive Distinction. *Mind*, *106*(423), 375–409.

Blair, J.A. (2012). Rhetoric, Dialectic, and Logic as Related to Argument. *Philosophy and Rhetoric*, *45*(2), 148–164.

Boghossian, P. (2014). What is Inference? *Philosophical Studies*, *169*(1), 1–18.

Brandom, R. (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

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

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

Carston, R. (2005). Relevance Theory, Grice, and the Neo-Griceans: A Response to Laurence Horn's "Current Issues in Neo-Gricean Pragmatics". *Intercultural Pragmatics*, *2*(3), 303–319.

Corredor, C. (2017). Presumptions in Speech Acts. *Argumentation*, *31*(3), 573–589.

Corredor, C. (2018). The Dynamics of Conversation: Fixing the Force in Irony. In M. Witek, I. Witczak-Plisiecka (Eds.), *Normativity and Variety of Speech Actions* (pp. 140–158). Leiden: Brill.

Corredor, C. (2020). Deliberative Speech Acts: An Interactional Approach. *Language and Communication*, *71*, 136–148.

Eemeren, F., van. Grootendorst, R. (2004). *A Systematic Theory of Argumentation*. Cambridge: Cambridge University Press.

Frege, G. (1979). Logic. In H. Hermes, F. Kambartel, F. Kaulbach (Eds.), *Posthumous Writings* (pp. 1–8). Chicago: University of Chicago Press.

Frankish, K. (2010). Dual-Process and Dual-System Theories of Reasoning. *Philosophy Compass*, *5*(10), 914–926.

Green, M. (2018). Assertion and Convention. In S. Goldberg (Ed.), *The Oxford Handbook of Assertion*. Oxford: Oxford University Press.

Grice, H.P. (1969). Utterer's Meaning and Intention. *The Philosophical Review*, *78*(2), 147–177.

Grice, H. P. (1975). Logic and Conversation. In P. Cole, J. Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41–58). New York: Academic Press.

Grice, H. P. (2001). *Aspects of Reasons*. Oxford: Clarendon Press.

Hitchcock, D. (1985). Enthymematic Arguments. *Informal Logic*, *7*(2), 83–97.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Strauss, Giroux.

Labinaz, P., Sbisà, M. (2018). Argumentation as a Dimension of Discourse. *Pragmatics & Cognition*, *25*(3), 602–630.

McCarthy, M. 2001. *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.

Mohammed, D. (2016). Goals in Argumentation: A Proposal for the Analysis and Evaluation of Public Political Arguments. *Argumentation*, *30*(3), 221–245.

Neale, S. (1992). Paul Grice and the Philosophy of Language. *Linguistics and Philosophy*, *15*(5), 509–559.

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

Sbisà, M. (2002). Speech Acts in Context. *Language and Communication*, *22*(4), 421–436.

Sbisà, M. (2006). Communicating Citizenship in Verbal Interaction. In H. Hausendorf, A. Bora (Eds.), *Analysing Citizenship Talk* (pp. 151–180). Amsterdam: John Benjamins.

Sbisà, M. (2009). Uptake and Conventionality in Illocution. *Lodz Papers in Pragmatics*, *5*(1), 33–52.

Sperber, D., Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Toulmin, S. (2003). *The Uses of Argument*. Cambridge: Cambridge University Press.

Toulmin, S. E., Rieke, R. D., Janik, A. (1984). *An Introduction to Reasoning*. New York: Macmillan.

Sellars, W. (1953). Inference and Meaning. *Mind*, *62*(247), 313–338.

Wenzel, J. (1992) Perspectives on Argument. In W. L. Benoit, D. Hample, P. J. Benoit (Eds.), *Readings in Argumentation* (pp. 121–143). Berlin: Foris.

Witek, M. (2015). An Interactional Account of Illocutionary Practice. *Language Sciences*, *47*, 43–55.

Witek, M. (2019). Coordination and Norms in Illocutionary Interaction. In M. Witek, I. Witczak-Plisiecka (Eds.), *Normativity and Variety of Speech Actions* (pp. 66–98). Boston: Brill.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 65–80
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.05
Article

PALLE LETH [*]

SPEAKER'S REFERENT AND SEMANTIC REFERENT IN INTERPRETIVE INTERACTION

SUMMARY: In this paper I argue that the notions of speaker's reference and semantic reference—used by Kripke in order to counter the contentious consequences of Donnellan's distinction between the referential use and the attributive use of definite descriptions—do not have any application in the interpretive interaction between speaker and hearer. Hearers are always concerned with speaker's reference. Either, in cases of cooperation, as presented as such by the speaker or, in cases of conflict, as perceived as such by the hearer. Any claim as to semantic reference is irrelevant for the purposes of communication and conversation. To the extent that the purpose of semantic theory is to account for linguistic communication, there is no reason to take definite descriptions to have semantic reference.

KEYWORDS: definite descriptions, speaker's referent, semantic referent, semantics/pragmatics, conversational interaction, interpretation.

Introduction

There are two controversial things suggested by Donnellan in his paper *Reference and Definite Descriptions*. First, the claim that the distinction between the referential and the attributive uses of definite descriptions amounts to a semantic distinction. Second, the claim that a speaker may succeed in saying something true despite using a definite description which does not apply to the referent she had in mind. These claims are counter to Russell's influential analysis. For Russell, the surface form of sentences containing definite descriptions should not

[*] Stockholm University, Department of Philosophy. E-mail: palle.leth@philosophy.su.se. ORCID: 0000-0001-6265-2205.

mislead us into thinking that they are about particular objects. A statement of the form “The *F* is *G*” amounts, at the logical level, to a general existential statement of the form “There is one and only one entity which is *F* and that entity is *G*”. Semantically, sentences containing definite descriptions are not referential at all. In case there is no entity which corresponds to the description “the *F*”, the sentence is simply false (Russell, 1905). Strawson, in reaction to Russell’s account, certainly takes sentences containing definite descriptions to be genuinely referring. However, his official position regarding faulty descriptions is that in case there is no entity which corresponds to the description, the sentence lacks a truth value, that there is such an entity being a presupposition of the sentence (Strawson, 1950).

Donnellan invites us first to imagine that the speaker is at the site of Smith’s murder. The circumstances of the scene lead to her to the belief that the person who murdered Smith, of whom nothing further is known, is insane. Second, we are invited to imagine that a certain person called Jones is accused of the murder of Smith and that the speaker is at Jones’s trial. Jones’s behaviour in court leads her to the belief that Jones is insane. Would not the speaker’s utterance of the sentence “Smith’s murderer is insane” in these two imagined cases make two distinct claims? In the first case, the speaker would not be concerned with any particular person; she would be concerned with whomever murdered Smith. In the second case, the speaker would be concerned with a particular person, namely Jones, and her claim would be about him, whether or not he actually murdered Smith. In the latter case, the speaker uses the description “Smith’s murderer” only as a device to pick out the particular person she has in mind, namely Jones, and about whom she wants to say something, namely that he is insane. In the former case, the speaker uses the definite description to say something about the person, whoever she or he is, who murdered Smith, namely that she or he is insane, to judge from the details of the crime scence.

These are thus the intuitions which motivate Donnellan’s distinction between the referential use (the latter case) and the attributive use (the former case) of definite descriptions. This distinction does not only contradict Russell’s unitary semantic account of definite descriptions, but adds also to Strawson’s criticism of Russell. Donnellan insists that, precisely because the speaker uses the description to refer to some object that she has in mind, she may very well succeed in saying something true, even though the description does not apply to the object. So, it seems that, for Donnellan, sentences containing definite descriptions which do not properly apply to their intended referents are neither false (Russell) nor lacking a truth value (Strawson), but may actually be true.

In his paper *Speaker’s Reference and Semantic Reference*, Kripke contends that Donnellan does not present any conclusive argument against Russell’s semantic analysis of defintie descriptions. Donnellan’s referential use should be conceived of as a thoroughly pragmatic phenomenon. The semantic referent and the speaker’s referent of a definite description should be firmly distinguished. In this paper, I shall argue that speakers and hearers engaged in conversation and communication are not concerned with any such thing as the semantic reference of definite descriptions. Hearers are solely concerned with speaker's reference. Either, in cases of cooperation, as presented as such by the speaker or, in cases of conflict, as perceived as such by the hearer. Any claim as to semantic reference is irrelevant for the purposes of communication and conversation. First, I shall review Kripke's arguments for semantic reference. Second, I shall look at Kripke's so-called complex cases from the viewpoint of the interpretive interaction of speakers and hearers. I shall also have a brief look at some more recent approaches to the referential/attributive distinction where there is an unnecessary concern with semantic reference too. I shall conclude that to the extent that the purpose of semantic theory is to account for linguistic communication there is no reason to take definite descriptions to have semantic reference.

Part I

One influential way of restoring Russell's analysis of definite descriptions is due to Kripke. In his paper *Speaker's Reference and Semantic Reference*, Kripke counters Donnellan's referential/attributive distinction by distinguishing between simple and complex cases of uses of definite descriptions. In the simple case, which corresponds to Donnellan's attributive use, the speaker has the intention to refer to the unique satisfier of the description. In the complex case, which corresponds to Donnellan's referential use, the speaker has the intention to refer to a particular object in her mind. This object may or may not be the object, if any, which satisfies the description she uses. Thus, the speaker's referent—the object the speaker has a referential intention about—may or may not coincide with the semantic referent of a given definite description, i.e. the unique satisfier of the definite description. This distinction has the virtue of applying not only to definite descriptions, but also to proper names. Speakers regularly utter definite descriptions, as well as proper names, while having particular objects in mind to which they want to refer. These particular objects need not fit, nor have, the definite descriptions, or names, which speakers use. That speakers often succeed in making themselves understood in accordance with their intentions is, however, a matter of pragmatics. There is no reason to refute Russell's account from a semantic point of view.

The simple/complex distinction is supported by a distinction between what words mean, what words mean on a given occasion, and what speakers mean. This distinction is intuitive and plausible. A sentence seldom means all that the speaker wants to convey by uttering it. Kripke says:

> The notion of what words can mean, in the language, is semantical: it is given by the conventions of our language. What they mean, on a given occasion, is determined, on a given occasion, by these conventions, together with the intentions of the speaker and various contextual features. Finally what the speaker meant, on a given occasion, in saying certain words, derives from various further special in-

> tentions of the speaker, together with various general principles, applicable to all human languages regardless of their special conventions. (Kripke, 1977, p. 263)

The first level is the inherent meaning of lexical items and syntactical constructions. It is the meaning which items and constructions carry with them to each individual occasion of use. This meaning is a matter of conventions and past use. It is created by speakers collectively and therefore unaffected by the habits, idiosyncrasies and occasional intentions of individual speakers. Each item or construction has, as it were, a certain meaning potential: it can mean this or that. On an occasion of use the question is not, however, what a sentence can mean according to the conventions of the language. The question is what it means here and now, what contribution it makes to the communicative purposes at hand. This is the second level distinguished by Kripke. He says that the meaning at this level is determined by three factors: convention, intention and context. Which reasons are there to distinguish between what a sentence can mean and what it does mean on a given occasion of use? Apart from ambiguity, the most conspicuous reason is perhaps to do with indexicality. The function of some terms is not to contribute their inherent standing meaning, but to pick out particular objects or values at their occasions of use. Their conventional or linguistic meaning provides us with general rules as to how to determine their occasional reference. These terms thus mean one thing according to the conventions of the language and another thing according to their contexts of use (cf. Kaplan's [1977] distinction between character and content). The third level is about what speakers mean when using sentences. It is clear that speakers may mean more than can be read off from their words, even if these are complemented by intention and context. Much additional meaning which hearers perceive utterances to have is not to be tied to the words of the sentence but to general considerations about the speaker's intentions. These are what Kripke calls special intentions.

Kripke's own example will illustrate these levels. One burglar says to another: "The cops are inside the bank". The word "bank", according to the conventions of the language, can mean commercial bank as well as river bank. This is relevant to the first level above. What does the word mean on this occasion of use? This is the second level. It is determined by convention (either commercial or river bank) together with context and intention. In this case, the word "bank" is used to mean "commercial bank", whether this is conceived of as determined by context or intention. Moreover, the burglar in uttering this sentence might well have a further purpose. For instance, by uttering the sentence he might want to propose to the other burglar to split. But "this is no part of the meaning of his words" (Kripke, 1977, p. 263). In this case, it is by knowing first the meaning of the speaker's words that the hearer can understand the speaker's further purpose in uttering them.

Kripke suggests that it is the last level which is relevant in order to address the referential/attributive distinction. According to the conventional meaning of a sentence containing a definite description, it means "There is a unique object such that it is *F* and *G*". The occasional meaning of such a sentence includes as its semantic referent whatever object happens to fit the description. But, of course, a speaker may use a definite description in order to refer to the particular object which she wants to talk about. This object may not even satisfy the description. Kripke suggests that the speaker's referent belongs to the speaker's meaning and is no part of the meaning of the speaker's words, no more than the burglar's proposal to split is part of the meaning of "The cops are inside the bank".

The simple/complex distinction is also supported by another general distinction, namely the distinction between a speaker's general and specific intentions.

> In a given idiolect, the semantic referent of a designator (without indexicals) is given by a general intention of the speaker to refer to a certain object whenever the designator is used. The speaker's referent is given by a specific intention, on a given occasion, to refer to a certain object. If the speaker believes that the object he wants to talk about, on a given occasion, fulfills the conditions for being the semantic referent, then he believes that there is no clash between his general intentions and his specific intentions. (Kripke, 1977, p. 264)

This distinction is also very plausible. Certainly a speaker has, with regard to the designators of her language, general intentions such that she uses this designator to refer to that object and that designator to refer to this object. Certainly she also has, whenever she is about to use one of her designators in order to talk about a particular object, the specific intention to refer to the particular object she wants to talk about. In most cases, she will use the designator which according to her general intentions refers to the particular object she has the specific intention to refer to. But, naturally, it may happen, for various reasons, that she uses a designator which according to her general intentions refers to an object different from the one she now wants to talk about. If so, there will be a clash between her different kinds of intentions.

The distinctions of levels of meaning and of intentions which Kriple identifies are intuitive and plausible. The account of simple and complex cases presents us with a picture as to how Donnellan's intuitions can be handled while preserving Russell's analysis of definite descriptions. They seemingly permit us to relegate the referential/attributive distinction to the realm of pragmatics. There are indeed good arguments for the view that there is no reason to count the referential/attributive distinction as anything but pragmatic. However, these arguments do not by themselves establish that there is any reason to take definite descriptions to have semantic reference.

Kripke and many other theorists take it as a matter of course that definite descriptions have semantic meaning or reference. The reason is probably that the semantic reference of definite descriptions appears to be due to certain matters of fact. First, there are certain facts of linguistic meaning. To use the classic example which we will soon come back to, the definite description "the man drinking martini" does as a matter of fact mean the man drinking martini, in the sense that the community of English speakers regularly use these words in such a way that they have acquired the meanings they have, and syntax or rules of composition tell us the certain meaning the whole phrase has. Second, to whom this description applies in the context is also a factual matter. For of the people present at the party at which the sentence is uttered it is either true or false that they are the man drinking martini. Linguistic meaning and factual circumstances are both independent of the speaker's referential intention. It seems then that a definite description comes to acquire its referent in a way similar to the way pure indexicals often are thought to acquire their semantic values. "I", "now" and "here" pick out persons (speakers), times and places by virtue of meaning and circumstances. This speaks in favour of taking the reference of definite descriptions to be factually determined. The facts of meaning and of circumstances are certainly indubitable. Given the propositional content of a sentence, its truth value is a factual matter. The question here, however, is what should be taken as the content of the sentence. This is possibly not a factual matter. Perhaps definite descriptions, when used by speakers to refer to objects which they have in mind, do not have anything but speaker reference. This is what I shall attempt to show by considering the use of definite descriptions in interpretive interaction.

Part II

The Primacy of Speaker Intentions

Let us now take a look at the referential/attributive distinction from the point of view of speakers and hearers engaged in communication. Confronted with a speaker's utterance of a sentence containing a definite description, the hearer will hardly be concerned with the linguistic meaning of the sentence as such. The hearer's concern is not with what the sentence means according to the conventions of language, lexical content and syntactical rules. The hearer's concern is with what the sentence means here and now. The occasional meaning of the sentence which the hearer is concerned with seemingly corresponds to the second of the levels which Kripke distinguished. How does the hearer conceive of this occasional meaning? Does it appear to her as the conventional meaning of the sentence which is to be determined and complemented by the speaker's intention and the context of the utterance, as Kripke suggests? Rather, the hearer takes a direct interest in what the speaker wants to convey. Her goal is to know what contribution the speaker is making to the ongoing conversation, and the communicative purposes that the speaker and the hearer are involved in. For the hearer, the occasional meaning of the sentence is the speaker's intended meaning. The speaker's intention is, as such, inaccessible to her. The hearer uses what she knows about the conventional meaning of the sentence and about the context in order to come up with a hypothesis about the speaker's intention.

This interpretive procedure does not imply the unimportance of linguistic meaning. Linguistic meaning is in most cases the principal clue to the speaker's intended meaning. But it does imply that the hearer does not proceed at determining the meaning of the sentence independently of coming up with an hypothesis concerning the speaker's intended meaning. For the hearer to know that definite descriptions may be used to state things about whatever satisfies the description, and also to state things about a certain object the speaker has in mind, is certainly important in order to come up with a hypothesis regarding the speaker's intended meaning. But for this purpose it is completely unnecessary to determine whether the semantic meaning of definite descriptions is attributive or referential. The sentence is not truth evaluated in abstraction from the speaker's intention. The hearer's question is not whether the sentence expresses something true in the context at hand, but whether the speaker expresses something true. In the case of definite descriptions it is not incumbent on the hearer first to tell what is said and then reason from what is said to the speaker's meaning.

This is true also of Kripke's distinction between general and specific intentions. The hearer may be convinced that the speaker has general intentions concerning the designators in her language. These intentions will not however interest her as such. The hearer's interest is oriented towards the speaker's specific intention, i.e. what the speaker wants to refer to by her use of the designator here and now. Her interest in the speaker's general intention is only to the extent that it contributes to the satisfaction of her interest in the specific intention.

Similar remarks apply to some more recent theorists' referential approach to definite descriptions. Devitt takes the fact that definite descriptions are regularly used as referring devices to speak in favour of their referentiality's being a feature of their conventional meaning. Definite descriptions are thus to be regarded as ambiguous: the semantic meaning of definite descriptions is attributive as well as referential (Devitt, 2007). Jaszczolt goes a step further. Even if definite descriptions at the linguistic level can be used quantificationally as well as referentially, they are most often used referentially. The referential reading is thus not only conventional, but actually default (cf. also Capone, 2011). Jaszczolt uses the following example:

The best architect designed this church (Jaszczolt, 2005, p. 106).

She comments:

> [I]n [this sentence], "the best architect" normally refers to a particular, known, identifiable individual. In the context of conversation, such as, for example, when the interlocutors are looking at the Sagrada Família in Barcelona, this salient reading is the one where the description refers to Antoni Gaudí. (Jaszczolt, 2005, p. 106)

It is perhaps the case that "the best architect" normally is referential. Nevertheless, for the hearer engaged in communication with the speaker, what meaning is default and what meaning is semantic is of limited concern. The hearer's question is what the speaker uses it for here and now. To know what the default interpretation of definite descriptions is does not answer the question of what the speaker uses the definite description to say on a particular occasion. To know that the literal meaning of definite descriptions is attributive as well as referential does not help the hearer in knowing what the speaker means by her use of a given definite description. It is certainly important for hearers to know that definite descriptions are used by speakers, both to make general existential statements and singular statements, but whether the latter kind of use is semantic or pragmatic is not important. It might also be helpful for hearers to know that definite descriptions most frequently are used to make singular statements. That piece of knowledge might be useful when coming up with an hypothesis about the speaker's intention. To go from the empirical observation that definite descriptions regularly or even most frequently are used as referential devices to the theoretical claim that the referential reading is default could certainly be important, but should not eclipse the fact that this has no regulatory role to play in the interpretive interaction of speakers and hearers.

In sum, the hearer does not have to determine what is literally or semantically said by a sentence containing a definite description in order to make an hypothesis about what object the speaker wants to refer to. In other words, it is not necessary to determine the semantic referent of a definite description in order come up with a hypothesis about the speaker's referent. It is not at all necessary to let the linguistic meaning of the sentence give rise to a semantically expressed referent in order to calculate the speaker's referent. For the hearer, the question as to whether definite descriptions at the semantic level express general or singular propositions and, if singular, whether the attributive or referential are uninteresting for the hearer engaged in understanding what the speaker means. Linguistic meaning serves as no more than an important clue as to what the speaker means here and now. The hearer's interest in the conventional meaning of the sentence is no more than instrumental.

Complex Cases

So far I have insisted that hearers take a direct interest in the speaker's intended meaning. In those cases which Kripke describes as simple cases, this interpretive attitude will not have any distinctive consequences. For in those cases there is no difference between what the words mean on this occasion and what the speaker wants to mean by them. We must therefore consider what Kripke describes as complex cases.

In complex cases, there is a clash between the speaker's general and specific intentions. The speaker has the specific intention to refer to Jones, who is not drinking martini. Due to faulty knowledge, however, she uses the definite description "the man drinking martini" concerning which she has the general intention that it refers to unique martini drinkers. What if the speaker uses a definite description to refer to a person which does not satisfy the description? How does such a clash appear to the hearer? Let us take Sainsbury's depiction of the scenario as the background of our discussion:

> Donnellan […] argued that we could recognize a referential use of a definite description "the *F*" by the fact that the speaker could thereby refer to something which is not *F*. If one takes this line, one will be tempted to count an utterance of "The man drinking martini is drunk" as true if Jones is drunk and is the object of the speaker's referential intentions, even if Jones has nothing but water in his martini glass. This ruling is not compulsory. In such a case, assuming the circumstances to be of the most ordinary kind, the speaker intended to refer to a martini-drinker but failed. We are not compelled to say that this failure really amounts to success in referring to a non-martini-drinker. […] Suppose (as before) that Jones is the object of the speaker's intentions and that there is also a unique martini drinker, Smith. One could not fault a hearer who took the utterance to be true just if Smith is drunk. If this is a faultless interpretation, it must have correctly identified what the speaker said. (Sainsbury, 2006, p. 415)

The speaker wants to say about a certain person whom she knows under the name of Jones that he is drunk. She thinks that the hearer does not know that the person whom she wants to talk about is called Jones. Therefore the speaker has recourse to a description of Jones which she thinks will help the hearer to identify the person she wants to talk about. Luckily, Jones, unlike the other guests, as far as the speaker knows, has a martini glass in his hand. So the speaker thinks that the utterance of the sentence "The man drinking martini is drunk" will do the job. However, the hearer is better informed about the real distribution of glasses and liquids among the guests. She knows that there are two martini glasses around. One is filled with martini and is in the hands of a certain person called Smith. The other martini glass is filled with water and is handled by a person which is otherwise unknown to the hearer. This is the person whom the speaker knows as Jones and wants to talk about. In such a scenario it is clear that the definite description "the man drinking martini" properly applies to Smith and not to Jones, as the speaker falsely believes.

When describing the scenario, Sainsbury uses notions such as "counting as true", "failure", "success" and "correct identification of what is said". How will the hearer handle this scenario and what notions will she have recourse to? There are several possibilities which we will consider in turn.

Jones as Referent

Let us first imagine a scenario where the hearer directly takes the referent to be Jones. The hearer certainly thinks that the description "the man drinking martini" applies to Smith. However, the hearer is also presented with simultaneous additional evidence as to the speaker's intended referent. These factors speak against the speaker's wanting to refer to Smith. For instance, Smith is not in the vicinity and the speaker gestures in the direction of the person holding a martini glass with water in his hand. The overall evidence suggests to the hearer that the speaker wants to talk about Jones. So she takes the predication of drunkenness to regard this person.

Did the speaker in such a case fail or succeed to refer to Jones? Did the hearer correctly identify what the speaker said? Will the hearer make the distinction between the meaning of the speaker's words on this occasion and what the speaker meant? I doubt that the hearer will put things in these terms. In case her hypothesis that the speaker wanted to talk about the water-drinking person is not contradicted by their future conversation, the hearer will probably think that she managed to guess at the speaker's intended referent, even though the linguistic means used by the speaker were not adequate. The hearer may be perfectly aware that the words of the speaker, understood along conventional lines, indicate a different referent. But the hearer will hardly be concerned with determining the meaning of the speaker's words on this occasion. That would be irrelevant to her purpose. The natural interest of the hearer is in the speaker's intention. Therefore, her whole effort will be directed at the speaker's referent. In order to arrive at the speaker's referent, it is not necessary to establish the semantic referent or unique satisfier of the definite description. The hearer takes an interest in what the speaker's words, according to the conventions of language and various contextual factors could mean on the occasion in question, because that serves her ultimate purpose, which is to know what the speaker means by those words. But it would be very peculiar for the hearer to proceed in determining what the speaker's words do mean, as a matter of semantical fact somehow composed of convention, intention and context. For, what purpose would that serve? The semantic referent, i.e. the object satisfying the definite description, is of no concern for the hearer taking an interest in the speaker's referent.

This interpretive attitude is considered by Strawson, as Donnellan points out in an interesting footnote. In a reply to Sellars, Strawson says that in some cases "if forced to choose between calling what was said true or false, we shall be more inclined to say that it was true" (Strawson, 1954, p. 227).

Strawson continues by means of an example:

> if I say, "The United States Chamber of Deputies contains representatives of two major parties", I shall be allowed to have said something true even though I have used the wrong title, a title, in fact, which applies to nothing. (Strawson, 1954, p. 227)

In this case the speaker is misdescribing the United States Congress. Strawson proposes to deal with cases like this by the notion of an amended statement. The hearer understands that the speaker by her use of the misnaming description "the United States Chamber of Deputies" wants to refer to the United States Congress. The hearer amends the speaker's original statement accordingly. It is the amended statement which is assessed for truth or falsity; the original statement is left aside: "we are not awarding a truth-value at all to the original statement" (Strawson, 1954, p. 227).

Donnellan presents two objections to the notion of amended statement. First, he points out that it is unclear which description the hearer will be using in her amended statement. The description which according to the hearer is suited for picking out the speaker's intended referent may be a description which the speaker is unaware of. For example, because she is misinformed about the correct designation. It is thus very difficult, if not impossible, to establish any amended statement. Donnellan's second point is, however, that this is inconsequential, in so far as "the notion of the amended statement really plays no role anyway" (Donnellan, 1966, p. 294*n*).

When setting out to understand the speaker's original statement, the hearer goes directly for the speaker's intended referent. The hearer's first question is what the speaker wanted to refer to. Once she thinks she knows this, she directly asks whether the speaker's referent has the properties the speaker ascribes to it. There is no reason at that point to go back and amend the original statement and evaluate it for truth or falsity. The speaker's original statement is only used as a springboard for arriving at the speaker's intention. Not only is the original statement not truth evaluated, as Strawson admits, but neither is any amended statement's truth evaluated. It is the speaker's intended meaning which is directly truth evaluated. The role of the original statement is purely instrumental; it is not even amended, it is simply left aside.

It should be stressed, of course, that generally the hearer's getting at the speaker's specific intention will be facilitated by the speaker's using the designator in accordance with her general intention. But if, for some reason or other, there is a clash between her general intention with the designator and her specific intention with it, the hearer, in most cases, is not particularly concerned with the speaker's general intention.

Smith as Referent

Let us now consider a different kind of scenario. It is, of course, equally possible that the hearer takes the predication to be about the person she knows as Smith. The hearer knows that there is one unique martini drinker at the party and the description used by the speaker, "the man drinking martini", accordingly applies to him. There is, as far as the hearer is aware, no evidence which points in a different direction. Consequently, the hearer takes the referent of the definite description directly to be Smith. Is this not a case where the hearer is concerned with the semantic referent of the speaker's definite description? As we will see, she is rather concerned with Smith *qua* intended referent.

Imagine now that, even though the hearer initially takes the predication to be about Smith, the continuation of the conversation makes the hearer aware that the speaker, by her use of the definite description in question, wanted to refer to the water-drinking person. The most natural thing for the hearer to do is to adapt her previous understanding. She might certainly think that the speaker was mistaken about who is drinking martini, and that she herself is better informed. She might also think that the speaker's expression of her thought was faulty and that she herself had the best reasons to take the speaker to be talking about Smith. She might even think that this certainly was an incorrect use of the definite description. But now, given that she knows whom the speaker wanted to refer to, such issues are of little importance. The question for many hearers is not whether they understand speakers according to the rules of language, but whether they understand speakers according to their wishes. Once the hearer is confident that she understands what the speaker means, there is no further issue as to what the meaning of the speaker's words on this occasion of use is. The hearer did not understand the speaker as the speaker wanted to be understood initially, and even though the fault was entirely with the speaker, there is no particular reason to insist on that fact. After all, even though the hearer initially took the referent to be Smith, unlike the previous case that we considered, she eventually behaves in the same way as when she immediately took the referent to be Jones.

Conflict

We must now consider whether hearers always leave the issue of the proper satisfaction of the definite description aside to the benefit of the speaker's intention. Are hearers always adapting to speakers? Strawson said that in some cases hearers, forced to choose between calling the speaker's utterance true or false, say it is true and that a speaker may be allowed to have said something true, even though the description which she used is faulty. We have so far considered cases where the hearer does precisely this. Strawson admits though that this hearer attitude is not universal, even if he does not say anything about in which cases hearers take this attitude. What forces hearers? When are speakers allowed to have spoken the truth? Sainsbury seems to have a different hearer attitude in mind when he says that hearers could not be faulted for understanding definite descriptions according to their strict content. If the hearer insists on the faultlessness of her interpretation, she may not be up to allowing the speaker to have spoken the truth. There are of course cases where the hearer is interested in pointing out to the speaker that there is a difference between the speaker's intended referent and the referent according to the content of the description used. Interpretation is not always collaborative, cooperative and charitable; it may be antagonistic and conflictual (see, e.g., Marmor, 2008; Lee & Pinker, 2010; Asher & Lascarides, 2013). Would the hearer for that reason claim that the speaker attempted to refer to Jones but failed, or that she the hearer correctly identified what the speaker said? Would the hearer say that one thing is what the speaker meant, another thing is what the words of the speaker meant on a given occasion? Would she make the distinction between the speaker's referent and the semantic referent of the definite description? It is now time to speak of these cases.

Let us then imagine that the hearer wants to insist that there was a difference between whom the speaker wanted to refer to and whom the definite description that she used actually applied to. The hearer points out to the speaker that, the description having the linguistic content that it has and the circumstances being as she knows them to be, as a matter of fact, the speaker referred to Smith or the semantic referent of her definite description was Smith. What would the speaker say in response to this claim?

The speaker might admit that, as a matter of fact, the semantic referent of her definite description was not the person she intended to refer to. She might even admit that, as a matter of fact, she had, unbeknownst to herself, referred to Smith. But after having granted this point, the speaker would presumably draw the hearer's attention to other matters of fact. First, as a matter of fact, her intention was to refer to Jones, Jones being the person in her mind. The speaker is, of course, aware that this matter of fact cannot appear as such to the hearer. The definite description that she used was not, after all, particularly helpful in displaying this matter of fact to the hearer. Still, it is an important matter of fact. And now, at last, it is made manifest to the hearer. Second, the speaker would certainly allege as another matter of fact that conversation and communication are about getting at the speaker's point. Given the hearer's engagement in communication with the speaker, it would be quite difficult for the hearer to deny. Given that the hearer now is informed about the speaker's intention, why should she insist that the speaker originally did not convey accurately the referent that she had in mind? Third, the speaker may question the foundation of the hearer's claim. To whom the description "the man drinking martini" actually applies is a factual matter. What is the guarantee that the hearer is right about who is drinking martini and who is not? Are they going to have sips in order to ascertain the semantic referent? Is it not obvious to everyone concerned that such a manœuvre would be ridiculous and serve no sensible purpose at all? In short, the speaker's response to the hearer's claim that the semantic referent of her definite description was Smith would be that this claim is possibly false and in any case irrelevant.

The upshot is that the hearer insisting on the semantic reference of a definite description would have to motivate the interest she takes in semantic reference. It seems to me that the hearer's only reason for insisting on the semantic reference is to justify her own interpretation. The hearer's insistence on the semantic reference is a way to enforce the faultlessness of her interpretation. But if this is the hearer's purpose in insisting on semantic reference, it should be stressed that this purpose can be served without invoking the notion of semantic reference. By avoiding semantic reference, the hearer would escape the charges of possible falsity and irrelevance.

The important point for the hearer in the kind of case we are considering is that the linguistic meaning of the definite description and the circumstances being what they were, the hearer was completely justified in taking the referent to be Smith. This point can be made in a straightforward way by the claim that the hearer was justified in thinking that the speaker wanted to refer to Smith. As such it is a claim about the speaker's reference. It is highly relevant and the possible falsity of the claim as to the satisfier of the description is inessential. If the hearer construes her taking the reference to be Smith as a claim about the semantic reference, she will run the risk of irrelevance and also be challenged as to the foundation of this contention. The hearer might easily avoid all this by construing her taking the reference to be Smith as a claim about the speaker's reference.

Discussion

Sainsbury ends his considerations with the following remark: "A hearer is not obliged, in order to reach a proper understanding, to chase through the various possible errors of which a speaker might be guilty" (Sainsbury, 2006, p. 415).

The notion of proper understanding here is intriguing. We are to imagine the hearer telling the speaker that her interpretation of the speaker's utterance represents the correct identification of what the speaker said, despite the speaker's protestations that her intended meaning was different. An understanding unrelated to the speaker's intention may perhaps in some sense be proper, but in any case it is hardly appropriate. For what purpose would it serve the hearer to have reached an understanding which has no function in the conversational interaction? A hearer is certainly not obliged to chase through the various possible errors of which the speaker might be guilty. Neither is she obliged to listen to the speaker at all. But if she listens to the speaker, if she is engaged in conversation with her, what errors is she not prepared to chase through?

Capone says, in the same vein as Sainsbury: "a speaker who says 'The man drinking a martini' intending to refer to the man drinking water is literally saying something false (however charitably interpreted)" (Capone, 2011, p. 157*n*; cf. Bontly, 2005). But what is literal meaning to a charitable interpreter? Even uncharitable interpreters ought to couch their claims in terms of what they (pretend to) perceive as speaker reference, as I suggested above. If not, they will not appear to be engaged in conversation at all, in which case their interpretations will hardly be given any weight. The interpretive attitude suggested by Sainsbury and Capone is, in fact, opposed to the natural interests of a hearer.

I have attempted to show that speakers and hearers engaged in conversational interaction do not take interest in such a thing as the semantic reference of a definite description. They are solely concerned with the speaker's reference. When the hearer is confronted with a definite description, she wants to know whether the speaker intends an attributive use or a referential use, and in the latter case, the hearer's question is not what satisfies the description, but what is in the speaker's mind. This is the case also in conflictual interpretation. Even if an object different from the one intended by the speaker satisfies the description and the hearer on that account holds the speaker responsible for referring to the particular object uniquely satisfying the description, the hearer is not concerned with the actual semantic reference of the definite description. She is rather concerned with the faultlessness of her interpretation, which is not to be concerned with semantic reference, but with what the hearer had good reasons to perceive as the speaker's reference. A claim about semantic reference is not a sensible move in conversational interaction.

But does the possible fact that speakers and hearers engaged in conversation take no interest in such a thing as the semantic reference of definite descriptions prove that there is no such thing? It might be pointed out that in general, it is not the case that our lack of interest in a thing is an argument for the inexistence of the thing. However, the relevance of such an objection seems to presuppose that the reference of definite descriptions is something of a natural kind. The traditional question concerning definite descriptions is what the structure of their meaning is, what kind of contribution they make to sentences containing them. Do definite descriptions contribute to existential general statements of singular propositions? If the latter, are they about what satisfies the descriptive conditions or what the speaker is thinking of? It is perhaps possible that such a meaning could be discovered by semantical investigations. But in many statements the question appears rather as a matter of decision than as a matter of discovery. Witness Sainsbury, who says that "[w]e are not compelled to say that this failure really amounts to success" (Sainsbury, 2016, p. 415) and Strawson, on the other hand, saying that "if forced to choose between calling what was said true or false, we shall be more inclined to say that it was true" (Strawson, 1954, p. 227). If the purpose of semantic theorizing about definite descriptions is to account for linguistic communication by means of them and it is independently established that the determination of semantic refence has no role to play in the hearer's arriving at the speaker's intended referent, nor in the conversational interaction between speaker and hearer, it seems, in any case, that any discovery in this regard would be inconsequential and, consequently, no decision is called for. We had better stop asking what the semantic reference of definite descriptions is, for such a notion plays neither a theoretical nor a practical role.

Conclusion

In order to counter Donnellan's contentious suggestions that the referential/attributive distinction is semantic and that a speaker may say something true although the object she wants to talk about does not satisfy the definite description she uses, Kripke has recourse to the distinction between semantic reference and speaker's reference. I have argued that the category of semantic reference is not applicable to the interpretive interaction between speaker and hearer. Most hearers on most occasions of their interpretive career are cooperative: they want to know what speakers mean. Therefore, semantic reference is of no concern for them. Even when hearers take a conflictual approach to interpretation insisting that the words meant something different from what the speaker meant, they had better not invoke semantic reference. For in order for their claim to be of concern for the speaker, they had better couch it in terms of what they perceived as the speaker's reference. If the task is one of "handling ordinary discourse" (Kripke, 1977, p. 255), as Kripke himself says, I think semantic reference is unnecessary.

There is no reason to say that the referential/attributive distinction is semantic. At the linguistic level, definite descriptions are items which speakers use to make general existential statements as well as singular statements (cf. Moldovan, 2019). There is no reason to be concerned with any semantic level at all. Whether a given use of a definite description is referential or attributive is a question of what the speaker means, i.e. it is a wholly pragmatic issue. In so far as the question for the hearer is not whom the description applies to as a matter of fact, but to whom the speaker wanted to refer, the hearer does not ask whether the speaker's sentence is true, but truth-evaluates the speaker's intended meaning. The speaker might say something true by the utterance of a sentence containing a faulty description, in the sense that she is taken by the hearer to convey something true.

REFERENCES

Asher, N., Lascarides, A. (2013). Strategic Conversation. *Semantics & Pragmatics*, *6*(0), 1–62.

Bontly, T. D. (2005). Conversational Implicature and the Referential Use of Descriptions. *Philosophical Studies*, *125*(1), 1–25

Capone, A. (2011). The Attributive/Referential Distinction, Pragmatics, Modularity of Mind and Modularization. *Australian Journal of Linguistics*, *31*(2), 153–186.

Devitt, M. (2007). Referential Descriptions and Conversational Implicatures. *European Journal of Analytic Philosophy*, *3*(2), 7–32.

Donnellan, K. S. (1966). Reference and Definite Descriptions. *The Philosophical Review*, *75*(3), 281–304.

Jaszczolt, K. M. (2005). *Default Semantics: Foundations of a Compositional Theory of Acts of Communication*. Oxford: Oxford University Press.

Kaplan, D. (1977). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In J. Almog, J. Perry, H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). New York & Oxford: Oxford University Press.

Kripke, S. (1977). Speaker's Reference and Semantic Reference. *Midwest Studies in Philosophy*, *2*, 255–76.

Lee, J. J., Pinker, S. (2010). Rationales for Indirect Speech: The Theory of the Strategic Speaker. *Psychological Review*, *117*(3), 785–807.

Marmor, A. (2008). The Pragmatics of Legal Language. *Ratio Juris*, *21*(4), 423–452.

Moldovan, A. (2019). Descriptions and Tests for Polysemy. *Axiomathes*. doi:10.1007/s10516-019-09445-y

Russell, B. (1905). On Denoting. *Mind*, *14*(56), 479–493.

Sainsbury, R. M. (2006). The Essence of Reference. In E. Lepore, B. C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 393–421). Oxford: Clarendon Press.

Strawson, P. F. (1950). On Referring, *Mind*, *59*(235), 320–344.

Strawson, P. F. (1954). A Reply to Mr. Sellars. *The Philosophical Review*, *63*(2), 216–231.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 81–103
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.06

Article

ANDREA RAIMONDI [*]

AGAINST THE QUOTATIONAL THEORY OF MEANING ASCRIPTIONS[1]

SUMMARY: According to the quotational theory of meaning ascriptions, sentences like “‘Bruder (in German) means brother” are abbreviated synonymy claims, such as “‘Bruder (in German) means the same as ‘brother’”. After discussing a problem with Harman’s (1999) version of the quotational theory, I present an amended version defended by Field (2001; 2017). Then, I address Field’s responses to two arguments against the theory that revolve around translation and the understanding of foreign expressions. Afterwards, I formulate two original arguments against both Harman’s and Field’s versions of the theory. One of them targets the hyperintensionality of quotations and the other raises a problem pertaining to variant spellings of words.

KEYWORDS: meaning ascriptions, use/mention distinction, pure quotation, translation, hyperintensionality, variant spellings.

1. Introduction: The Problem of Special Occurrence

We use language to talk about individuals, events, times, and states of affairs. But we can also use it to talk about letters, words, sentences, and utterances. When language is used this way, a linguistic device is needed that turns language on itself. Pure quotation (henceforth, simply quotation) is one such device.

[*] University of Nottingham, Department of Philosophy. E-mail: andrea.raimondi1@nottingham.ac.uk. ORCID: 0000-0001-9125-8645.

[1] I would like to thank for their comments on previous versions Stefano Predelli, Mark Jago, Steven Barker, Andrea Bianchi, Alice Monypenny, Martina Rosola, Jansan Favazzo, the referees for this journal, and my audience at the conference *PhiLang 2019* and the University of Nottingham Postgraduate Research Seminar.

By enclosing an expression in quotation marks, we refer to (or, more colloquially, we mention) that expression. Thanks to this device, then, we can say that a certain expression has such and such properties, among which is having a certain meaning. In ordinary communication, we ascribe meanings with such sentences as the following: "'Brother' means male sibling", "'Procrastinate' means to put things off", and "'Bruder' (in German) means brother".[2, 3] In each one of these meaning ascriptions, a quotation referring to a linguistic expression[4] is the subject of "means", which is followed by an expression of our own language. The latter expression plays the role of a "linguistic exemplar" (Field, 2017, p. 8) serving the purpose of providing an example able to display the meaning of the expression referred to by the quotation on the left-hand side.

Meaning ascriptions are worth discussing for three main reasons. First of all, they seem to challenge some widespread assumptions about the traditional use/mention distinction, which will be the topic of this paper. Secondly, they involve non-extensional linguistic environments, as they do not allow substitution of coextensional expressions after "means".[5] These non-extensional environments do not necessarily involve "that"-clauses. Thirdly, meaning ascriptions are sentences that speakers use for a variety of purposes in ordinary linguistic exchanges: explaining the meaning of an expression, providing a definition, disambiguating among different meanings of a single expression, etc. However,

[2] My quotation conventions are these. Double quotation marks are used to quote expressions; single quotation marks are used to quote expressions that occur inside a quotation. So, while "my" is a possessive, "'my'" is not—it is the quotation of a possessive. Notice that in accordance with standard usage, I use double quotation marks also to report another's writing or speech; the context will always make it clear how I am using double quotation marks.

[3] There are other kinds of sentences that we use to ascribe meanings: "The meaning of 'Bruder' (in German) is brother" and "Male sibling is what 'brother' means". My discussion applies to them in the same way in which it applies to sentences like "'Bruder' (in German) means brother".

[4] Although nothing in my discussion hinges on this, my preferred view is that quotations occurring in meaning ascriptions refer to morphologically and graphically marked realization types of word types. For example, the quotations "'Bruder'" and "'Brüder'" refer to two different realization types of the same word type BRUDER, i.e., the singular and the plural realization type, respectively. This is my preferred view because meaning ascriptions are sensitive to morphological variations: if I want to specify the meaning of "Bruder" in German, I should say that it means brother, not brothers, for the latter is the meaning of "Brüder". Nevertheless, since quotations can be used to refer to tokens, nominalistically-minded philosophers may well read meaning ascriptions as involving quotations referring to word tokens (either simple tokens or realization tokens).

[5] As Kripke (1982, pp. 9–10, fn. 8) underlines, "means" may be used as synonymous with "refers to". For instance, in legal contracts, "means" is often used to specify what the technical terms stand for (e.g., "'Programme' means a programme of study for which you have received an offer"). In these cases, we have straightforward extensional environments. However, in this paper I am not concerned with this reading of the sentences at stake.

as suggested above, the topic of this paper involves the use/mention distinction. What is puzzling about meaning ascriptions is the semantic status of the expressions figuring as linguistic exemplars: as Sellars (1956, p. 24; 1963; 1974), Black (1962, Chap. 2), Alston (1963a; 1963b), Garver (1965), and Christensen (1967) noticed in passing some time ago, the mode of occurrence of these expressions is very special, for they appear to be neither regularly mentioned nor regularly used.

By way of example, consider "'Bruder' (in German) means brother". If "brother" were mentioned, and the containing ascription were taken at face-value, then the ascription would be true if and only if what "Bruder" means is the linguistic expression "brother"—assuming, as is usually done, that quotations (unambiguously) refer to linguistic expressions, signs, or any other quotable item.[6] Given that the meaning of "Bruder" is not a linguistic expression, the sentence would incorrectly turn out to be false.

One might then be tempted to understand linguistic exemplars as being regularly used expressions, which contribute the semantic values they customarily have in sentences other than meaning ascriptions. But this cannot be correct, for (again, if the ascriptions are taken at face-value) we would get wrong results. To illustrate, consider "'Procrastinate' means to put things off". If the complex expression on the right-hand side were used, in uttering the sentence we would be saying that the word "procrastinate" intends to delay certain things. In some special and perhaps bizarre contexts, we may want to say something like this; but normally, the intended interpretation of the aforementioned sentence is not this one. Hence, the view that the complement of "means" is regularly used is unable to account for the intended reading of the ascription.

To put it in a nutshell, the way linguistic exemplars occur on the right-hand side of meaning ascriptions is *sui generis*, at least *prima facie*: if these sentences are taken at face-value, such exemplars are neither regularly used nor regularly mentioned expressions. The question as to the mode of occurrence of these expressions is what I call the Problem of Special Occurrence.

[6] This is a widespread assumption among theories of the semantics of quotation. It is so widespread that it would be difficult to list all the people who endorse it; however, see Pagin & Westerståhl 2010 for a detailed discussion of the standard view of quotations and its motivations. Yet, it should be mentioned that some theorists reject this assumption and, accordingly, maintain that quotations are multiply ambiguous or context-sensitive. The views they endorse appear to be quite hospitable to the idea that quotations, in some linguistic environments, can refer to the meanings of the quoted expressions. In the final section of the paper, I shall argue that one way of solving the problem I am describing may be to embrace one such view of quotation; however, in this paper I shall not discuss this kind of solution.

Another important assumption in my arguments will be that pure quotations are semantically inert: the content of the quoted linguistic material is segregated from the content of the containing sentence. As far as I know, this assumption is accepted in the debate, and it is usually treated as an essential feature of pure quotations, as opposed to, e.g., mixed quotations and scare quotations.

According to Abbott (2003, p. 21), their awkward mode of occurrence is evidenced also by the fact that people often choose a different punctuation device when they wish to write them. For instance, Washington (1992, Ex. 12), Perry (2001, p. 59), Whiting (2013, p. 6), Glüer & Wikforss (2015) use italics, Field (2001; 2017) employs corner quotes, Kaplan (1969) introduces special meaning marks, and Garver (1965) proposes a dedicated notation (attributed ultimately to Black, 1962, Chap. 2).[7] However, Abbott also observes that there is "an intuitively obvious way to express the idea" conveyed by, e.g., "'Bruder' means brother": we may well say "'Bruder' means the same as 'brother'". This suggests a seemingly sensible way to dismiss the Problem of Special Occurrence: ascriptions should not be taken at face-value, but rather as shorthand for synonymy claims between the expressions figuring in subject and complement positions.[8] On this account, linguistic exemplars are nothing but regularly mentioned expressions. Yet, Abbott says that "usually we do not want to be so wordy" (Abbott, 2003, p. 21). However, I maintain that the point is more substantial than this: the quotational theory of meaning ascriptions is wrong and therefore it does not really provide a solution to the Problem of Special Occurrence. In the remainder of the paper, I shall present two versions of the theory, and then discuss some arguments against them.

2. Quotational Solutions

A clear exemplification of the quotational theory is suggested by Harman: "['means'] abbreviates a relational predicate *S* together with a pair of quotation marks surrounding what follows […] where *S* is such that for every expression $\ulcorner e \urcorner$, $\ulcorner e\ S\ e \urcorner$ is true" (Harman, 1999, p. 265).[9] He then suggests that *S* may be interpreted as "is synonymous with", "means the same as", or the like. More recently, Neale (2018) has observed that meaning ascriptions are "completely metalinguistic", that is, they involve reference to two linguistic expressions. In short, (1) is to be analysed as (2):

(1) "Bruder" (in German) means brother.

(2) "Bruder" (in German) means the same as "brother" (in English).

[7] Following the suggestion of an anonymous reviewer, I have decided not to use italics or any sort of punctuation mechanism (according to the reviewer, the use of italics or special punctuation after "means" is extremely rare in actual lay language).

[8] To my knowledge, the first appearance of this view is due to Johnson (1921, p. 90), mentioned by Moore in the lectures he gave at Cambridge in 1933–1934 (Moore, 1966). As far as I know, Moore is the first one to attack the quotational view. In §4 I discuss an argument inspired by one of Moore's remarks on the topic.

[9] I have added here Quine's quasi-quotation marks. I use them to mention variables ranging over expressions.

Even though in the subsequent discussion I make use of the informal rendering of the quotational analysis, here is a formal representation of (2):

(3) $\forall x$ (x is a meaning $\rightarrow$ ("Bruder" has x in German $\leftrightarrow$ "brother" has x in English))[10]

This representation literally quantifies over meanings. For argument's sake, I assume that it is possible to provide an equally good representation that does not quantify over such entities, but rather unpacks the notion of "meaning the same as" in a certain way. For example, according to Field (2001), the relation between two synonymous words is that of having equivalent meaning characteristics, which include "the inferences that govern certain kinds of sentences containing the words and […] the worldly conditions that typically lead to the assent to other kinds of sentences involving the words" (Field 2001, p. 159). On a more Quinean view, we may avoid mentioning equivalence relations, and just say that two words mean the same exactly if their meaning characteristics make it appropriate to translate one into the other (or to use them interchangeably, if they belong to the same language).[11]

The issues to which the foregoing paragraph alludes are worthy of further independent scrutiny. Yet, their topic is not the problem at stake here. Rather, the topic I want to discuss is whether or not the quotational view provides a correct solution to the Problem of Special Occurrence. If it does, (1) is nothing but a shortened version of (2) (or (3)), and hence the proposition expressed by (1) is nothing more and nothing less than the proposition expressed by (2) (or (3))—regardless of how the notion of "meaning the same as" (or some equivalent notion) is to be spelled out in detail.

[10] This formal representation is a simplified version of Field's analysis in Field (2017). I mention his actual version later on in this section, after introducing an amendment in the quotational theory.

[11] Let us set aside worries about intra-linguistic synonymy too. Mates (1950, pp. 215ff) argued that no two expressions of a single language are synonymous. Sceptics *á la* Mates have two options. On the one hand, they may embrace an error theory of meaning ascriptions, according to which all such sentences are false or lack a truth-value—perhaps with the exception of homophonic ascriptions, like "'Brother' means brother". On the other hand, they can focus on inter-linguistic synonymy (see (2)) and the related ascriptions ascribing meanings to foreign expressions (see (1)). Those who also have doubts concerning inter-linguistic synonymy can either hold an error theory or focus on more artificial examples involving English words and expressions of an argot, i.e., a language in which words in a given natural language are altered according to certain rules, like Pig-Latin. One of its rules is that for English words that begin with consonant sounds, all letters before the initial vowel are placed at the end of the word sequence, and then the suffix "ay" is added. No change in meaning occurs: "pig" becomes "igpay" in Pig-Latin, and by stipulation "pig" is synonymous with "igpay". Thus, people who are doubtful about inter-linguistic synonymy can focus on ascriptions such as "'Igpay' (in Pig-Latin) means pig".

For argument's sake, I shall also grant that the supporter of the quotational theory can provide a plausible story for the unvocalised linguistic material in (1), i.e., for the occurrence of "the same as" at some semantically relevant level of syntactic complexity. However, that story may well not be easy to come by. For one thing, the postulated relationship is surely not one of syntactic ellipsis, for (1) need not be uttered after a previous occurrence of "the same as", from which the alleged hidden occurrence in (1) would then be recovered.

Let us leave these difficulties aside and turn our attention to Harman's quotational theory. As it stands, the theory needs a refinement, because clearly (1) and (2) are not truth-conditionally identical. For instance, (1) is false and (2) is true with respect to a world in which (a) the use of "brother" and "sister" by the counterparts of actual English speakers is swapped, and (b) the use of "Bruder" and "Schwester" (i.e., the actual German translation of "sister") by the counterparts of actual German speakers is swapped. With respect to such a world, "Bruder" and "brother" are synonymous, but they mean something completely different to what they actually do, and hence "Bruder" does not mean brother (indeed, we actual speakers would say that it means sister). In order to solve this problem, one may rigidify the right-hand side of (2). In this spirit, Field (2001, pp. 158ff; 2017, pp. 6ff) suggests that (1) should be analysed as (4), of which (5) is a more formal version:

(4) "Bruder" (in German) means what "brother" actually means (in English).[12]

(5) $\forall x$ (x is a meaning $\rightarrow$ ("Bruder" has x in German $\leftrightarrow$ actually ("brother" has x in English)))

Despite the fact that reference to actuality in Field's analysis solves the problem raised for Harman's theory (and granting that Field can provide a story about the unvocalised linguistic material in (1)), there are independent arguments against both versions. Given that these arguments do not hinge on the presence or absence of the actuality operator, in my discussion I shall focus on the simpler version of the theory, i.e., Harman's. Two of the four arguments I present have been allegedly rejected by Field (2001; 2017). I shall start with them, showing that Field's replies can be challenged (§§3–4). Later, I provide two arguments that neither Field nor Harman has discussed (§§5–6).

[12] Note that Field (2001) focuses on the individual speaker with their own idiolect. For example, he claims that to say that a word means brother "is just to say that it has meaning-characteristics that are […] equivalent to the actual meaning characteristics of my term ['brother']" (Field, 2001, p. 59). However, in Field (2017), the focus of the discussion is mainly on public languages.

3. Translation

A well-known objection against sententialist theories of belief ascriptions is Church's (1950) translation argument. A parallel argument can be formulated against the quotational theory of meaning ascriptions.[13] If this theory is correct, (1) is analysed as (2), here reported:

(1) "Bruder" (in German) means brother.

(2) "Bruder" (in German) means the same as "brother" (in English).

If (1) is analysed as (2), they express the same proposition. The Italian translations of (1) and (2) are (6) and (7), respectively:

(6) "Bruder" (in tedesco) significa fratello.

(7) "Bruder" (in tedesco) significa lo stesso di "brother" (in inglese).

According to the quotational theory, (6) and (7) are thus translations of two sentences that express the same proposition. Then, also (6) and (7) express the same proposition. But this is patently false.[14]

Field objects that the argument relies on standards of translation that require reference-preservation of the parts, but "these are not the proper standards of translation in this case" (Field, 2017, p. 8). Thus, (7) does not translate (2); rather, its correct translation is (8):

(8) "Bruder" (in tedesco) significa lo stesso di "fratello" (in italiano).

Field (2001, p. 161; 2017, p. 7–8) urges that when we translate (2) we are not interested in literal translation, but rather in quasi-translation, which involves the translation of the quoted expression on the right-hand side. Similarly, he holds that the quotation marks surrounding the latter expression "don't behave quite like ordinary quotation marks" (Field, 2001, p. 161). Field does not define

[13] The target of the original argument was Carnap's analysis of belief ascriptions. Church ultimately attributes the argument to Langford (1937, p. 61). This type of argument has been used by a number of authors (for different purposes), like Lewy (1947, p. 26), Strawson (1949, p. 84) and Kneale & Kneale (1962, pp. 50–51).

[14] As for Church's original argument, Putnam (1953), Davidson (1963), and Richard (1997) observe that Carnap's analysis of belief ascriptions is not intended to capture their meaning, but only something logically equivalent, thus making Church's objection inapplicable. This problem is irrelevant here: the quotational view of meaning ascriptions is meant to be a solution to the Problem of Special Occurrence, and thus must provide semantically equivalent sentences (§2). However, the problem of hyperintensionality (§5) raises an objection against the view that the theory provides sentences that are even just logically equivalent to meaning ascriptions.

the notion of quasi-translation,[15] nor does he elaborate on the special quotation marks he mentions, except for saying that "we want [their quoted material] to be quasi-translated rather than 'literally translated'" (Field, 2001, p. 161). Field's remarks are reminiscent of a reply put forth by a number of philosophers to Church's original argument. For example, Geach (1957, p. 91–92), Dummett (1973), Burge (1978, p. 141–145), and Kripke (1979, p. 139, fn. 5) hold that what counts as correct, actual translation often includes translation of quoted expressions, in order to convey the point of the source sentence. On this view, what is crucial to meaning ascriptions and synonymy claims is not part of their semantics, but it is better seen as involved "in a convention presupposed in [their] use and understanding" (Burge, 1978, p. 146). The convention in connection with (1) and (2) directs one to interpret the sentences in the specified manner; this yields the result that (8), and not (7), translates (2).

Let me reply as follows. It is completely inessential to the argument whether or not (8) is an actually acceptable translation of (2). The argument is concerned exclusively with the semantics of (1) and (2), and not with any pragmatic "conventions presupposed in their use". Sometimes we may be interested, as Field says, in quasi-translation, but it is surely possible to be interested in literal translation as well. Literal translation requires at least preservation of character (in the sense of Kaplan, 1989), so that an expression e_1 of a language L_1 literally translates an expression e_2 of a language L_2 only if e_1 and e_2 have the same character.[16] Any notion of translation that does not meet this condition is not literal. Assuming that "brother" and "fratello" have the same character, the latter is a literal translation of the former. On the other hand, the quotation "'brother'" and the quotation "'fratello'" do not have the same character; therefore, the latter is not a literal translation of the former. If we assume that translation is compositional (at least in the case at stake), we can conclude that (8) is not a literal translation of (2). Moreover, given that literal translation requires at least character-preservation (and *a fortiori* reference-preservation, given a context), (7) is the correct literal translation of (2). And literal translation, as opposed to non-literal

[15] Nevertheless, he offers a nice non-linguistic analogy that should help us see the point: "[s]uppose that a witness before the Warren Commission described the impact of the decisive bullet by pointing at the place on his own head "where the bullet hit", i.e. analogous to the place on Kennedy's head where it hit. And suppose that in some future investigation someone is asked to give a literal account of the Warren Commission testimony; she will do so by pointing to a spot on her own head" (Field, 2017, p. 8).

[16] I am giving only a necessary condition for literal translation because there are clear cases of literal translation requiring more than character-preservation. Consider proper names: usually people think that the Italian literal translation of "Hesperus" is "Espero", and not "Fosforo", though the character of "Fosforo" is the same as that of "Hesperus" (maybe here another condition involves the history of the name, or something along these lines).

or quasi-translation, is the kind of translation we should use for the purpose of drawing semantic conclusions.[17]

In addition, echoing Salmon's (2001) remarks on Church's translation argument, I suggest that translation is here invoked merely as a device to facilitate our seeing the semantic difference between certain sentences. The argument aims at showing that (6) and (7) differ semantically, as they have different literal meanings. That is, they differ in character and therefore, given a context, they express different contents or propositions. Notice that the two sentences have different characters no matter whether or not the quotational theory is correct. Now, literal meaning should be opposed to whatever kind of information that may be inferred from it together with knowledge of English—in particular, knowledge of what "brother" means. By showing the semantic difference between (6) and (7), the argument establishes that (1) and (2) are semantically different too, even if the proposition expressed by the former may be easily inferred from the proposition expressed by the latter.

There is a desperate move that the supporter of the quotational theory may make: denying the intuitive claim that (1) is translated into Italian by (6). Since in (1) reference is made to the word "brother", any literal translation of (1) must include an expression referring to that word. Hence, despite our intuitions, (9) counts as the literal translation of (1):

(9) "Bruder" (in tedesco) significa brother.

For argument's sake, I grant that (9) is a grammatical sentence.[18] Yet, a more controversial result is obtained by reformulating the argument focusing on ascriptions of meaning to declarative sentences. Consider, for instance, (10):

[17] My distinction between literal and non-literal translations exemplifies one way of substantiating Salmon's (1986, p. 58–59, 84–85; 2001, p. 586) distinction between—on the one hand—translations that aim at preserving the semantically encoded information of a piece of linguistic material, and—on the other hand—translations that aim at preserving its pragmatically imparted information. According to Salmon, only the latter may depart from mere semantic constraints. Now, (8), contrary to (7), is a pragmatically, though not a semantically, correct translation of (2). If there is a semantically adequate translation of a sentence, we should be concerned with that kind of translation when assessing semantic aspects of the sentence, as is the case here. (Quine, 1960, pp. 26–79, 211–216; Tarski, 1983; Davidson, 1967; 1968; 1973; and Kaplan, 2004, §3 have employed this kind of translation in drawing some of their semantic conclusions). I stress the "if" because there are sentences whose translations cannot preserve all their semantic aspects: "The sentence hereby uttered is in English" cannot be translated into German by preserving both its character and the content it expresses in a context. Anyway, (1) does not include self-referential elements that may trigger impossibility results.

[18] Although I originally suspected that there may be some controversy as regards the grammaticality of (9), an anonymous reviewer suggests that the grammaticality of this sentence is totally unobjectionable, as so-called noun switches are extremely frequent in code-switching between all sorts of pairs of languages.

(10) “Hunde bellen” (in German) means that dogs bark.

Field observes that “the presence of ‘that’ is enough to indicate the special role that the sentence ‘dogs bark’ plays, [namely] its role as a content-indicator” (Field, 2017, p. 3). According to him, there is no relevant difference between meaning ascriptions for sentential and sub-sentential expressions, except for the fact that in the latter case “we have no analog of ‘that’” (Field, 2017, p. 3). If Field is right,[19] the quotational theory applies straightforwardly to meaning ascriptions for sentential expressions (as he explicitly says). Thus, (10) is analysed as (11):

(11) “Hunde bellen” (in German) means the same as “Dogs bark” (in English).

Now, the alternative version of the translation argument would rely on the idea that the Italian translations of (10) and (11) are (12) and (13), respectively:

(12) “Hunde bellen” (in tedesco) significa che i cani abbaiano.

(13) “Hunde bellen” (in tedesco) significa lo stesso di “Dogs bark” (in inglese).

If (1) is not literally translated by (6) (as the desperate objection holds), then (10) is not literally translated by (12). Rather, its literal translation is (14):

(14) *“Hunde bellen” (in tedesco) significa che dogs bark.

But (14) appear to be ungrammatical. As the advocate of the objection and I are both assuming, a literal translation is character-preserving and hence reference-preserving. If the reference of a sentence is its truth-value, as is commonly thought, (14) is not the literal translation of (10). Indeed, given that ungrammatical sentences do not express propositions, (14) does not express a proposition; *a fortiori* it lacks a truth-value, and hence is not true, contrary to (10). Obviously, if one thinks that the reference of a sentence is not a truth-value, it will suffice to say that (14) lacks a character altogether (because of its ungrammaticality), and therefore it is not a literal translation of (10).[20]

[19] Field may be right as regards ascriptions of meaning to non-indexical sentences. Things are more complicated with indexicals. If I were to attribute a meaning to the Italian sentence “Spero di mangiare presto” I could do two things. I may say that it means “I hope to eat soon”, or that the sentence, as uttered by me, means that I hope to eat soon. In the first case, I would be attributing a character to the sentence, whereas in the second case, a content (i.e., the proposition expressed on that occasion).

[20] The example in my reply to the desperate objection involves a language switch between a complementizer and a content clause. An anonymous reviewer observes that this phenomenon has been discussed (and variably assessed) in Spanish-English code-switching. Unfortunately, I was unable to find literature on Italian-English code-switching. However, in Spanish-English code-switching, there is some controversy as to

4. Knowing What a Word Means

Field attributes the second argument he rejects to Schiffer (1987, pp. 33–35; 2003, p. 47; 2008, p. 289). Schiffer's original argument did not concern meaning ascriptions; rather, it was meant to provide an objection against Davidson's (1968) paratactic account of indirect reports. However, I think that the idea at the heart of Schiffer's argument (as applied to meaning ascriptions) is ultimately to be found in a remark advanced by Moore (1966). He observes that if in saying that "Bruder" means brother all you are saying were that "Bruder" means the same as "brother", "you would not be telling anyone what the meaning of ['Bruder'] is […]. If this were all, it is an assertion you might make, even if you hadn't the least idea what ['brother'] meant" (Moore, 1966, p. 57). Thus, in saying that "Bruder" means brother, you are not just saying that "Bruder" and "brother" are synonymous expressions.[21] The thought here is that the quotational theory incorrectly implies that one can understand what an ascription says without knowing the meaning of the expression that works as a linguistic exemplar. Or, to put it more accurately, the quotational theory makes wrong predictions when applied to occurrences of meaning ascriptions that are embedded in knowledge or belief ascriptions.

To see this point clearly, we may arrange Moore's remarks in the form of an argument. If the quotational theory is correct, then (1) is shorthand for (2), here reported:

(1) "Bruder" (in German) means brother.

(2) "Bruder" (in German) means the same as "brother" (in English).

Then, if Pablo knows that (2), he knows that (1). But if Pablo is a monolingual speaker of Spanish, he does not understand the word "brother". Thus, the quotational theory predicts that Pablo knows that (1) without knowing what "Bruder" means in German. But saying that Pablo knows that (1) seems to imply exactly that Pablo does know what "Bruder" means.

Field (2001, pp. 160ff) discusses a different version of the argument. Suppose that Anna and Marco are monolingual speakers of Italian. Anna believes that "Bruder" (in German) means what "fratello" means (in Italian), while Marco believes that (2). The quotational theory apparently implies that it will be Marco, rather than Anna, who believes that (1).

whether a switch between a complementizer and a content clause is allowed (see González-Vilbazo, 2005; Hoot, 2011; Ebert & Hoot, 2018; Sande Piñeiro, 2018). Be that as it may, I should not take for granted that (14) is surely ungrammatical. Hence, the desperate move that the supporter of the quotational theory makes may not be so desperate.

[21] As mentioned in footnote 8, Moore makes this remark in his discussion of Johnson's quotational view. The original example draws on an intra-linguistic meaning ascription. I have adapted it to make it consistent with the other examples. However, the difference between inter- and intra-linguistic ascriptions is here irrelevant.

Field replies by applying the notion of quasi-translation to belief and knowledge ascriptions. Roughly, the idea is that, for every rational agent *S* and English sentence ⌜ *P* ⌝, ⌜ *S* believes/knows that *P* ⌝ is true in English if and only if *S* stands in some appropriate relation with a quasi-translation of ⌜ *P* ⌝ in a language *S* understands (Field, 2001, p. 162; 2017, p. 7–8).[22] So, if Anna believes that (1), she is in a relation of, say, acceptance with an appropriate quasi-translation of (1) or of an equivalent sentence, like (2).[23] An Italian quasi-translation of (2) is (8), here reported:

(8) "Bruder" (in tedesco) significa lo stesso di "fratello" (in italiano).

Thus, the theory correctly predicts that it is Anna, and not Marco, who believes that (1). Similarly, Pablo knows what "Bruder" means in German because he assents to a Spanish quasi-translation of (2).

However, the only rationale for preferring the notion of quasi-translation over that of literal translation is that it helps Field to handle apparent problems in his theory. Moreover, the notion of quasi-translation appears to be so coarse-grained that it may be used for too many different purposes, which casts doubts on the very notion itself. For instance, it may be used to argue that co-referring proper names are truth-preservingly substitutable in the complement clauses of belief ascriptions. One may say that, e.g., "Espero brilla nel cielo" (i.e., the Italian literal translation of "Hesperus shines in the sky") is a quasi-translation of "Phosphorus shines in the sky"; then, if Anna assents to "Espero brilla nel cielo", she believes that Phosphorus shines in the sky, even if she assents to neither "Espero = Fosforo" nor "Hesperus = Phosphorus". The conclusion may be correct, but certainly not in virtue of the arbitrary choice of treating "Espero brilla nel cielo" as a quasi-translation of "Phosphorus shines in the sky".

Field may reply that the example of Anna and Marco involves the quasi-translation of a quoted expression, while the Hesperus/Phosphorus case involves the quasi-translation of a regularly used expression. He may then stress that we are allowed to, and should, prefer quasi-translation only for cases involving quoted expressions. Notice, though, that the example of Anna and Marco does involve the quasi-translation of a regularly used expression, namely "English",

[22] Variables like ⌜ *P* ⌝ are assumed to range over declarative sentences that lack indexicals, ambiguities and pronominal devices. Field sketches the view not in terms of quasi-translation but in terms of "quasi-meaning", so that to believe that *P* is to be in a certain relation with a sentence that quasi-means the same as ⌜ *P* ⌝. However, the notion of quasi-meaning is defined in terms of that of quasi-translation: "We don't care much about 'literal meaning', if that is what is preserved in 'literal translation' […] what we care about, rather, is what is preserved in quasi-translation" which we may call "quasi-meaning, though I think it is what most would simply call meaning" (Field, 2001, p. 161).

[23] For the sake of precision, note that since Field's actual analysis of (1) is (4) (or (5)), to believe that (1) is (for him) to assent to an appropriate quasi-translation of (4) (or (5)). Nothing in my discussion hinges on this.

which is quasi-translated into Italian as "italiano" (see (2) and (8)). At this point, Field cannot object that references to English and Italian may be omitted. A word means something in a given language, and since it may mean one thing in a language and a different thing (or nothing) in another one, the analysis must make clear which language is at stake, in order to get the right truth-conditions. The same applies to any translation or quasi-translation of them.[24]

Let us take stock. The two arguments discussed so far concern translation and the understanding of foreign expressions. There are further arguments against the quotational theory that are independent of such issues, arguments that neither Harman nor Field addresses. In the next sections, I present two of them. The first one targets the hyperintensionality of quotations; the second one concerns how the quotational theory deals with ascriptions involving variant spellings of a word. Again, since they apply to Harman's version of the theory as well as to Field's, I shall focus on the former, given its greater simplicity.

5. Hyperintensionality

Pure quotations are standardly thought to be the clearest examples of hyperintionsional positions that we have in natural languages.[25] I argue that this raises a problem for the quotational theory.

First of all, a bit of terminology. For every English sentence $\ulcorner P \urcorner$, a position in $\ulcorner P \urcorner$ is hyperintensional if and only if synonymous (and hence necessarily co-extensive) English expressions are not replaceable in that position without changing the truth-value of $\ulcorner P \urcorner$. Although "lawyer" is synonymous with "attorney", the former cannot be truth-preservingly replaced by the latter in the sentence "The word 'lawyer' has six letters".

Now, the quotational theory implies that the expressions figuring on the right-hand side of meaning ascriptions are not truth-preservingly replaceable by synonymous expressions. But this implication is incorrect. Consider (15) and its quotational analysis:

(15) "Archäologie" (in German) means archaeology.

(16) "Archäologie" (in German) means the same as 'archaeology' (in English).

Call (C_1) the claim that "archaeology" and "archeology" are synonymous expressions in English, and assume it is true (according to *Collins English Diction-*

[24] Notice, however, that in using (1) we may (and usually do) omit the complement "(in German)" in ordinary language exchanges; but, again, if we want the truth-conditions to be given correctly, we should make it explicit.

[25] See Cappelen & Lepore (2007, p. 4). Predelli (2013, p. 174–177) is the only exception I am aware of: truth-preserving substitutivity of synonymous (and hence necessarily co-extensive) expressions within pure quotation marks is a corollary of his defence of truth-preserving substitutivity of all strings within pure quotation marks.

ary, “archeology” is a variant spelling of “archaeology”, in both British and American English). Then, “archaeology” in (16) cannot be truth-preservingly replaced by “archeology”, since the former expression occurs in a hyperintensional position. Hence, (17) cannot be validly inferred from (16) and (C_1):

(17) “Archäologie” (in German) means the same as “archeology” (in English).

Nevertheless, “means” does not trigger a hyperintensional position. If (15) and (C_1) are true, then (18) is true as well:

(18) “Archäologie” (in German) means archeology.

Similarly, if “Bruder” (in German) means brother, and “brother” and “male sibling” are synonymous expressions, “Bruder” (in German) means male sibling. However, we should not jump too quickly to the conclusion that meaning ascriptions are intensional in the position following “means”, where a position is intensional in $\ulcorner P \urcorner$ if and only if only necessarily co-extensive expressions are replaceable in that position without changing the truth-value of $\ulcorner P \urcorner$. Even though “to be German” and “to be German and to be English or not English” are necessarily co-extensive expressions, the former cannot be truth-preservingly replaced by the latter in the sentence “‘Essere tedesco’ (in Italian) means to be German”. Meaning ascriptions allow for the truth-preserving substitution only of synonymous expressions, where the relevant notion of synonymy is more fine-grained than that of necessarily co-extensionality; spelling out in detail what this notion exactly amounts to is a tough job, about which much has been said. However, this is an issue for another discussion; for our purposes, we just need to acknowledge that “means” does not trigger a hyperintensional position and, most importantly, that at least some substitutions are allowed in the complement position.[26] By contrast, no substitutions at all are allowed in pure quotations.

Let us go back to the argument. If the quotational theory is correct, (a) (15) and (16) stand in the analysis relation, and (b) (17) cannot be validly inferred from (16) and (C_1). Since (18) and (17) stand in the analysis relation too, (18) cannot be validly inferred from (15) and (C_1). But, as argued above, (18) can be

[26] Some recent developments in truthmaker semantics for exact entailment may be useful here. Fine and Jago (2019) offer a system in which $\ulcorner P \urcorner$ and $\ulcorner Q \urcorner$ are semantically equivalent when, roughly, they share all their truthmakers in all truthmaker models. This is one of the few systems that draws semantic distinctions between, say, $\ulcorner P \urcorner$, on the one hand, and $\ulcorner P \wedge (Q \vee \neg Q) \urcorner$ and $\ulcorner P \vee (Q \wedge \neg Q) \urcorner$ on the other (which are all classically, intuitionistically, and relevantly equivalent)—*ditto* for predicates. This kind of view may allow us to account for the intuitively obvious semantic difference between the ascriptions “‘Essere tedesco’ (in Italian) means to be German” and “‘Essere tedesco’ (in Italian) means to be German and to be English or not English”. In turn, this could help us provide a criterion for substitutivity (in the complement position) in terms of exact truthmaking.

validly inferred from (15) and (C_1): therefore, the quotational theory incorrectly invalidates the inference from (15) and (C_1) to (18).

One might reply that this objection implies the absurd conclusion that the quotation in subject position is not a pure quotation. Let us pretend, for the sake of the argument, that also in German there are two words for archaeology, namely, "Archäologie" and "Arkäologie".[27] Call (C_2) the claim that "Arkäologie" is a German expression that is not only necessarily co-extensive with "Archäologie", but also synonymous with it (whatever synonymy might be), and assume it is true. If we replace "Archäologie" with "Arkäologie" in (15), we obtain the following true meaning ascription:

(19) "Arkäologie" (in German) means archaeology.

In general, synonymous expressions appear to be truth-preservingly replaceable in the quotation figuring in subject position. This validates the inference from (15) and (C_2) to (19). Consequently, (15) is not hyperintensional in the position occupied by the quotation "'Archäologie'"; hence, the latter is not a pure quotation, as pure quotations are hyperintensional positions.

This line of reasoning is wrong. To illustrate why, suppose that (15) allows the truth-preserving substitution of synonymous expressions in the position occupied by the quotation "'Archäologie'", and thus that the quoted expression in subject position can be truth-preservingly replaced by "Arkäologie". As a consequence, any sentence resulting from the conjunction of (15) with another sentence allows the truth-preserving replacement of the quoted expression in subject position with "Arkäologie". Thus, the inference from (20) and (C_2) to (21) is valid:

(20) "Archäologie" has eleven letters and means archaeology (in German).

(21) "Arkäologie" has eleven letters and means archaeology (in German).

But (21) is false, contrary to (20). Hence, the inference is not valid. Therefore, we have no reason to think that my argument against the quotational theory can be applied to the subject position.

Recall that in his discussion of the translation argument (§3), Field maintains that quotation marks occurring on the right-hand side of meaning ascriptions "don't behave quite like ordinary quotation marks [since] we want [their quoted material] to be quasi-translated rather than 'literally translated'" (Field, 2001, p. 161). Perhaps Field would use a similar strategy to deal with the problem of

[27] I am just pretending that "Arkäologie" is an actual German word. Of course, the fact that German does not actually have two words for archaeology is irrelevant, as we may find realistic examples in other languages. I have decided not to change the example in order to be consistent with the remainder of the section and the paper, in which I extensively make use of that example.

hyperintensionality. For instance, he may argue that the quotations at stake are so special that meaning ascriptions are not hyperintensional in such positions. However, I think that introducing a special semantic category just to save the quotational theory from apparent problems puts the supporter of the theory in a bad dialectical position.

Apart from this, appealing to a special semantic category will not help us solve the Problem of Special Occurrence: the puzzle consists exactly in understanding what semantic contribution is made by an expression occurring after "means". Saying that it is a quotation that does not behave as a regular quotation clearly is not an answer to the puzzle; rather, it is a way of restating the puzzle once we have assumed that there is something quotational in the way the relevant expression occurs.

6. Variant Spellings

As already mentioned, "archeology" is a variant spelling of "archaeology", in both British and American English. Suppose that Tim and Sam are speakers of the former. Tim has always come across "archaeology", and never "archeology", while Sam the opposite. Suppose that Tim and Sam read in a German monolingual dictionary that the definition of "Archäologie" is such-and-such. Then, they use a bilingual dictionary to translate the definition as follows: "the study of human activity through the recovery and analysis of material culture". Tim and Sam understand this definition the same way. Now consider the following sentences:

(22) a. There is a word in German that means archaeology.
b. There is a word in German that means archeology.

On the basis of the procedure Tim has followed (i.e., using a monolingual and a bilingual dictionary) and his knowledge of English, he accepts (22a). Hence, on the basis of that procedure and his linguistic knowledge, Tim has learnt something about German. The same line of reasoning applies, *mutatis mutandis*, to Sam and (22b).

The procedure Tim has followed is identical to the one Sam has followed. Thus, it seems that the thing about German that Tim has learnt is identical to the thing about German that Sam has learnt. What distinguishes Tim from Sam is how they would express that thing: Tim would express it with (22a), while Sam would express it with (22b). Thus, (22a) and (22b) intuitively express the same proposition. But the quotational theory conflicts with this conclusion: the analysis of the former makes reference to "archaeology" while the analysis of the latter makes reference to "archeology".[28]

[28] If we reformulate the example focusing on the idiolects spoken by Tim and Sam, we may raise a problem pertaining again to translation. While the lexicon of Tim's idio-

Why, however, should one not insist that (22a) and (22b) are semantically different? First of all, if these two sentences expressed different propositions, perhaps we should hold something analogous as regards, say, an utterance of (22a) made by someone with rhotacism and an utterance of the same sentence made by someone without rhotacism. After all, one might maintain that if a small spelling variation of a word affects the semantics of the containing sentence, then there is no reason for us not to say that the pronunciation affects it as well; but the conclusion that it does would be patently absurd.[29]

An advocate of the quotational theory may reply as follows. Two different ways of pronouncing "archaeology" do not count as utterances of two different words; on the contrary, "archaeology" and "archeology" are two different words. Therefore, (22a) and (22b) express different propositions because they make reference to two different words. Here there may be room to argue that this reply relies on a controversial assumption concerning word individuation, namely, that words are not individuated (among other things) by their phonetic properties, or that words are not so individuated in the case at issue. Be that as it may, I do not wish to push in this direction. Instead, I want to challenge one of the implications of the view that (22a) and (22b) do not express the same proposition. To do this, I need to introduce an assumption concerning the relation between propositions and beliefs:

(PB) For every atomic English sentence ⌜ *P* ⌝ and ⌜ *Q* ⌝, if the proposition that *P* is not the proposition that *Q*, then it is possible for a rational agent to believe that *P* (in a context *c*) without believing that *Q* (in *c*).

If (PB) is true, then the view (implied by the quotational theory) that (22a) and (22b) express different propositions implies that it is possible for a rational agent to believe that (22a) (in a context *c*) without believing that (22b) (in *c*). For instance, since Tim does not know that "archaeology" is a variant spelling of "archeology", he may believe that (22a) without believing that (22b).

This does not seem correct to me. Imagine the following situation. After Tim finds out the meaning of "Archäologie" via the translation of the definition he found in the monolingual dictionary, we ask him: "So, what does 'Archäologie' mean?". He replies by uttering certain sounds: 'arçɛolo′gi in "ʤɜ:mən mi:nz ˌɑ:ki′ɒləʤi". How could we write in English what he is saying? We have two options:

lect contains "archaeology", but not "archeology", the lexicon of Sam's idiolect contains "archeology", but not "archaeology". If the quotational theory is correct, (22b) is not a sentence of Sam's idiolect that literally translates (22a), i.e., a sentence of Tim's idiolect. This result is strongly counterintuitive.

[29] Or it would be patently absurd in the case at stake. In other cases, the pronunciation may affect the proposition expressed, e.g., when the sentence contains an indexical that refers to the way the utterer utters that very sentence or some part of it ("In order to sound like a posh nobleman, you need to speak like so").

(23) a. “Archäologie” (in German) means archaeology.
b. “Archäologie” (in German) means archeology.

Even if we do not know whether he is aware of one or both spellings, we understand what he is saying, and we note no ambiguity. But if the same sounds corresponded to sentences expressing different propositions, we would notice some degree of ambiguity.[30] Since Tim’s utterance is not ambiguous, and we assume that he is speaking sincerely and in English, we would ascribe to him the belief that (23a) and the belief that (23b), as they are one and the same belief[31]—*ditto* for the belief that (22a) and the belief that (22b).

One may notice that despite the wide acceptance of (PB) (i.e., the principle I have invoked in my example), the latter is not universally endorsed. For instance, Richard maintains that “believes” expresses a “triadic relation among a person, a proposition, and a sentential meaning, the latter entity a different sort of thing than a proposition” (Richard, 1983, p. 425), a sort of “Kaplanesque character” (1983, p. 429). On this view, to believe that *P* is to be in a relation with the proposition that *P*, under a certain sentential meaning (see also Richard, 1990). One may draw on this view to argue that (22a) and (22b) differ in sentential meaning, and thus, even if the proposition that (22a) and the proposition that (22b) are not the same, it is possible to believe that (22a) without believing that (22b). However, if the supporter of the quotational theory maintains that (PB) is false, my argument may be seen as showing that they are committed to a minoritarian view of the relation between propositions and beliefs, as the majority of philosophers accept (PB).

Moreover, we may rephrase one of my observations in the form of an independent objection that does not make use of (PB). If (23a) and (23b) express different propositions, an utterance of “arçɛolo′gi” in “ʤɜːmən miːnz ˌɑːki′ɒləʤi” should be ambiguous; but such an utterance is not ambiguous—at least, for most

[30] The supporter of the quotational theory may insist that what the example shows is that we should ask Tim to disambiguate. This is in sharp contrast with our intuitions about the difference between the case at stake and one in which Tim utters certain sounds that correspond both to “I’m writing a paper on intentionality” and “I’m writing a paper on intensionality”. We would regard his utterance as ambiguous.

[31] I am assuming some form or another of disquotational principle, according to which from Tim’s assent to the uttered string of sounds we can jump to conclusions about his beliefs. The point can be restated by means of a different and unobjectionable disquotational principle conditionally linking assertion to utterance of a (string of sounds that counts as a) sentence. Regardless of whether one’s sincere assent to a (string of sounds that counts as a) sentence does or does not imply that one believes the proposition expressed, it certainly does imply that one asserts that proposition. So, we could say that in the described scenario, Tim asserted that (23a) or, analogously, that he asserted that (23b).

ordinary speakers, supporters of the quotational theory being exceptions. Hence, (23a) and (23b) are semantically equivalent—*ditto* for (22a) and (22b).[32]

7. Conclusion

If taken at face-value, sentences of the form ⌜ e (in L) means x ⌝ raise a puzzle about the way ⌜ x ⌝ occurs in them (at least on one reading), as this expression appears to be neither regularly used nor regularly mentioned. This is the Problem of Special Occurrence. Quotational approaches attempt to show that there is no problem at all; rather, the illusion of such a problem is generated by failing to see that the predicate "means" (as it occurs in meaning ascriptions) is shorthand for "means the same as" or some other predicate that expresses a relation between linguistic expressions. Once we see "means" in the right way, we have an unproblematic answer to the alleged puzzle: ⌜ x ⌝ occurs in ⌜ e (in L) means x ⌝ exactly how ⌜ e ⌝ does. In this paper, I have considered two versions of the quotational theory and I have discussed some arguments against them. In particular, I have replied to Field's responses to two arguments that revolve around translation and the understanding of foreign expressions. Then, I have provided two original arguments involving hyperintensionality and variant spellings. If these arguments are correct, this theory is wrong, and thus the Problem of Special Occurrence persists.

One might notice that the phenomenon shown by this problem is somehow opposite to the phenomenon known as mixed quotation, of which the following sentence contains a paradigmatic example: "Quine said that 'quotation has a certain anomalous feature'". Intuitively, what this sentence says is true if and only if Quine said (expressed the proposition) that quotation has an anomalous feature, and did so by uttering the words "has a certain anomalous fea-

[32] With respect to my discussion in this section, an anonymous reviewer notices that it is crucial to make a distinction between two different issues. One is whether the words "archeology" and "archaeology" quote each other; the other one is whether one knows that they quote each other. The former is a problem of semantics of quotation, whereas the latter is an epistemic issue. According to the reviewer, here the relevant issue is the epistemic one, and it requires more formal and conceptual work on modality (as applied to quotation) to be implemented in the discussion.

However, I am not completely sure that the issue here is epistemic, although I acknowledge that my example involving Tim and Sam may give the impression that it is. In my view, the relevant point is that two meaning ascriptions that differ only in that one involves "archeology", while the other one "archaeology" (in their complement position) do not differ semantically. However, if the reader thinks that the issue here is only epistemic, then they might construe the argument as one that shows that the quotational theory has unwelcome consequences as regards (what we might call) the epistemology of meaning, the issue of what one knows when one knows the meaning of a word. Intuitively, Tim and Sam know the very same thing about a certain German word; but this fact is denied by the quotational theory.

ture".[33] Hence, here we have some words that are simultaneously used and mentioned. Then, one may be wondering why there is no Problem of Special Occurrence for mixed quotation.

The answer is pretty straightforward: there is no such problem because we have, on the one side, theories of the semantics of regularly used expressions, and, on the other side, semantic accounts of quotation. A theory of mixed quotation, then, is an attempt to put the two things together, so to speak. By contrast, there are no theories that are possibly combined with one another to explain how expressions can meaningfully occur in sentences without being used nor mentioned. In other words: while we know how expressions occur in mixed quotations, we have only a negative description of how they occur when they figure as linguistic exemplars in meaning ascriptions.

Let me conclude by stressing that, as suggested in footnote 5, one way of dismissing the Problem of Special Occurrence is to reject standard accounts of quotation, according to which quotations always refer to linguistic expressions, signs, and the like. In light of my discussion of the quotational theory, one may conclude that we should embrace an account that makes quotations ambiguous or somehow context-sensitive. [34] On one such account, quotations can refer to a variety of things, including linguistic expressions, sounds, typographic forms, and—one may urge—also meanings. Therefore, one might advocate this kind of account and then take meaning ascriptions at face-value by arguing that the complement position is occupied by a quotation referring to a meaning (presumably, the meaning that the quoted expression actually has). Given that we were led to the quotational theory because of a problem concerning the mode of occurrence of the complement expression, one may say that my arguments against that theory suggest that other views of quotations need to be endorsed. Be that as it may, theories assuming that quotations can refer to a variety of things are worthy of independent scrutiny. Moreover, if we want to use such theories to address the Problem of Special Occurrence, we shall need a detailed account of how the right interpretation of the quotation in complement position is to be obtained (that is, the interpretation in which the quotation refers to a meaning, as opposed to an expression). These are topics for another discussion.

[33] Mixed quotation had not been much discussed prior to Davidson (1979), but it has recently taken centre stage in discussions of quotation. See De Brabanter (2010) for a survey of the issue from a linguist's point of view. Maier's (2017) article presents some of the most important formal semantic theories. For recent philosophical theories, see Cappelen & Lepore (1997), Recanati (2001), Gómez-Torrente (2005), and McCullagh (2017).

[34] See, for instance, Davidson (1979), Clark and Gerrig (1990), Saka (1998; 2006), García-Carpintero (2004; 2017; 2018), Gómez-Torrente (2017), Johnson (2018). Although all these theories agree that quotations can refer to a variety of things, they differ in various respects.

REFERENCES

Abbott, B. (2003). Some Notes on Quotation. *Belgian Journal of Linguistics*, *17*(1), 13–26.

Alston, W. P. (1963a). The Quest for Meanings. *Mind*, *72*(285), 79–87.

Alston, W. P. (1963b). Meaning and Use. *The Philosophical Quarterly*, *13*(51), 107–124.

Black, M. (1962). *Models and Metaphors*. Ithaca: Cornell University Press.

Cappelen, H., Lepore, E. (1997). On an Alleged Distinction Between Semantic Theory and Indirect Quotation. *Mind and Language*, *12*(3–4), 278–296.

Cappelen, H., Lepore, E. (2007). *Language Turned on Itself: The Semantics and Pragmatics of Metalinguistic Discourse*. New York: Oxford University Press.

Christensen, N. E. (1967). The Alleged Distinction Between Use and Mention. *The Philosophical Review*, *76*(3), 358–367.

Clark, H. H., Gerrig, R. J. (1990). Quotations as Demonstrations. *Language,* *66*(4), 764–805.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*(23), 685–700.

Davidson, D. (1968). On Saying That. *Synthese*, *19*(1/2), 130–46.

Davidson, D. (1973). Radical Interpretation. *Dialectica*, *27*(1), 314–28.

Davidson, D. (1979). Quotation. *Theory and Decision*, *11*(1), 27–40.

De Brabanter, P. (2010). The Semantics and Pragmatics of Hybrid Quotations. *Language and Linguistics Compass*, *4*(2), 107–120.

Ebert, S., Hoot, B. (2018). *That*-trace Effects in Spanish-English Code-Switching. In L. López (Ed.), *Code-Switching—Experimental Answers to Theoretical Questions: In Honor of Kay González-Vibazo* (Vol. 19: *Issues in Hispanic and Lusophone Linguistics*, pp. 101–145). Amsterdam: John Benjamins Publishing Company.

Field, H. (2001). Attributions of Meaning and Content. In H. Field (Ed.), *Truth and the Absence of Fact* (pp. 157–174). Oxford: Oxford University Press.

Field, H. (2017). Egocentric Content. *Nôus*, *51*(3), 1–26.

Fine, K., Jago, M. (2019). Logic for Exact Entailment. *The Review of Symbolic Logic*, *12*(3), 536–556.

García-Carpintero, M. (2004). The Deferred Ostension Theory of Quotation. *Noûs*, *38*(4), 674–692.

García-Carpintero, M. (2017). Reference and Reference-Fixing in Pure Quotation. In P. Saka, M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation* (pp. 169–194). Dordrecht: Springer.

García-Carpintero, M. (2018). Pure Quotation Is Demonstrative Reference. *The Journal of Philosophy*, *115*(7), 361–381.

Garver, N. (1965). Varieties of Use and Mention. *Philosophy and Phenomenological Research*, *26*(2), 230–238.

Glüer, K., Wikforss, Å. (2015). Meaning Normativism: Against the Simple Argument. *Organon F*, *22*(Supplementary Issue), 63–73.

Gómez-Torrente, M. (2005). Remarks on Impure Quotation. In: P. De Brabanter, (Ed.), *Hybrid Quotations* (*Belgian Journal of Linguistics* 2003 Yearbook, Vol. 17, pp. 129–151). Amsterdam: John Benjamins.

Gómez-Torrente, M. (2017). Semantics vs. Pragmatics in Impure Quotation. In P. Saka, M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation* (pp. 135–167). Dordrecht: Springer.

González-Vilbazo, K. (2005). *Die Syntax des Code-Switching. Esplugisch: Sprachwechsel an der Deutschen Schule Barcelona.* (Unpublished Ph.D. thesis). University of Cologne, Cologne, Germany.

Harman, G. (1999). Immanent and Transcendent Approaches to Meaning and Mind. In G. Harman (Ed.), *Reasoning, Meaning, and Mind* (pp. 262–276). Oxford: Oxford University Press.

Hoot, B. (2011). Complementizer Asymmetry in Spanish/English Code-Switching [Presentation at *International Symposium on Bilingualism 8*]. Oslo, Norway.

Johnson, M. (2018). Pure Quotation and Natural Naming. *The Journal of Philosophy*, *115*(10), 550–566.

Johnson, W. E. (1921). *Logic. Part 1*. Cambridge: Cambridge University Press.

Kaplan, D. (1969). Quantifying In. *Synthese*, *19*(1/2), 178–214.

Kaplan, D. (1989) Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In J. Almog, J. Perry, H. Wettstein H (Eds.), *Themes From Kaplan* (pp. 481–563). New York: Oxford University Press.

Kaplan, D. (2004). The Meaning of Ouch and Oops. Retrieved from: http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf

Kneale, W., Kneale, M. (1962). *The Development of Logic.* Oxford: Clarendon Press.

Kripke, S. (1982). *Wittgenstein on Rules and Private Language. An Elementary Exposition.* Harvard: Harvard University Press.

Langford, C. H. (1937). Review: Bertrand Russell, *The Limits of Empiricism*. *Journal of Symbolic Logic*, *2*(1), 61.

Lewy, C. (1946). Truth and Significance. *Analysis*, *8*(2), 24–27.

Maier, E. (2017). The Pragmatics of Attraction. Explaining Unquotation in Direct and Free Indirect Discourse. In P. Saka, M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation* (pp. 259–302). Dordrecht: Springer.

Mates, B. (1950). Synonymity. In L. Linsky (Ed.), *Semantics and the Philosophy of Language* (pp. 136–167). Urbana: The University of Illinois Press.

McCullagh, M. (2017). Scare-Quoting and Incorporation. In P. Saka, M. Johnson (Eds.), *The Semantics and Pragmatics of Quotation* (pp. 3–34). Dordrecht: Springer.

Moore, G. E. (1966). *Lectures on Philosophy*. London: Routledge.

Neale, S. (2018). Means Means Means [Talk given at the University of London, 26th Nov. 2018]. Retrieved from: https://backdoorbroadcasting.net/2018/11/stephen-neale-means-means-means/

Pagin, P., Westerståhl, D. (2010). Pure Quotation and General Compositionality. *Linguistics and Philosophy*, *33*(5), 381–415.

Perry, J. (2001). *Reference and Reflexivity*. Stanford: Center for the Study of Language and Information.

Predelli, S. (2013). *Meaning Without Truth*. Oxford: Oxford University Press.

Putnam, H. (1953). Synonymity, and the Analysis of Belief Sentences. *Analysis*, *14*(5), 114–122.

Quine, W. V. O. (1960). *Words and Object*. Cambridge (MA): The MIT Press.

Recanati, F. (2001). Open Quotation. *Mind*, *110*(439), 637–687.

Richard, M. (1983). Direct Reference and Ascriptions of Belief. *Journal of Philosophical Logic*, *12*(4), 425–452.

Richard, M. (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge: Cambridge University Press.

Richard, M. (1997). Attitudes in Context. *Linguistics and Philosophy*, *16*(2), 123–148.

Saka, P. (1998). Quotation and the Use-Mention Distinction. *Mind*, *107*(425), 113–135.

Saka, P. (2006). The Demonstrative and Identity Theories of Quotation. *The Journal of Philosophy*, *103*(9), 452–471.

Salmon, N. (1986). *Frege's Puzzle*. Cambridge: The MIT Press.

Salmon, N. (2001). The Very Possibility of a Language. In C. A. Anderson, M. Zeleny (Eds.), *Logic, Meaning and Computation: Essays in Memory of Alonzo Church* (pp. 573–595). Boston: Kluver.

Sande Piñeiro, A. (2018). *C Plus T as a Necessary Condition for Pro-Drop: Evidence From Code-Switching*. (Unpublished Ph.D. thesis). University of Illinois, Chicago, USA.

Schiffer, S. (1987). *Remnants of Meaning*. Cambridge: MIT Press.

Schiffer, S. (2003). *The Things We Mean*. Oxford: Clarendon Press.

Schiffer, N. (2008). Propositional Content. In E. Lepore, B. C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 267–294). Oxford: Clarendon Press.

Sellars, W. S. (1956). Empiricism and the Philosophy of Mind. *Minnesota Studies in the Philosophy of Science*, *1*, 253–329.

Sellars, W. S. (1963). Abstract Entities. *Review of Metaphysics*, *16*(4), 627–671.

Sellars, W. S. (1974). Meaning as Functional Classification (A Perspective on the Relation Between Syntax and Semantics). *Synthese*, *27*(3–4), 417–37.

Strawson, P. F. (1949). Truth. *Analysis*, *9*(6), 83–97.

Tarski, A. (1983). The Concept of Truth in Formalized Languages. In A. Tarski, *Logic, Semantics, Metamathematics* (pp. 152–278). Indianapolis: Hackett.

Washington, C. (1992). The Identity Theory of Quotation. *The Journal of Philosophy*, *115*(10), 582–605.

Whiting, D. (2013). What Is the Normativity of Meaning? *Inquiry*, *56*(2), 219–238.

STUDIA SEMIOTYCZNE (SEMIOTIC STUDIES), t. XXXIV, nr 2 (2020), s. 105–127
ISSN 0137-6608, e-ISSN 2544-073X
DOI: 10.26333/sts.xxxiv2.07
Article

TOMOO UEDA [*]

KANTIAN PRAGMATISM AND THE HABERMASIAN ANTI-DEFLATIONIST ACCOUNT OF TRUTH

SUMMARY: In this paper, I aim to characterize the pragmatist and anti-deflationist notions of truth. I take Habermas's rather recent discussion (1999) and present the interpretation that his notion of truth relies on the reliabilist conception of knowledge rather than the internalist conception that defines knowledge as a justified true belief. Then, I show that my interpretation is consistent with Habermas's project of weak naturalism. Finally, I draw some more general implications about the pragmatist notion of truth.

KEYWORDS: pragmatism, truth, deflationism, consensus theory of truth, reliabilism, weak naturalism, Habermas.

1. Introduction

The focus of this paper is on the anti-deflationist notion of truth and its role in what Habermas calls "Kantian pragmatics".[1] Specifically, I shall discuss Habermas's version of Kantian pragmatics; Habermas (1999, Einleitung) characterizes Kantian pragmatism in the following way:

[*] Hosei University, Faculty of Law. E-mail: ueda.tomoo@hosei.ac.jp. ORCID: 0000-0001-8493-5636.

[1] This work is supported by JSPS KAKENHI JP20K12781. An earlier draft of some parts of this paper was presented at the 2019 PhiLang Conference (Łódź, Poland). I would like to thank the organizers and anonymous reviewers of the conference as well as the audience who provided good discussions in and after the session. I would also like to thank two anonymous reviewers of the journal as well as Lena Kästner, Shuhei Kimoto, Shigeki Noya and Kichiro Oishi for helpful comments on earlier versions of this paper.

> Kantian pragmatism […] relies on the transcendental fact that subjects capable of speech and action, who can be affected by reasons, can learn—and in the long run even "cannot not learn [*nicht nicht lernen können*]". And they learn just as much in the moral-cognitive dimension of interacting with one another as they do in the cognitive dimension of interacting with the world. By the same token, the transcendental formulation of the issue expresses the postmetaphysical awareness that even the best results of these fallible learning processes remain, in a significant sense, our insights. Even true assertions can realize only those ways of knowing that our sociocultural forms of life make available to us. (Habermas, 2003, pp. 8–9; see Habermas, 1999, p. 16)[2]

This citation needs three clarifications: first, one may well wonder how "Kantian" Habermas's position actually is; certainly, historians of ideas have been interested in whether and to what degree Habermas's position qualifies as Kantian. For example, Bernstein (2018, p. 194) claims, "Habermas's Kantianism is far removed from the "historical" Kant, but he appropriates what he takes to be the core insight of Kant's transcendental project".

Second, putting these interpretative issues aside, if we inquire solely after the theoretical content of the "transcendental formulation", one may also wonder what he intends by predicating something as "transcendental"; this point is an important one. The question must be asked: what is the "transcendental fact" Habermas is talking about? According to Habermas, it concerns the subject's ability of learning speech and action (or inability thereof).[3] This learning capacity is called a "transcendental fact" because it depends on a reflective capacity of the subject, which constitutes the "background assumptions that for Kant ensured the status of the unavoidable conditions of the possibility of cognition as rational and as atemporal" (Habermas, 2003, p. 9; see 1999, 17). It is this transcendental fact that affects the notion of truth.

Finally, the concept of "postmetaphysical awareness" must be explained. Habermas's conception of "metaphysical philosophy" is very broad, but, according to Baynes's interpretation, it involves the view that

> there is a form of inquiry and knowledge proper to philosophy that is, on the one hand, quite distinct from that found in the natural and social sciences and, on the other, one that can nonetheless yield a special and authoritative insight into questions concerning both the meaning of life and how the world, in the broadest sense, "hangs together". (Baynes, 2018, p. 72)

[2] Passages from *Wahrheit and Rechtfertigung* (Habermas, 1999), except from Chapters 2 and 5, are translated by Barbara Fultner (Habermas, 2003). I shall quote her translation unless otherwise indicated and refer to both page numbers.

[3] Oishi (in personal conversations and email correspondence) pointed out that Habermas stresses the "reflexivity" of natural languages (Habermas, 1971, p. 122). This suggestion is of importance since the theme of this paper might have a much broader scope and concern Habermas's whole philosophical project. However interesting this is, I cannot go into this larger topic in this paper.

It is this task that Habermas resists ascribing to philosophy. In Habermas's postmetaphysical thinking (1988), the task of philosophy rather consists in "its persistent tenacity in posing questions universalistically, and its procedure of rationally reconstructing the intuitive pretheoretical knowledge of competently speaking, acting, and judging subjects—yet in such a way that Platonic anamnesis sheds its nondiscursive character" (Habermas, 1992, p. 38; see 1988, p. 46).

To sum up: Habermas is pursuing a sort of pragmatist project that seeks to identify the universal and unavoidable (*unhintergebar*) conditions for communicative rationality.

1.1. Varieties of Validity Claims and Tuth

It is well known that Habermas's consensus theory (2009a; 1981, esp. Chap. 3) applies not only to the truth (*Wahrheit*) of factual statements, but also to the rightness (*Richtigkeit*) of normative statements, and to the truthfulness (*Wahrhaftigkeit*) of statements about subjective experience.[4] According to Habermas, each class of statements raises a distinct validity claim (namely, that of truth, rightness or truthfulness). And each must be justified in a discourse, a special sort of dialogue, in which the validity claim is directly questioned and its justification is required. Although the focus of this paper is restricted to the notion of truth, we should bear in mind that Habermas's discussion of truth is applicable to the other validity claims.[5]

In this paper, we shall pick the notion of truth because Habermas (1999) explicitly discusses this validity claim and its relationship to his Kantian pragmatism.[6] A descriptive statement is true if and only if the statement's validity claim of truth is justified in a discourse. Theories of truth that satisfy this formulation belong to the "consensus theory of truth".

1.2. The Traditional Notion of Truth

Since the central focus of this paper is the notion of truth, let me begin by characterizing the traditional notion of truth for the comparison with Habermas's notion.

The traditional notion of truth, which Habermas calls the "semantic truth concept", consists of three assumptions:

[4] Initially (Habermas, 2009a), there were four sorts of validity claims, but since *Theorie der kommunikativen Handelns* (1981), Habermas only names these three validity claims. So, in this paper, I shall only discuss these three sorts of validity claims.

[5] We shall come back to this point later.

[6] It is also worth noting that Habermas keeps this basic structure in his later works (1999) that we are going to analyze, although there are substantial differences between his early works (such as in 1981; 2009a) and later ones. See Section 4.1.1. below.

> (i) Truth is a property; some beliefs and statements exemplify it and some don't. (ii) It's a substantive property, in that we can reasonably expect an account of what truth is, of its underlying nature. And (iii) this account should provide explanations of various important things about truth: including, why the methods appropriate for its detection are what they are, and why we are well-advised to pursue it—that is, to strive for true belief. (Horwich, 2010, p. 13, original emphasis)

There are many important arguments against these three, but what Habermas focuses on is the paradox of self-reference. For example, imagine a card that says "What is written on the reverse of this card is false" on one side, and "What is written on the reverse of this card is true" on the other. If we assume the semantic concept of truth, this card leads to an obvious paradox. This paradox is caused by the fact that the semantic concept of truth does not differentiate between object languages and metalanguages.

This counterargument leads to the conclusion that there is no such abstract property as truth. This thesis is commonly assumed by the positions I shall examine below.

1.3. The Pragmatist Notion of Truth

This paper mainly discusses the pragmatist notion of truth, which states that if there is a role that the notion of truth plays, it should play a pragmatic role (or a role in a discourse). In other words, the question of truth concerns the proper use of truth predicates such as "… is true," rather than characterizing truth as an abstract property.

Let me briefly clarify three issues concerning the pragmatist notion of truth: first, there is a debate about whether the notion of truth plays any role in pragmatism at all (Misak, 2013, p. 66; Okochi, 2017); however, in the course of this paper, it should become clear that neither Habermas nor I agree to the deflationist claim. Second, this label of the "pragmatist notion of truth" is often associated with the notion of utility. For example, the slogans "The truth is what works" (Brandom, 1994, p. 285) or "What we find it helpful in practice to believe" (Horwich, 2010, p. 3) illustrate the utilitarian character of the pragmatist notion of truth. Although it is popular to ascribe such definitions to pragmatism, I shall not consider the problem of utility here.

Thus understood, the final and most important point is that the notion of truth is characterized in terms of justification. Therefore, the following discussions will be about the relationship between the universal and atemporal notion of truth, and justifications that are given in a particular time and place. According to Habermas, the notion of truth "transcends justification although it is always already operatively effective in the realm of action" (2000, p. 49; original emphasis; see 1999, p. 264).

1.4. Outline

I am going to examine the following questions: first, what is the relationship between knowledge and truth according to the pragmatist notion of truth? Second, is truth deflationist?

In the first part of this paper, I shall argue against the deflationist account of truth. First, I shall characterize the relationship between truth and justification (Section 2), and analyze the deflationist account of truth, especially disquotationalism (Section 3.1). Then, I shall formulate Habermas's criticisms of disquotationalism (Section 3.2). Finally, I shall sketch my answer to the first question and then argue that it is faithful to Habermas's position (Section 4).

2. Justification and Truth

According to a pragmatist stance towards content, the truth predicate "… is true" applies to the content of an utterance (which is called a "statement [*Aussage*]") rather than of a sentence. Furthermore, pragmatists analyze the usage of truth predicates in terms of the notion of truth.

In this section, I shall first sketch what the pragmatist notion of truth looks like (Section 2.1). Then, in order to locate Habermas's Kantian pragmatism in this context, I shall lay out his argument against the correspondence theory of truth (Section 2.2).

2.1. The Pragmatist Notion of Truth

From the pragmatist stance, it immediately follows that the classical correspondence theory of truth, which is a version of the semantic concept of truth, is inappropriate for the analysis. It claims that a content (or a statement) is true if and only if it corresponds to reality. However, this theoretical dependence on reality presupposes certain metaphysical assumptions that pragmatists are profoundly against. For, the pragmatists are "best viewed as pursuing the speech-act and justification projects. Pragmatic accounts of truth have often focused on how the concept of truth is used and what speakers are doing when describing statements as true" (Capps, 2019, Sec. 4).

So, the question for pragmatists is: when is the truth predicate used properly? The answer is that it is used properly in our first-order practice of justification: "[All pragmatist perspectives] claim that we should not add anything metaphysical to the first-order research. We must distill the concept of truth out of our practices of research, reason-giving, and consideration" (Misak, 2013, p. 66; my translation).[7]

The distinction between first-order and second-order justifications corresponds, for example, to the justifications for "Elephants have long noses" on one

[7] The original English version has yet to be published.

hand, and "'Elephants have long noses' is true" on the other. Using a truth predicate is pragmatically justified only if there is a proper first-order justification for the relevant statement (i.e., in the above example, for elephants having long noses).[8]

The notion of first-order justification is directly related to the content of the relevant utterance, such as "Elephants have long noses," rather than the content of "'Elephants have long noses' is true". The relevant first-order justification, of course, includes scientific justifications about elephants and their noses, but it is not restricted to these. Testimonies, for example, in television documentaries or encyclopedias, provide such justifications too.

The relationship between truth and justification is not straightforward, since the notion of truth is universal and atemporal while justifications are made at a specific time and place. In other words, first-order justifications are necessary for using the truth predicate. However, they are not sufficient to define truth because the notion of truth goes beyond each individual justification, as Habermas (2000, p. 49; see also 1999, p. 264) puts it, truth "transcends justification". Actual justifications can, in principle, turn out to be wrong, even if they are made in a very controlled setting. This fallibilist assumption, that the best justification can turn out to be wrong, is important for pragmatists.

2.2. Realism After the Linguistic Turn

Habermas's critique of the correspondence theory stems from his critique of conceptual realism. According to conceptual realism, the objective world is conceptually articulated independently of our minds. Thus, the knowledge about the objective world assumed in conceptual realism is a sort of knowledge that is specific to philosophy rather than the natural sciences.

One apparent problem with conceptual realism is the ontological status of abstract entities such as propositions or the property of truth; in other words, conceptual realism is committed to the view that such abstract entities belong to the objective world. However, for Habermas, this is not acceptable:

> If the "world" that is presupposed according to formal pragmatics is all that is the case—"the totality of facts, not of things" [(Wittgenstein, 1922, 1.1)]—then abstract entities such as propositional contents or propositions must also be taken to be "something in the world". (Habermas, 2003, pp. 30–31; my insertion; see also 1999, p. 41)

Habermas argues against this kind of ontological commitment on many occasions (i.e., Habermas, 1999, Einleitung).

[8] Because of this characteristic, the pragmatic notion of truth is a type of epistemic notion of truth (see, for example, Wrenn, 2015, Chap. 4; esp. Sec. 4.4). It is worth noting that the deficiencies that Wrenn (2015, p. 64) ascribes to Peirce and James will be resolved in the Habermasian consensus theory of truth.

A similar, rather historical argument against conceptual realism might be constructed from Habermas's (1988, Chap. 2) critique of "metaphysical thinking". Metaphysical thinking consists of the idea that the world is already structured in such a way that "theoretical reason will rediscover itself in the rationally structured world, or that nature and history are given a rational structure by reason itself" (Habermas, 1992, p. 34; my emphasis; see 1988, p. 42). While Habermas (1988, p. 36) ascribes this position to a wide range of philosophers, one typical view he associates with conceptual realism is the philosophy of subjectivity or *Bewusstseinsphilosophie*. Both the metaphysical thinking found in *Bewusstseinsphilosophie* and conceptual realism share a certain important feature: an ignorance of intersubjectivity. Inattention to the role of intersubjectivity in a theory of truth is unacceptable for Habermas because he (1999, sec. 5.2) accepts the linguistic turn (of which he relies on Rorty's characterizations). Once this is accepted, intersubjectivity becomes an integral part of the pragmatist theory of truth. A true statement is a statement about the objective world, but, as Habermas says,

> [t]he objective world is no longer something to be reflected but is simply the common reference point for a process of communication [*Verständigung*] between members of a communication community who come to an understanding with one another with regard to something. (Habermas, 2000, p. 35; see also 1999, p. 237)

However, it is essential to distinguish between conceptual realism and the kind of realism Habermas commits himself to:

> Because acting subjects have to cope with "the" [objective] world, they cannot avoid being realists in the context of their lifeworld. Moreover, they are allowed to be realists because their language games and practices, so long as they function in a way that is proof against disappointment, "prove their truth" [*sich bewahren*] in being carried on. (Habermas, 2000, p. 48; my insertion; see also 1999, p. 262)

This kind of realism is not conceptual realism because the objective world is not articulated conceptually; rather, it is merely constituted by objects and events that are not conceptual.

For Habermas, after the linguistic turn, the notion of truth became associated with his Kantian pragmatism, in which linguistic interaction and communication (*Verständigung*) constitute the intersubjective conditions of experiencing the objective world.

Despite the transition from the transcendental (or reflective) subjectivity of consciousness to the detranscendentalized (or prereflective) intersubjectivity of the life-world (cf. Habermas, 1981, p. 451), the problem of conceptual realism remains. As Habermas writes,

> Only the realist presupposition of an intersubjectively accessible objective world can reconcile the epistemic priority of the linguistically articulated horizon of

> the lifeworld, which we cannot transcend, with the ontological priority of a language-independent reality, which imposes constraints on our practices. (Habermas, 2003, p. 30; original emphasis; see also 1999, p. 41)

Thus, Habermas introduces a dichotomy between the life-world and the objective world. The discursive participants (or us) are, as language users, always already (*immer schon*) in the life-world. The life-world is articulated linguistically or conceptually while the objective world is not (otherwise, it would be conceptual realism).

As far as subjective experience is concerned, Habermas (2003, p. 30; see also 1999, p. 41) commits himself to the thesis that "[a]ll experience is linguistically saturated such that no grasp of reality is possible that is not filtered through language". The necessary conditions for objective knowledge through experience are the intersubjective conditions for linguistic interpretation and communication (*Verständigung*).

Thus construed, truth, for Habermas, involves two principles: first, truth is related to referents in the objective world, but not to the facts that belong to the life-world. Second, intersubjective communication is a necessary condition for the possibility of objective knowledge. From these conditions, Habermas characterizes the notion of truth in terms of (intersubjective) justification; hence, the Habermasian notion of truth is a pragmatist notion of truth.

3. Deflationism

Using truth predicates is justified only if the speaker has a first-order justification for claiming the relevant content. Let's call positions on truth that follow this central idea "pragmatist". Beyond agreeing on this central idea, there are disagreements among pragmatists about whether uses of the truth predicate play a distinctive role (see Misak, 2013, p. 66).

One such pragmatist understanding of truth is defended by deflationists. Deflationist theories "are characterized by a cluster of four interlocking ideas about the truth predicate" (Horwich, 2010, p. 14): the truth predicate (1) has a special kind of utility, (2) is non-predicative and non-explanatory, (3) is not naturalistic and (4) is not significant.

(1) The truth predicate has a special kind of utility although theorists' views vary over which kind of utility has priority (e.g., the truth predicate can be used as a device for emphasis, concession, generalization or anaphora; see Horwich, 2010, p.14).

(2) According to deflationism, the meaning of the truth predicate is not empirical. That is, there is no empirical predicate F for any content x such that x is true if and only if x is F. Nor is the truth predicate an abbreviation of a complex non-empirical expression such that it will explain the use of the truth predicate.

(3) Because the truth predicate does not have any empirical characteristics, there cannot be any naturalistic reduction. Furthermore, no natural laws can relate truth with empirical experience (Horwich, 2010, p. 15).

(4) Truth is not “a deep concept and should not be given a pivotal role in philosophical theorizing” (Horwich, 2010, p. 16; original emphasis).[9] That is, the notion of truth cannot be the basis for our conception of meaning.

The central conclusion that deflationists draw from the above discussion is two-fold. First, since claiming something to be true is nothing more than claiming it, “truth is not a concept that has an important explanatory role to play in philosophy” (Brandom, 2009, p. 158; original emphasis; see also Horwich, 2010, pp. 14–16).

Second, since the notion of truth does not play any explanatory role in philosophical theories, there is no such property as truth. That is, unlike other predicates, the truth predicate (“… is true”) does not designate any real property. If this is the case, the logical form of the truth predicate must be different from first-order predicates.

3.1. Disquotationalism

One kind of deflationist theory of truth is disquotationalism (DQ), which is based on the following disquotational scheme:

DQ: For every *S*, “*S*” is true if and only if *S*.

DQ presupposes a distinction between an object language and a metalanguage; that is, while the occurrence of *S* on the left-hand side of DQ is simply mentioned, that on the right-hand side of DQ is being used. Hence, the paradox of self-reference does not occur in this formulation of truth. This makes the disquotationalist theory of truth more attractive than the correspondence theory.

Disquotationalists respect the common pragmatist assumption that the justification for the assertion “‘Snow is white’ is true” is the same as the first-order justification for the assertion “Snow is white”. From this, they infer that claiming something to be true is nothing but claiming it. So, there is no pragmatic difference between these two claims.[10] Hence, disquotationalists conclude that there is no distinctive nature of truth to be investigated. In this sense, the disquotationalist theory of truth is deflationist. In the following section, I argue that this theory is inappropriate.

[9] Rorty (2000b, p. 56) also made a similar point in his reply to Habermas (2000).

[10] Interestingly, Frege (1983, p. 211 [“Einleitung in die Logik”]) also makes a similar point.

3.2. Habermas's Argument Against Disquotationalism

In his discussion of Rorty's paper (1994),[11] Habermas criticizes the disquotationalist position.[12, 13] Habermas's criticism is twofold:

(1) Habermas criticizes the uninformativeness of the disquotational function "because it already presupposes the representational function [*Darstellungsfunktion*]" (Habermas, 2000, p. 43; my insertion; see also 1999, p. 252). In order to understand the expression "… is true," you should understand the right-hand side of DQ: For every *S*, "*S*" is true if and only if *S*. The right-hand side of DQ is used in the metalanguage. This requires that "[b]efore an assertion can be quoted it must be 'put forward'" (Habermas, 2000, p. 43; see also 1999, p. 252). The content must be asserted before it can be quoted in the use of "… is true". There is a problem here because of fallibilism, according to which actual justifications can, in principle, turn out to be wrong, even if made in a very controlled setting. Fallibilism implies that the right-hand side of DQ could turn out to be false even if it were asserted with the best possible argumentation.

The difference between deflationists and Habermas consists in the question of whether fallibilism is applied to a second-order claim, that is, a claim of the form "'S' is true". While deflationists answer the question affirmatively, Habermas should answer it in a negative way. Thus, for Habermas, there must be something more to the notion of truth than just an assertability of the quoted sentence; that is, there is an important distinction between a statement's being justified and its being known.

In this respect, it is helpful to understand Habermas's critique of Rorty's deflationism. According to Rorty (2000, p. 4), "it is no more necessary to have a philosophical theory about the nature of truth, or the meaning of the word 'true', than it is to have one about the nature of danger, or the meaning of the word 'danger'". Habermas argues against this position by referring to the regulative role of truth.[14] The notion of truth is said to be regulative if truth works as

[11] This paper is a shortened version of Rorty's (2000a; for the explanation, see p. 25, fn. 1).

[12] Habermas (1999, Sec. 5.5) includes this position in the "semantic conception of truth". However, as I argued in Section 3.1, the disquotationalist position is actually a sort of pragmatist but deflationist position.

[13] Misak (2007, Sec. 3) discusses another version of deflationism (the so-called prosententialist position) that is introduced by Grover, et al. (1975) and defended by Brandom (1994, Chap. 5; see also 2009b). It would be very interesting to explore whether Habermas's critique of disquotationalism applies to prosententialism too; however, this requires another paper because he does not explicitly discuss prosententialism.

[14] Wellmer (2007a) criticizes Putnam, Habermas and Apel, and argues that truth is not regulative. His assumption is that, in order to discern justifications and truth, the consensus theorists of truth must introduce the notion of "the last consensus". However, it should be noted that the position of Habermas that Wellmer aims to criticize is his former position (esp. in 1981; 2009a), and Wellmer's criticisms of truth as a regulative idea do not apply to Habermas's current position (1999).

the "reference point" in such a way that it gives "the fallibilist consciousness that we can err even in the case of well-justified belief" (Habermas, 2000, p. 48; see also 1999, p. 262). This reference point stems from the life-worldly distinction between believing and knowing that "relies on the supposition, anchored in the communicative use of language, of a single objective world" (Habermas, 2000, p. 48; see also 1999, p. 262).

This criticism implies that, for Habermas, the notion of truth "transcends justification although it is always already operatively effective in the realm of action" (2000, p. 49; original emphasis; see also 1999, p. 264). From this, it follows that truth claims, namely, claims of the form "'*S*' is true," are second-order claims and they are not fallibilistic.

(2) Habermas's second criticism of disquotationalism concerns the use of "… is true". Of course, truth plays an essential role in scientific knowledge, which requires absolute justifications, but truth is also used in everyday situations. And even if the scientific uses could be explained in terms of disquotationalism, Habermas claims, "a semantic [i.e., disquotational] conception of truth simply does not help us at all" because of the contextual expressions in pretheoretical uses of truth predicates (Habermas, 2000, 43; my insertion; see also 1999, p. 253). If disquotationalism were right, there would have to be an instance of DQ for everyday uses of the truth predicate; however, it is hardly clear how to formulate any instance of DQ in a way disquotationalists would accept. Take "'I am in good health now' is true" as an example. If DQ is applied to this example, it is true if and only if I am in good health. However, if there are plenty of cases where the utterer of the embedded statement, "I am in good health," differs from the truth claim, then DQ does not work. Hence, an integral part of the analysis of the use of truth predicates is an examination of the discursive situation shared by the speaker and the hearer.

3.3. Proper Uses of Truth Predicates

According to Habermas, claiming "*S*" is (pragmatically) distinct from claiming "'*S*' is true". The question is, then: what is the difference? The key to answering this is in his notion of validity claims. In the following, I shall briefly intro-

Independent of applicability, Wellmer's claim (2007a, Sec. 9) that a regulative idea is not necessary for truth is wrong. He argues that the trans-subjective rational consensus is the telos of argumentation, but that it falls short of truth because "rationality" just means "justified" (*begründet*). However, Habermas (1999, Chap. 5) requires reached consensus to function as the shared foundation for future intersubjective actions. Here, the notion of truth is understood performatively and regulates the rational actions of both interlocutors because "[a]n assertion that has been disposed of argumentatively in this way and returned to the realm of action takes its place in an intersubjectively shared lifeworld from within whose horizon we, the actors, refer to something in a single objective world" (Habermas, 2003, p. 47; see also 1999, p. 261).

See also Apel's (2011) for his critical discussion of (Wellmer, 2007a).

duce the Habermasian notion of validity claims and then sketch the pragmatic role of the truth predicate using the notion of validity claims.

3.3.1 Validity claims.

According to Habermas (esp. in 2009a), when a person makes a claim, for example, "Susan is clever," in addition to making the claim, the speaker also claims that the content of their claim is true, and claiming it is normatively right, and he claims it sincerely. The speaker asks the listener to agree to these claims. These three sorts of claims are called "validity claims" (*Geltungsansprüche*), and they play an essential role in Habermas's consensus theory. Namely, what Habermas calls "discourse" in which justifications for the validity claims are given only if the listener questions (one of) the validity claims and requires justification for the relevant validity claim(s).

Let's focus on the validity claim of truth. If you claim "Susan is clever" and your interlocutor raises a question as to whether the claim is true, a discourse kicks in and you have to give justification for the claim. In such a discourse, both you and your listener will only focus on justifications and take hypothetical attitudes towards the truth of the claim. If you and your listener reach a consensus that the validity claim of truth has been justified in the discourse, the truth of the claim is established.[15]

3.3.2. Pragmatic roles of the truth predicate.

Although truth claims are indeed about providing justification, they only deliver one specific sort of justification about one specific validity claim; however, claiming something (e.g., "Susan is clever") without any specifications might be subject to the responsibility to provide a variety of justifications. In other words, "'Snow is white' is true" should be rewritten as "In terms of a truth-related justification,[16] snow is white". Therefore, the use of truth predicate is not deflationist.

This is clearly in the spirit of Habermas when he says, "All of these utterances imply validity claims" (2009a, p. 229, my translation), and every discourse has one of these validity claims about which justifications are given (see Section

[15] As we shall see later (in Section 4), Habermas changed his opinion concerning the conditions in which the relevant consensus will be reached. In his original position (2009a), this consensus can only be reached in the ideal speech situation, while in the version that we are analyzing (1999), the consensus has to be reached in an actual discursive situation.

[16] By a "truth-related justification," I mean a justification about the objective world as the "system of possible referents—as a totality of objects" (Habermas, 2003, p. 27; see also 1999, p. 37). Note that the objective world consists of objects and events rather than facts. It is not clear from his writing whether properties will count as a subclass of referents, but I assume that should be so. In contrast, facts clearly do not constitute the objective world (Habermas, 1999, p. 42).

1.1). A truth claim only focuses on the validity claim of truth in the first instance.[17]

4. The Pragmatic Role of Truth: A Provisional Externalist Interpretation[18]

In the previous sections, we examined Habermas's truth theory from pragmatist perspectives. In this section, I shall present a provisional but positive account of truth claims and their role, which I argue is a natural extension of what Habermas (1999, Chap. 5) defends. Essentially, Habermas's revised position (1999, Chap. 5) is a consensus theory of truth in which the notion of truth is characterized in terms of a consensus reached in an actual discourse rather than a consensus in an ideal speech situation. So, the issue for this revised consensus theory of truth is how to discern truth from justification; for truth is universal in character while justifications are bound to a specific time and place.

This provisional account of truth must meet two conditions: first, it must be anti-deflationist; second, it must characterize the conditions in which truth claims are justified (Section 4.1). I shall further argue (in Section 4.2) that my provisional account is Habermasian in spirit.

4.1. Anti-Deflationism and Justification for Truth Claims

Anti-deflationism claims that the concept of truth cannot be explained away; rather, it plays an essential role in our daily (linguistic) practice. In other words, the justification for asserting "'S' is true" is not the same as the justification for asserting "S".

What, then, does the justification for a truth claim look like?

To answer this, I shall discuss the consensus theory of truth, which is characterized in terms of the life-world, which, in turn, depends on the reliabilist conception of knowledge (Section 4.1.1). This will make the notion of justification for a truth claim naturalistic (Section 4.1.2). The naturalistic notion of truth is not deflationist (Section 4.1.3).

4.1.1. Reliable life-worlds.

In a previous paper of mine (Ueda, 2019), I argued that Habermas's theory of truth relies on the reliabilist conception of knowledge (Goldman, 1979).[19]

[17] Of course, one can easily imagine a context in which the truth claim itself will be criticized from the normative point of view. However, as Habermas (2009a) argues, this kind of criticism constitutes a metadiscourse.

[18] This interpretation is provisional, partly because it does not yet cover other discursive properties, such as the rightness (*Richtigkeit*) of normative assertions (see Section 5.2).

[19] Brandom (1994, Chap. 4) also relies on the reliabilist notion of knowledge. However, for Brandom, the reliabilist notion of knowledge is the source of entitlement to form perceptual beliefs.

To recap, Habermas (1999, Chap. 5) defends a new consensus theory of truth, according to which truth is defined in terms of the actually reached consensus between interlocutors in a concrete discursive situation rather than in an ideal one (as defended in Habermas's earlier paper [2009a]). In order for Habermas's theory of truth to accommodate the fallibilist nature of well-justified statements, he points out that the discursive participants rely on life-worlds. The notion of a life-world, which he inherits from the work of Schutz and Luckmann (1973), is characterized as a class of background beliefs that are shared among the participants of a discourse. Constituents of a life-world, which are background information shared between the participants of a discourse, stay implicit between the interlocutors, but can be made explicit merely by asking an explicit question about them and asking for justification. This is exactly the same sort of question about validity claims that a speaker makes when she assertively utters some statement. And we (as the active participants of any discourse) rely on the background and implicit beliefs contained in the life-world shared between us.[20]

Our reliance should, then, be based on the fact that the life-world as a whole counts as knowledge without being explicitly justified; rather, life-worldly beliefs remain implicit. That is to say, one needs to make a belief shared among interlocutors in the life-world explicit if it is to be justified. That is exactly the function of a discourse.

In each discourse, we can rely on the life-world as a whole because life-worldly implicit beliefs are formed through reliable processes from the objective world or natural world (Schutz, Luckmann, 1973, Sec. 1.A). Namely, "The everyday life-world is the region of reality in which man can engage himself and which he can change while he operates in it by means of his animate organism" (Schutz, Luckmann, 1973, p. 3), and the totality of life-worldly beliefs is given "as a certain reliable ground of every situationally determined explication" (Schutz, Luckmann, 1973, p. 9).

To characterize implicit life-worldly knowledge, it is a natural step to attribute to Habermas an externalist notion of knowledge, namely, the reliabilist one (Goldman, 1979). To be more precise, the justification condition of the reliabilist knowledge concept is relevant. For, according to the reliabilist definition of knowledge, the justification condition of the internalist conception of knowledge, which is defined as a justified true belief (JTB), will be substituted by the external and reliable process of forming a belief. This is exactly the type of justification process that we take for granted to form life-worldly knowledge. So, as a natural extension of Habermas's theory of truth, the consensus theory should be based on the reliabilist conceptions of justification rather than the internalist notion used in JTB.

[20] The first-person plural nature of the discourse is important. Habermas (2000) criticizes Brandom (1994) on the grounds that the game of asking for and giving reasons, or discursive score-taking, does not capture the first-person plural nature of the discourse. Interestingly, Brandom (1994) acknowledges this criticism, saying he is "not really doing justice to the specific role of the second person" (2000, p. 362).

I shall argue for two theses: first, that my account of truth is consistent with Habermas's text (1999, Chap. 5). Second, that it answers Wellmer's criticism (2007b) of Habermas's (1999).

(1) My ought claims here are consistent with Habermas's new defense of the consensus theory of truth (1999, Chap. 5). This new defense represents a significant revision of his former position on consensus theory of truth (2009a), which is based on the classical internalist notion of knowledge as JTB.

The justification condition in the definition of knowledge as JTB is closely related to the idealization of discursive situations. As discussed above, the idealization of a speech situation is needed in order to distinguish truth from justification, the latter of which always occurs in a specific time and place. That is to say, if we characterize the notion of truth in terms of knowledge as JTB, we need idealized justifications; or the consensus that is reached between the speaker and the listener in a discourse remains fallible (i.e., it does not reach the truth that transcends justifications) regardless of how rigorous the arguments given by the speaker are.

However, the idealization of discursive situations has been heavily criticized; in particular, Wellmer (1993; 2004) argues that beliefs that are only justified in idealized discursive situations cannot count as knowledge that human beings can have. Habermas (1999, Chap. 5) accepts Wellmer's points.

The reliabilist construal of implicit life-worldly knowledge is consistent with Habermas's fallibilist position. In the reliabilist understanding, a statement "*S*" is justified reliably if S occurs in a (e.g., statistically) reliable manner. The famous example of "barn-façade" shows this type of justification is fallible. Hence, it makes sense to talk about a statement that is reliably justified, but false, nonetheless.

(2) My interpretation of the consensus theory of truth seeks to answer Wellmer's criticism (2007b) of the revised consensus theory of truth. His criticism is directed against the objective world as a "[w]orld objectified in a natural-scientific way [*naturwissenschaftlich objectierte Welt*]" (Wellmer, 2007b, p. 211; my translation) and is stated in the form of a dilemma:

> Either the objective world refers to the domain of scientific objectifiables—then, it is useless to stake out the region of statements that are capable of truth and discourse; or this region also contains the historical-cultural reality—then, the word "objective" is nothing much more than just a word centrifuge against contextualism. (Wellmer, 2007b, pp. 211–212; my translation)

The first horn of this dilemma is especially important. According to Wellmer, the notion of truth is only applicable to the domain that can be examined through natural science. However, Wellmer claims that Habermas acknowledges that historical-cultural concepts such as personhood are constituents of the objective world because we can talk about persons; if so, the objective world characterized by Habermas is not sufficient for defining truth.

I think Wellmer confuses Habermas's uses of "truth" (*Wahrheit*) and "falsehood" (*Falschheit*) with their everyday uses; while we talk about something normative (or belonging to the "social world") as "*wahr*" or "*falsch*," Habermas evaluates these kinds of uses on the basis of the pair of "rightness" (*Richtigkeit*) and "wrongness" (*Falschheit*) in his theory. As I mentioned above, Habermas tries to capture the everyday uses of "*wahr*" and "*falsch*" in addition to the uses of these terms in the natural sciences. However, capturing the everyday uses does not mean that he should cover all the everyday uses under one single sort of discourse in which one single sort of justification should be given and, accordingly, one single pair of evaluative concepts should be applied;[21] indeed, he should not. There is a distinction between factual and normative statements even in everyday discursive situations; and we give a different sort of justification in each of these discursive situations.

To sum up, the internalist conception of knowledge as JTB leads to the idealization of discursive situations, however, I have argued that the idealization of discursive situations ought to be abandoned. It is, therefore, plausible to interpret Habermas's more recent defense of his theory of truth as depending on an externalist conception of knowledge.

So far, I have argued that Habermas (1999, Chap. 5) revised his position significantly from his former one and that the revision consists in the use of an externalist and reliabilist notion of knowledge.[22] This is exactly the same version of the consensus theory of truth as the one I have put forward (Ueda, 2019), according to which a statement *p* is true only if

1) the discursive participants actually reach an agreement on the justification of the validity claim about *p* in the discourse,
2) the agreement makes some of the life-worldly and implicit background assumptions explicit, and
3) the agreement, as a whole, captures the objective world in a reliable manner.

4.1.2. Weak naturalism about truth.

The externalist conception of knowledge has a certain affinity with naturalism, a version of which Habermas indeed commits himself to (1999, Einleitung).[23] In the following sections, I shall first characterize Habermas's distinction

[21] This point should not be confused with another essential point concerning statements and their evaluation, namely, that everyday statements often raise multiple validity claims at the same time (see the example in Section 3.3.1 and also, Habermas, 2009a). The point here is that there are several different sorts of discourse in accordance with which there must be distinctions among evaluative terms, and Habermas is quite right in distinguishing them.

[22] One significant consequence of this revised consensus theory of truth is that the externalist justification relations are holistic. Habermas (1999, Chap. 5) is aware of this consequence, and I think it is a positive feature (cf. Ueda, 2019).

[23] According to Misak's understanding of pragmatism, "All the pragmatist viewpoints on truth are naturalistic viewpoints, too" (2013, p. 19; my translation). However, the

between strong and weak naturalism. Then, I shall argue that my proposal is consistent with weak naturalism.[24]

(1) Habermas (1999, p. 37) draws a distinction between strong and weak naturalism.[25] Naturalism is strong if it is reductionist; in Habermas's words:

> All cognition is ultimately to be reducible to empirical processes. The transcendental architectonic drops out, as does the difference between the conditions of how the world is constituted (or of world disclosure), which call for conceptual analysis, on the one hand, and states of affairs and events in the world, which can be explained causally, on the other. (Habermas, 2003, p. 23; see also 1999, pp. 32–33)

In contrast, weak naturalism "makes no reductionistic claims" (*Ansprüche*; Habermas, 2003, p. 27; see also 1999, p. 38); again, as Habermas states:

> [W]eak naturalism contents itself with the basic background assumption that the biological endowment and the cultural way of life of *Homo sapiens* have a "natural" origin and can in principle be explained in terms of evolutionary theory. (Habermas, 2003, pp. 27–28; see also 1999, p. 38)

There are two important points for the current discussion: first, the learning process plays a central role in weak naturalism. Habermas assumes

> that "our" learning processes, that are possible within the framework of sociocultural forms of life, are in a sense simply the continuation of prior "evolutionary learning processes" that in turn gave rise to our forms of life. (Habermas, 2003, p. 27; see also 1999, p. 37)

This is the kind of learning process that was necessary for Kantian pragmatism (see the quote in the introduction).

Second, Habermas's view is not a reductionist position on discursive argumentation, which is characteristic of the consensus theory of truth. Every time you make a statement, for example, "It is snowy today," you raise validity claims. If your interlocutor does not agree with you, she can ask for the explicit justification for your validity claim of truth. In such a case, you have to justify your validity claim by making the implicit background (or life-worldly) assumptions explicit and, thereby, explaining how the objective world is constituted.

Hence, one can argue that the explicit discursive argumentation relevant for the notion of truth plays a distinct role that refutes the reductionist project.

---

question remains open as to whether Habermas's weak naturalism is consistent with Misak's interpretation.

[24] Note that I shall defend a rather weak thesis here. I shall not go so far as to argue that the consensus theory of truth should always be interpreted as weakly naturalist.

[25] The discussion of the relationship between naturalism and religion will be a central theme of Habermas's later works (see Habermas, 2005). However, this paper does not cover the topic of religion.

(2) As I have stressed in the introduction to this paper, it is central to Habermas's Kantian pragmatics to evaluate how the subject's ability to learn (or inability thereto) affects the notion of truth.

Since the relevant ability to learn is characterized in terms of an evolutionary process and the learning process is the process of acquiring the relevant knowledge, the notion of knowledge used here must be possible to naturalize. The reliabilist notion of knowledge aims to externalize the justification condition. The reliabilist way of naturalizing the justification condition does not have to be reductionist; for, instead of reducing the explicit justification, reliabilism requires that the belief-forming process is reliable and it plays a role as the implicit justification that can be made explicit by way of explicit justification. Of course, according to reliabilism, explicit justifications are not necessary for defining knowledge.[26] However, justifications play a different role in intersubjective discourse; namely, they make the implicit background beliefs (which are the constituents of the life-world) explicit.

I claim that the reliabilist theory of knowledge is consistent with the transcendental distinction between the rational realm of justifications (the life-world) and the causal realm of the objective world. Taking the example of perception, Goldman characterizes the formation of justified perceptual beliefs as follows: "(6a) If *S*'s belief in *p* at *t* results ("immediately") from a belief-independent process that is (unconditionally) reliable, then *S*'s belief in *p* at *t* is justified" (Goldman, 1979, p. 13; original numeration). Thus, Goldman states the connection between two sets of relationships: the causal and cognitive relationship between the objective world and the subject on the one hand, and the intersubjective justification relations in the life-world on the other.

My proposal that the pragmatist notion of truth is characterized in terms of the externalist notion of truth clearly respects these points.

4.1.3. Anti-deflationism about truth.

So far, I have established two theses that are consistent with Habermas (1999, Chap. 5). First, the Kantian-pragmatic notion of truth should be characterized in terms of the externalist notion of knowledge, especially the reliabilist one (regardless of what Habermas's own position might be). Second, the notion of knowledge in question is weakly naturalistic (which is consistent with Habermas's text, especially in 1999). From these theses, I shall conclude that the consensus theory of truth is (1) pragmatist and (2) not deflationist.

(1) The consensus theory of truth follows the core principles of pragmatism and hence is a pragmatist notion of truth. First, the consensus theory of truth applies to the content of an utterance (or statement) and is defined in terms of discursive justification and the consensus reached through justification. That is, truth is dependent on the speaker's speech act. Second, according to the consen-

[26] Of course, justified true belief is not sufficient for the definition (Gettier, 1963).

sus theory of truth, consensus is reached only if explicit justifications for the validity claim in question are provided. From these two points, it follows that truth plays a pragmatic role in discourse.

(2) The weakly naturalistic nature of (Kantian) pragmatism is clearly inconsistent with the classical deflationist stance because deflationism is committed to anti-naturalism; as discussed in Section 3, deflationism claims that truth cannot be substituted with any empirical property or any class of such properties.

A more positive point can be made for the anti-deflationist stance. Deflationism asserts that a truth claim, such as "'It snows in Tokyo' is true," does not claim anything more than "It snows in Tokyo". However, I argued (with Habermas) in Section 3.3.2 that truth claims play a distinct role in discursive situations (without committing to the existence of the abstract property of truth) in two ways. First, truth claims make life-worldly beliefs explicit and illustrate a speaker's readiness to justify the relevant validity claims. Second, the relevant justification is explicitly about truth rather than other sorts of validity claims.

4.2. Habermasian in Spirit

In Section 4.1, I advanced two theses: first, I argued that the consensus theory of truth relies on the externalist and reliabilist notion of knowledge. Second, I argued that the notion of truth is not deflationist because truth claims play a pragmatic role in discourse.

From these theses, it is almost straightforward to see that my provisional interpretation of the consensus theory of truth is consistent with Habermas's text (1999). It is especially important that the consensus theory of truth depends on the reliabilist notion of knowledge. This dependence is necessary not only for the pragmatist and anti-deflationist notion of truth, but it is also consistent with the transcendental nature of such a notion.

Of course, the consistency between my theory and Habermas's text does not necessarily mean that Habermas actually takes my theory as his own theory; therefore, I must say more. However, since I have already motivated the reliabilist theory of knowledge independently above, my theory must be seen as a plausible candidate for Kantian pragmatism.

5. Conclusion

This paper has argued for three theses: first, the externalist notion of knowledge (rather than the internalist one) is necessary for the Habermasian notion of truth. Second, Kantian pragmatism is an anti-deflationist theory of truth. Finally, I defended my version of Kantian pragmatism and showed that it is consistent with Habermas's weak naturalism. A full-fledged defense of Kantian pragmatism, of course, requires a more detailed examination of Habermas's position, which remains to be conducted.

In this last section, I would like to provide some broader perspectives on pragmatism and its relationship to the notion of truth (Section 5.1) and briefly suggest some future lines of inquiry (Section 5.2).

5.1. The Role of Truth in Pragmatism

As mentioned above, there is a debate about whether the notion of truth plays any role in pragmatism at all (Misak, 2013; Okochi, 2017). However, the present paper clearly indicates that the notion of truth is not dispensable in pragmatism (regardless of whether it is Kantian or not). This is exactly the interpretation of Peirce that Misak (2007, Sec. 4) defends.[27] According to Misak, Peirce "was very explicitly not interested in a reductive analysis of truth. And he was not focused on the ideas of total evidence, epistemically ideal conditions, and the solving of all questions" (Misak, 2007, p. 82).

In an important respect, Misak's interpretation of Peirce is consistent with my interpretation of Habermas (1999). According to her,

> Peirce never went anywhere near trying to spell out what epistemically ideal conditions might be, and he never went anywhere near that idea that an inquirer would know that she was in epistemically ideal conditions. In fact, his fallibilism explicitly has it that a person could never know that inquiry had been pursued as far as it could fruitfully go. (Misak, 2007, p. 83)

As I have shown in this paper, this interpretation is perfectly consistent with (my interpretation of) Habermas (1999). He does not commit himself to the ideal discursive situation as a necessary condition for defining truth anymore. He is explicit about not spelling out "what epistemically ideal conditions might be".

There are, of course, some inconsistencies between Habermas's views and Peirce's. For example, according to Misak (2007, p. 83), Peirce "thinks that truth is a property of beliefs" while Habermas thinks that truth is not a property of statements and the truth-predicate is applicable to statements rather than the semantic content of an utterance. Making the difference between them explicit requires another entire paper at least. Nonetheless, the interpretation advanced in this paper is consistent with Misak's interpretation in the most important respects, and if Misak is successful in defending the indispensability of truth in pragmatism, I think Habermas's Kantian pragmatism certainly counts as a pragmatist project.

5.2. Future Perspectives

I have proposed a provisional account of truth that is consistent with Kantian pragmatism, and I have also pointed out some theoretical claims to which Ha-

[27] The primary aim of this section is to examine Misak's evaluation of Wright's view (1992). However, I shall not get into the discussion laid out by Wright (1992).

bermas should be committed. However, so far, my reading is merely consistent with Habermas's views, and it remains to be shown whether Habermas is actually committed to the theoretical claims I have outlined. The latter task requires a far more detailed analysis of Habermas's works.

Another important issue that has remained undiscussed in this paper is the rightness (*Richtigkeit*) of normative statements, which Habermas must be able to analyze in a fully parallel way to truth.[28] Rightness constitutes the main domain of his social and moral theories and is the central notion at play in his discourse ethics (Habermas, 1991).

With regard to rightness, there is another important issue that has been left behind in this paper (and in my previous paper as well). This is the issue of the relationship between Habermas's current position (laid out in 1999) and Apel's position. They once worked together, and it is still common practice to treat both of them as defending the same consensus theory of truth. However, their positions cannot be seen as the same anymore (see, for example, Apel, 2011, Sec. 3, for Apel's criticism of Habermas's current theory of truth).[29] Explicitly differentiating their current theories would surely provide a better understanding of Habermas's Kantian pragmatism.

REFERENCES

Apel, K.-O. (2011). Wahrheit als regulative Idee. In K.-O. Apel, *Paradigmen der Ersten Philosophie* (pp. 322–349). Frankfurt a.M.: Suhrkamp.

Baynes, K. (2018). Postmetaphysical Thinking. In H. Brunkhorst, R. Kreide, C. Lafont (Eds.), *The Habermas Handbook* (pp. 71–74). New York: Columbia University Press.

Bernstein, R. J. (2018). Neopragmatism. In H. Brunkhorst, R. Kreide, C. Lafont (Eds.), *The Habermas Handbook* (pp. 188–195). New York: Columbia University Press.

Brandom, R. (1994). *Making It Explicit*. Cambridge, MA: Harvard University Press.

Brandom, R. (2000). Facts, Norms, and Normative Facts: A Reply to Habermas. *European Journal of Philosophy*, *8*(3), 356–374.

Brandom, R. (2009). Why Truth Is Not Important in Philosophy. In R. Brandom, *Reason in Philosophy: Animating Ideas* (pp. 156–176). Cambridge, MA: Belknap Press.

[28] On the recent development of the relationship between truth and rightness, see Habermas 1999, ch. 6.

[29] Kettner (2018, p. 43) analyzes the distinction between Apel and Habermas in terms of their reliance on "two different strains of American pragmatism represented by Charles Sanders Peirce, on the one hand, and John Dewey, on the other". These are certainly important theoretical points that need to be examined in detail.

Capps, J. (2019). The Pragmatic Theory of Truth. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2019 Edition). Retrieved from: https://plato.stanford.edu/archives/sum2019/entries/truth-pragmatic/

Frege, G. (1983). *Nachgelassene Schriften* (2nd ed.). Hamburg: Felix Meiner.

Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121–123.

Goldman, A. (1979). What is Justified Belief? In G. S. Pappas (Ed.), *Justification and Knowledge. Philosophical Studies Series in Philosophy* (vol. 17, pp. 1–23). Dordrecht: Springer.

Grover, D., Camp, J., Belnap, N. (1975). A Prosentential Theory of Truth. *Philosophical Studies*, *27*(2), 73–125.

Habermas, J. (1971). Der Universalitätsanspruch der Hermeneutik. In J. Habermas, D. Heinrich, J. Taubes (Eds.), *Hermeneutik und Ideologiekritik* (pp. 120–159). Frankfurt a.M.: Suhrkamp.

Habermas, J. (1981). *Theorie der kommunikativen Handelns, Band 1*. Frankfurt a.M.: Suhrkamp.

Habermas, J. (1988). *Nachmetaphysisches Denken: Philosophische Aufsätze*. Frankfurt a.M.: Suhrkamp.

Habermas, J. (1991). *Erläuterungen zur Diskursethik*. Frankfurt a.M.: Suhrkamp.

Habermas, J. (1992). *Postmetaphysical Thinking: Philosophical Essays*. Cambridge, MA: MIT Press.

Habermas, J. (1996). Rorty's pragmatische Wende. *Deutsche Zeitschrirft für Philosophie*, *44*(5), 715–741.

Habermas, J. (1999). *Wahrheit und Rechtfertigung*. Frankfurt: Suhrkamp.

Habermas, J. (2000). Richard Rorty's Pragmatic Turn. In R. Brandom (Ed.), *Rorty and His Critics* (pp. 31–55). Oxford: Blackwell.

Habermas, J. (2003). *Truth and Justification*. Cambridge, MA: MIT Press.

Habermas, J. (2005). *Zwischen Naturalismus und Religion.* Frankfurt a.M.: Suhrkamp.

Habermas, J. (2009a). Wahrheitstheorien. In J. Habermas, *Philosophische Texte* (2nd vol., pp. 208–269). Frankfurt a.M.: Suhrkamp.

Habermas, J. (2009b). Rationalitäts- und Sprachtheorie. In J. Habermas, *Philosophische Texte* (2nd vol., pp. 105–145). Frankfurt a.M.: Suhrkamp.

Horwich, P. (2010). *Truth-Meaning-Reality*. Oxford: Clarendon Press.

Kettner, M. (2018). Pragmatism and Ultimate Justification. In H. Brunkhorst, R. Kreide, C. Lafont (Eds.), *The Habermas Handbook* (pp. 43–48). New York: Columbia University Press.

Misak, C. (2007). Pragmatism and Deflationism. In C. Misak (Ed.), *New Pragmatists* (pp. 69–90). Oxford: Oxford University Press.

Habermas, J. (2013). Hndert Jahre Pragmatismus. In M. Hartmann, J. Liptow, M. Willaschek (Eds.), *Die Gegenwart des Pragmatismus* (pp. 62–80). Berlin: Suhrkamp.

Okochi, Y. (2015). Shinri to Kihan [Truth and Norm]. *Gendaishiso*, *43*(11), 208–223.

Rorty, R. (1994). Sind Aussage universelle Geltungsansprüche? *Deutsche Zeitschrift für Philosophie*, *42*(6), 975–988.

Rorty, R. (2000a). Universality and Truth. In R. Brandom (Ed.), *Rorty and His Critics* (pp. 1–30). Oxford: Blackwell.

Rorty, R. (2000b). Response to Jürgen Habermas. In R. Brandom (Ed.), *Rorty and His Critics* (pp. 56–64). Oxford: Blackwell.

Schutz, A., Luckmann, T. (1973). *The Structure of the Life-World*. London: Heinemann.

Ueda, T. (2019). Habermas-niokeru Shinri to Seitoka: Risokanashino Shinrino Goisetsu [Truth and Justification for Habermas: Consensus Theory of Truth without Idealization]. *Tokyo Ikashika Daigaku Kyoyobu Kenkyukiyo*, *49*, 37–49.

Wellmer, A. (1993). Wahrheit, Kontingenz, Moderne. In A. Wellmer, *Endspiele: Die unversönliche Moderne: Essays und Vorträge* (pp. 157–177). Frankfurt a. M.: Suhrkamp.

Wellmer, A. (2004). *Sprachphilosophie: Eine Vorlesung*. Frankfurt a.M.: Suhrkamp.

Wellmer, A. (2007a). Der Streit um die Wahrheit. Pragmatismus ohne regulative Ideen. In A. Wellmer, *Wie Worte Sinn machen* (pp. 180–207). Frankfurt a.M.: Suhrkamp.

Wellmer, A. (2007b). Die Wahrheit über Warheit? Zu Jürgen Habermas, Wahrheit und Rechtfertigung. Philosophische Aufsätze. In A. Wellmer, *Wie Worte Sinn machen* (pp. 208–215). Frankfurt a.M.: Suhrkamp.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. London: Kegan Paul, Trench, Trubner, and Co.

Wrenn, C. (2015). *Truth*. Cambridge: Polity.

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