Special Issue

The Origins of Meaning and the Nature of Speech Acts

Maciej Witek

Guest Editor

Contents

Maciej Witek: *Preface* 270

Research Articles (Special Issue)

Stina Bäckström: *Must Expression Be Instrumental?* 282

Marina Bakalova: *The Epistemic Value of Music* 303

Maciej Witek: *Self-Expression in Speech Acts* 326

Mateusz Włodarczyk: *Limitations of Non-Gricean Approaches to the Evolution of Human Communicative Abilities* 360

Felix Bräuer: *Common Ground, Conversational Roles and Epistemic Injustice* 399

Marcin Lewiński: *Conclusions of Practical Argument: A Speech Act Analysis* 420

Research Articles

Kyoung-Eun Yang: *Do Kuhn's Cases of the Theory-Change from Newtonian to Einsteinian Physics Support His Incommensurability Thesis?* 458

Miloš Kosterec: *Some Further Remarks on Hybrid View of Fictional Characters* 484

Book Review

Jaroslav Peregrin: S*anford Shieh: Necessity Lost: Modality and Logic in Early Analytic Philosophy, Vol. 1* 492

Organon F 28 (2) 2021: 270–281
https://doi.org/10.31577/orgf.2021.28201
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

Preface

For the past two decades, Mitchell S. Green (2003; 2007; 2009; 2019a) has been developing an original model of self-expression, the central idea of which is that expressing, understood as a behavior whereby we make our mental states public, is a species of signaling. In other words, in expressing ourselves we both show and signal our introspectable states. According to Green (2009, 141-143), to show something is to make it knowable to an appropriate observer. A signal, in turn, is defined as a "feature of an entity that conveys information (including misinformation) and that was designed for its ability to convey that information." (Green 2007, 26-27)

Next, to account for our tendency to take some natural objects and artefacts as if they expressed emotions, feelings and moods, Green distinguishes between *expressing* a mental state and *being expressive* of it. Unlike expression, expressiveness is not factive. For instance, my sneer expresses my contemptuous attitude towards what my interlocutor has said in that it both signals and shows my actual mental state. However, a sneer can be expressive of contempt independently of the current feelings and emotions of the sneering agent. According to Green (*Ibid.*, 40) "[t]his evidently means that his face has a configuration that would typically be used by one who is expressing their contempt." Next, Green (*Ibid.*, 178-180) refers to the phenomenon of cross-modal congruence and put forth a hypothesis according to which it plays a key role in the mechanism underlying our practice of attributing expressiveness to natural objects, artefacts, and art works.

The above-mentioned notions of signalling and showing play a key role in Green's (2007) signalling model of communication, within which he develops a three-part account of speaker meaning. Green distinguishes between acts of *factual*, *objectual*, and *illocutionary* speaker meaning and claims that to speaker-mean an item—a fact, an object, or a commitment, respectively—is to signal and overtly show it. For instance, to

factually speaker mean that p is to perform an action with an intention that in performing it, first, one enables knowledge about the fact that p in an appropriately endowed receiver and, second, makes it manifest that one has *this* intention; by analogy, to speaker mean φ'ly that p—where 'φ' stands for a certain illocutionary force—is to perform an action intending that, first, one makes it knowable that one is committed to the proposition that p under force φ and, second, makes it manifest that one has *this* intention. Green's notion of speaker-meaning, which is an alternative to Grice's original definition of non-natural meaning, can be easily integrated within the framework of the signalling model of communication and successfully used in theorising about the evolutionary emergence of uniquely human communicative skills.

In "Speech Acts, the Handicap Principle and the Expression of Psychological States", Green (2009) uses the framework of the evolutionary biology of communication to account for the expressive dimension of speech acts, i.e., their power to express the mental states that are their sincerity condition. He argues that expressive illocutions—i.e., speech acts that allow for insincerity—are handicaps: signals difficult to fake in virtue of being costly to produce. More specifically, Green claims that assertions, requests, promises, etc. are subject to expressive norms. As the corollary of this, a speaker who performs an expressive speech act incurs the risk of a loss of credibility; for instance, a speaker who asserts that p incurs the cost of closing off the option of not having the belief that p without exposure to the risk of being accused of insincerity.

In his more recent paper entitled "Organic Meaning: An Approach to Communication with Minimal Appeal to Minds", Green (2019b) introduces the notion of organic meaning which significantly enhances the explanatory power of the signalling model of communication. He discusses a number of examples of organic meaning and argues that they can be regarded as intermediate forms between mere natural signs and acts of intentional and inferential communication. According to him, communicative transactions involving organic meaning do not require from their participants the ability to form and reason about intentions and other propositional attitudes. Green concludes that the notion of organic meaning enables us to solve the so-called *cognitive load* problem, i.e., it enables us to to fill in the gap between the rich communicative skills of humans and those of our extant evolutionary relatives.

Green has also contributed to the development of the common-ground model of communication. In (Green 2017a) he argues that the common-ground framework, which has been originally devised and used by Robert Stalnaker (1998, 2014) to explain the functioning of assertions, can be extended to account for acts of asking a question. According to Stalnaker (1998, 5-6), the context of a conversation can be defined as a context set: the set of all possible worlds that are compatible with the body of information that the conversing agents mutually take for granted or, in other words, that are compatible with the propositions that constitute their common ground. An assertion, in turn, is understood as a proposal to reduce the context set by eliminating from it the worlds in which the asserted proposition is false. Stalnaker (2014, 141) also suggests that to ask a question is to put forth a proposal to divide the context set into partitions representing alternative answers to it. Like asserted propositions, than, asked questions can be accepted by conversing agents and absorbed into the common ground. Following this suggestion, Green develops a common-ground model of communication that accounts for acts of making an assertion as well as for acts of asking a question: to accept a question as a common ground component—i.e., to take it to be "a question worthy of investigation" (Green 2017a, 1591)—is "a matter of (…) structuring those worlds in a certain way" (*Ibid.*).

Green (2017a; 2019a; 2021) also argues that the common-ground framework can be further elaborated to accommodate what he calls the teleological perspective on conversation. Adopting this perspective, he arrives at a taxonomy of conversation types that are defined by reference to their purposes and the distribution of roles or, in other words, 'illocutionary entitlements' among the conversing agents. Using the first criterion, he distinguishes between inquiries and deliberations, i.e., between conversations aimed at determining how things are and conversations whose purpose is to formulate a course of action. Next, taking into account the conversational roles of the participants in a dialogue—i.e., the relations between their conversational goals and the distribution of their 'illocutionary entitlements' to add propositions or questions to the common ground—he distinguishes between symmetrical, didactic asymmetrical, and socratic asymmetrical conversations. Participants in a symmetrical conversation are allowed to make the same types of conversational moves; "in didactic conversations, [in turn,] one interlocutor aims

to lead others to accept an answer to a practical or theoretical question about which the speaker may already have an opinion or plan, while in socratic conversations, an interlocutor aims to lead others to answers by helping them to formulate their own views or plans of action." (Green 2017a, 1595)

The above-mentioned ideas of self-expression, expressiveness, signalling communication, speech acts as handicaps, organic meaning, and conversation types are systematically discussed in the papers collected in the present volume.

Stina Bäckström in "Must expression be instrumental?" critically examines Green's (2007) model of self-expression. She argues that Green adopts the 'instrumental perspective', the central idea of which is that expression is a means for transmitting information about mental states from one organism to the other. Bäckström contrasts the instrumental perspective with an alternative approach, which she calls the 'descriptive perspective' and articulates with the help of Merleau-Ponty and Wittgenstein. She argues that it is the descriptive view, not the instrumental one, that enables us to account for manifesting emotions, feelings, attitudes, and thoughts that are formed and take shape in novel forms of expression.

Viewed from the instrumental perspective, expression is an information-transferring process that enables knowledge of one's mental states. For instance, Green (2007) takes smiles, frowns, yelps, gestures, and other expressive manifestations of what is within to be signals: behavioral or physiological traits that were designed for their ability to convey the information that they do. According to the descriptive perspective, by contrast, expression is best understood as an essential aspect of the *mode of being* of human agents and other sentient creatures.

Bäckström argues that Green's instrumentalist model makes no room for cases in which sentient and thinking agents express thoughts that are formed in the course of being expressed. One example of such a situation comes from George Eliot's *Middlemarch*: Rosamond Vincy twists her neck and thereby expresses her obstinacy and determination. Discussing this example, Green (2007, 142) finds it difficult to account for Rosamond's twist of the neck as a case of signaling and, by the same token, as a case of expressing. According to Bäckström, however, the twist is a genuine expression even though it cannot be regarded as a signal, i.e., as a sign designed for its ability to convey Rosamond's obstinacy. Bäckström claims that the case under

discussion constitutes a counterexample to Green's theory, according to which to design the twist as an expressive signal Rosamond would have to access her obstinacy independently of the piece of behavior she used to manifest it; the obstinacy, however, takes shape in the twist construed as a novel form of expression.

According to the descriptive perspective, by contrast, embodying or manifesting mental states that take shape in being expressed constitute the most fundamental cases of expression. This is not to say that descriptivist models make no room for instrumental expression. For instance, Merleau-Ponty's theory of expression allows for instrumental cases (e.g., the 'second-order-speech'). Nevertheless—Bäckström argues—Merleau-Ponty took genuinely novel expressions (e.g., the 'first-hand' speech) to be primitive and paradigmatic. In a similar vein, Wittgenstein argued that to have pain is to express it. In sum, being expressed is a characteristic mode of existence of mental states. Expressing is not a matter of manifesting an independently existing mental states; rather, it is a matter of constituting the expressed states.

Marina Bakalova in "The Epistemic Value of Music" argues that music enables us to acquire knowledge of the phenomenal character of real or imaginary inner states that cannot be easily gained otherwise; in other words, due to its unique expressive dimension—which she calls *musical expressiveness*—music has epistemic value. It is instructive to stress that Bakalova adopts Green's (2007) distinction between expressing a state and being expressive of it and claims that what a piece of music is expressive of—that is, what it 'expresses' in the ordinary sense of this word—is not necessarily what its composer or performer intends to express.

Bakalova also argues that thanks to their expressive function, pieces of music can evoke concepts that cannot be expressed verbally, e.g., 'elegance', 'lyricism', 'drama', 'nostalgia', and 'melancholy'. According to Bakalova, a piece of music that expresses certain inner experiences can convey to its listener a concept whose content relates to how these experiences feel like. What is more, it can express the concept in question more fully than it can be done with the help of a verbal phrase that stands for it. As a result, the listener can gain new knowledge which enables her to harbour emotions or other phenomenal states which she has never experienced before.

To account for the above mentioned expressive powers of music, Bakalova adopts Green's (2007) multispace model of artistic expressions and

argues that it allows us to explain how a sound sequence can display phenomenal characteristics. Viewed from the perspective of this model, a piece of music can be represented by a trajectory of expressive stimuli in a multidimensional space involving the time dimension. The piece so represented can be regarded as expressing a mental episode construed as a temporal sequence of phenomenal states that flow one into the other only if the trajectory and the episode occupy roughly the same region of the multidimensional space. In other words, the piece of music can enable qualitative knowledge about the episode—that is, it can show an appropriately endowed listener how this episode feels (see Green 2007: 48 and 2009: 142)—provided there is a sufficient overlap between the regions they occupy.

Maciej Witek in "Self-expression in speech acts" discusses Green's (2007) notion of self-expression and examines the role it plays in his model of illocutionary communication (Green 2009; 2019a). He suggests that Green's three-part model of speaker-meaning can be extended by introducing the notion of proto-illocutionary speaker-meaning, which is necessary to account for cases of overtly showing general commitments that are not 'marked' as being specific to one or another illocutionary force. He also argues that the model of expressive norms presented in (Green 2009) involves a kind of circularity. Green claims, namely, that expressive norms enable us to indicate the force of a speech act—i.e., "how what is said is to be taken and what would count as an appropriate reply" (2009, 160)—by showing the psychological states that constitutes its sincerity condition. At the same time, however, Green asks the following question: "How can the use of an illocutionary force constitute strong enough evidence of a psychological state to enable knowledge in an appropriate observer—that is to express that state? (*Ibid.*, 148). In other words, Witek concludes, Green seems to assume that the use of a force shows a certain psychological state and thereby indicates itself. Finally, Witek elaborates on the idea of discourse-constituted thoughts (Jaszczolt and Witek 2018)—or, in other words, thoughts that exist in virtue of being expressed—and argues that it can be regarded as a useful amendment to Green's model of expressive illocutions. More specifically, he distinguishes between (*i*) expressing with the thinking-to-speaking direction of influence and (*ii*) expressing with the speaking-to-thinking direction of influence or, in other words, between (*i*) expressing discourse-independent thoughts and (*ii*) expressing discourse-

constituted thoughts. According to Witek, the notion of discourse-constituted thoughts—together with the corresponding idea of expressing with the speaking-to-thinking direction of influence—enables us to arrive at a more comprehensive account of self-expression in illocutionary communication.

Mateusz Włodarczyk in "Limitations of non-Gricean approaches to the evolution of human communicative abilities" discusses Green's (2017b; 2019b) conception of organic meaning and Dorit Bar-On's (2013) model of expressive communication. What these two proposals have in common is that they describe forms of non-Gricean communication. Gricean communication is both intentional and inferential. By contrast, organisms participating in communicative transactions involving cases of organic meaning as well as animals producing and reading expressive signals do not have to form audience-directed intentions and reason about mental states of others.

According to Green and Bar-On, the models they offer are not only adequate descriptions of non-Gricean forms of communication to be found among humans and non-human animals, but also provide an adequate basis for explaining the phylogenetic and ontogenetic emergence of uniquely human communicative skills. More specifically, Green and Bar-On take cases of organic meaning and expressive signals, respectively, to be intermediate stages between cases of natural meaning and acts of speaker meaning. In other words, they use their models of non-Gricean communication to solve the *cognitive load* problem. According to Włodarczyk, the solutions they propose are non-Gricean in that they use their models of non-intentional and non-inferential communication to develop a plausible explanation of the evolutionary emergence of ostensive-intentional communication; by contrast, Gricean accounts of the evolution of human communicative abilities posit the class of 'attenuated' or 'minimally Gricean' (Moore 2016) acts and argue that they constitute an intermediate form between cases of natural meaning and acts of fully-fledged Gricean communication.

Włodarczyk claims that the non-Gricean evolutionary accounts based on the notions of organic meaning and expressive communication fail to provide a sufficient basis for explaining the evolutionary emergence of uniquely human communicative skills. To justify his view, he refers to Niko Tinbergen's (1963) 'four questions' model of ethological explanation and argues that Green and Bar-On, who focus on the adaptive function of non-intentional forms of

communication, ignore questions about their underlying mechanisms, ontogeny and phylogeny. Next, Włodarczyk argues that examples of organic meaning examined by Green, rather than constituting intermediate forms between natural and non-natural meaning, are best understood as special cases of natural meaning. Finally, he discusses communicative gestures and verbal acts whereby one agent directs the attention of the other towards an external entity or event. According to Tomasello (2010), such triadic interactions are hallmarks of human communication. It turns out, however, that most examples of expressive signals and organic meaning discussed by Bar-On and Green, respectively, are dyadic rather than triadic. For this reason—Włodarczyk concludes—non-Gricean models developed by Green and Bar-On fails to offer a plausible reconstruction of the evolutionary transition from dyadic to triadic forms communication.

Felix Bräuer in "Common Ground, Conversational Roles and Epistemic Injustice" takes up the suggestion made by Green in "Conversation and Common Ground" to the effect that his extended common-ground framework can shed light on the phenomenon of conversational injustice. It is instructive to note, however, that Bräuer uses Miranda Fricker's (2007) term 'epistemic injustice', which stands for situations in which a speaker is unfairly discriminated against in his or her capacity as a knower based on prejudices about him or her; by analogy, we can speak of conversational injustice when a speaker is discriminated in that his or her capacity as a conversational agent is unjustly curtailed.

Bräuer distinguishes between three varieties of epistemic or conversational injustice: testimonial, inquiring and interpretative injustice. What they have in common is that their victims are, due to some negative identity stereotypes against them, unfairly curtailed in their ability to add to the common ground in the way intended by them. A speaker who suffers testimonial injustice is unjustly prevented from updating the common ground with the proposition she asserts; even though her utterance is taken to be an assertion, the proposition she expresses is not accepted due to a negative stereotype against her. By analogy, a speaker who suffers inquiring injustice is unfairly prevented from adding the question she asks to the common ground; as a result, her attempt to shape the course of the inquiry is unfairly thwarted: her question, though recognized and understood, is not accepted as something worthy of investigation. What Bräuer calls interpretative

injustice, in turn, consists in unjustly misunderstanding the force and meaning of the speaker's utterance or in misinterpreting the attitude expressed by her.

Bräuer discusses three examples of epistemic or conversational injustice. Using elements of Green's model of conversation types, he also argues that epistemic injustice plays a key role in the subordinating mechanisms that impairs the conversational agency of its victims: a speaker who suffers testimonial, inquiring, or interpretative injustice is curtailed in his or her ability to play a symmetrical role in the language game he or she participates in. Epistemic injustice, then, inflicts conversational harms on its victim and—Bräuer argues—has dehumanizing effects in that it compromises the speaker's conversational capacities which are crucial to his or her identity as a human being.

Marcin Lewiński in "Conclusions of Practical Argument: A Speech Act Analysis" uses a speech-act theoretic framework (Fogal et al. 2018; Green 2009; 2018; 2020) to account for a variety of illocutionary acts that conclude practical arguments. It is commonly agreed that conclusions of theoretical reasoning are expressed by speech acts belonging to what Green (2009, 160) calls the 'assertive family': assertions, conjectures, suggestions, educated guesses, suppositions, and the like. What they have in common is that they have the words-to-world direction of fit and involve commitment to a propositional content; however, "[t]hey differ from one another in the norms by which they are governed, and thereby in the nature of that commitment." (Green 2009, 157) According to Lewiński, a distinctive feature of practical arguments is that their conclusions are expressed by what he calls *action-inducing speech acts*: illocutions with the world-to-words direction of fit that involve commitment to or at least putting forth a future action. Depending on the agent of the action induced, Lewiński distinguishes between three types of action-inducing illocutions: *commissives*, *directives*, and their *hybrids* such as proposals and offers. A commissive takes effect by committing the speaker or a group to which she belongs to a future action, whereas the performance of a directive counts as an attempt to get or even oblige the addressee or a third party to a certain future action; a hybrid speech act—i.e., the illocution that combines commissive and directive elements—takes effect as an attempt to get or even oblige both the speaker and her addressee to the joint performance of a future collective action. Moreover, Lewiński argues that action-inducing acts differ

also with respects to their illocutionary strength; in particular, the strength with which a given speech act presents its illocutionary point—i.e., the inducing of a future action—can vary from weak through neutral to strong. As a result, Lewiński arrives at a three-by-three matrix that enables him to distinguish between nine types of speech acts that can be used to express conclusions of practical argumentation. To demonstrate the explanatory power of his taxonomy of action-inducing speech acts, he uses it to account for arguments and illocutions to be found in the Guardian's campaign to disinvest fossil fuels.

It is instructive to stress that in developing his model Lewiński adopts an argumentative perspective on reasoning (Mercier and Sperber 2011), the central idea of which is that the structure of a reasoning construed as a cognitive process is constituted in the course and for the sake of argumentative practice. In particular, Lewiński claims that "[p]ractical discourse (practical argumentation, deliberative practices) is (…) not only a display mechanism for inner practical reasoning but also an important entry point into the elements and standards of practical reasoning" (Lewiński 2021, 434). For this reason, Lewiński's externalist model of practical argumentation, together with his taxonomy of action-inducing illocutions, can be regarded as an substantial contribution to the discussion on the expressive dimension of communicative practice (Green 2007 and 2009).

I would like to thank the authors for contributing to this thematic volume and allowing their research programs to become part of it. I would like to express my gratitude to Mitchell S. Green for the stimulating discussions we had on the role of self-expression in speech acts and other topics of common philosophical interest. Last but not least, I would like to thank Martin Vacek and the Editorial Board of *Organon F* for making this project possible and offering space in the journal.

Maciej Witek

References

Bakalova, Marina. 2021. "The Epistemic Value of Music." *Organon F* 28 (2): 303–25. https://doi.org/10.31577/orgf.2021.28203

Bäckström, Stina. 2021. "Must Expression Be Instrumental?" *Organon F* 28 (2): 282–302. https://doi.org/10.31577/orgf.2021.28202

Bar-on, Dorit. 2013. "Origins of Meaning: Must We 'Go Gricean'?" *Mind & Language* 28 (3): 342–75. https://doi.org/10.1111/mila.12021

Bräuer, Felix. 2021. "Common Ground, Conversational Roles and Epistemic Injustice." *Organon F* 28 (2): 399–419. https://doi.org/10.31577/orgf.2021.28206

Fogal, Daniel, Daniel W. Harris, and Matt Moss, eds. 2018. *New Work on Speech Acts*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.001.0001

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Green, Mitchell. 2003. "Grice's Frown: On Meaning and Expression." In *Saying, Meaning, Implicating*, edited by Georg Meggle, and Christian Plunze, 200–19. Leipzig: University of Leipzig Press.

Green, Mitchell. 2007. *Self-Expression*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283781.001.0001

Green, Michell S. 2009. "Speech Acts, the Handicap Principle and the Expression of Psychological states." *Mind & Language* 24 (2): 139–63. https://doi.org/10.1111/j.1468-0017.2008.01357.x

Green, Mitchell. 2017a. "Conversation and Common Ground." *Philosophical Studies* 174: 1587–604. https://doi.org/10.1007/s11098-016-0779-z

Green, Mitchell. 2017b. "How Much Mentality is Needed for Meaning?" In *The Routledge Handbook of Philosophy of Animal Minds*, edited by Kristin Andrews and Jacob Beck, 313–23. London and New York: Routledge

Green, Mitchell S. 2018. "A Refinement and Defense of the Force/Content Distinction." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 99–122. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0004

Green, Mitchell S. 2019a. "From Signaling and Expression to Conversation and Fiction." *Grazer Philosophische Studien* 96 (3): 295–315. https://doi.org/10.1163/18756735-09603002

Green, Mitchell. 2019b. "Organic Meaning: An Approach to Communication with Minimal Appeal to Minds." In *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*, edited by Alessandro Capone, Marco Carapezza and Franco Lo Piparo, 211–28. Springer, Cham. https://doi.org/10.1007/978-3-030-00973-1_12

Green, Mitchell S. 2000. "Speech Acts." *Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), edited by Edward N. Zalta. Last updated September 24, 2020. https://plato.stanford.edu/archives/win2020/entries/speech-acts/

Green, Mitchell S. 2021. "Context and Conversation." In *The Wiley Blackwell Companion to Semantics*, *First Edition,* edited by Daniel Gutzmann, 1–23. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118788516.sem045

Jaszczolt, Kasia M., and Maciej Witek. 2018. "Expressing the Self. From Types of *de se* to Speech-act Types." In: *Expressing the Self. Cultural Diversity & Cognitive Universals*, edited by Minyao Huang & Kasia M. Jaszczolt, 187–222. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198786658.003.0010

Lewiński, Marcin, 2021. "Conclusions of Practical Argument: A Speech Act Analysis". *Organon F* 28 (2): 420–57. https://doi.org/10.31577/orgf.2021.28207

Mercier, Hugo, and Dan Sperber. 2011. "Why do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and Brain Sciences* 34 (2): 57–74. https://doi.org/10.1017/S0140525X10000968

Moore, Richard. 2016. "Gricean Communication and Cognitive Development." *The Philosophical Quarterly* 67 (267): 303–26. https://doi.org/10.1093/pq/pqw049

Stalnaker, Robert. 1998. "On the Representation of Context." *Journal of Logic, Language, and Information* 7: 3–19. https://doi.org/10.1023/A:1008254815298

Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199645169.001.0001

Tinbergen, Niko. 1963. "On Aims and Methods of Ethology." *Zeitschrift für Tierpsychologie* 20 (4): 410–33. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tomasello, Michael. 2010. *Origins of Human Communication*. Cambridge, Mass.: Bradford Books/MIT Press.

Witek, Maciej. 2021. "Self-Expression in Speech Acts." *Organon F* 28 (2): 326-59. https://doi.org/10.31577/orgf.2021.28204

Włodarczyk, Mateusz. 2021. "Limitations of Non-Gricean Approaches to the Evolution of Human Communicative Abilities." *Organon F* 28 (2): 360–98. https://doi.org/10.31577/orgf.2021.28205

Organon F 28 (2) 2021: 282–302
https://doi.org/10.31577/orgf.2021.28202

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Must Expression Be Instrumental?

Stina Bäckström*

Received: 16 April 2019 / Revised: 4 November 2019 / Accepted: 2 December 2019

Abstract: This article engages critically with the theory of expression proposed by Mitchell Green in his Self-Expression (2007). In this book, Green argues that expressions are signals designed to convey information about mental states. By putting pressure on one of the examples Green uses in his book, I will challenge this thesis. I will then deepen this challenge by developing a contrast between two philosophical perspectives on expression, which I name the 'instrumental' and the 'descriptive'. I take Green's theory of expression to be an exemplar of the instrumental perspective. Expression, in the instrumental perspective, is a means for transmitting information about mental states from organism to organism. The descriptive perspective I articulate with the help of Maurice Merleau-Ponty and Ludwig Wittgenstein. On the descriptive view, expression is (at least a part of) an answer to the question what it is so much as to have mental states and a living body. I suggest at the end of the article that if we remain within the instrumental perspective, we will not be able to use expression to satisfactorily answer this question.

Keywords: Expression; Green; Merleau-Ponty; signal; Wittgenstein.

* Södertörn University
https://orcid.org/0000-0002-1326-4844
Centre for Practical Knowledge, Södertörn University Alfred Nobels allé 7 Flemingsberg 141 89 Huddinge, Sweden
stina.backstrom@sh.se

0. Introduction

> Our view of man will remain superficial so long as we do not return to this origin, so long as we do not rediscover the primordial silence beneath the noise of words, and as long as we do not describe the noise that breaks the silence. Speech is a gesture, and its signification is a world. (Merleau-Ponty, 2012 [1945], 190)

When Mitchell S. Green's book *Self-Expression* was published in 2007, I was setting out to write a dissertation on expression. At that stage of research, and that stage of one's career, a new book length treatment of one's subject—especially when such treatments are rare—is rather anxiety provoking. I knew I had to respond to the book, and I was afraid it would say everything I had wanted to say. This did not prove to be true, not the least because I did not, at that time, know what I wanted to say. What turned out to be the case was that Green's book shaped what I wanted to say. It gave me something—a theoretical vocabulary, a set of examples—to which to respond.

It has taken me a while (I am an untimely slow reader) to articulate what I found problematic about the perspective on expression the book offers. I have responded to specific parts of the book elsewhere, to moments and particular theses (see Bäckström 2013a; 2013b; 2016), but I haven't been able to clearly position it, or myself with respect to it. This paper is an attempt to do so. I know Green's thinking has moved on in various interesting directions since then, but I hope he has not moved so far as to make my intervention here obsolete.

My primary aim in this article is to articulate a challenge to Green's thesis that an expression is a signal. I will do so by developing a contrast between two different perspectives on expression.[1] The first I call the 'instrumental perspective'. Green's book is an exemplar of this perspective. What characterizes the instrumental perspective is that its guiding idea is

[1] In calling them 'perspectives' rather than 'theories' I wish to indicate that the differences between them are not only or even primarily in the resulting account of expression but in the guiding questions and background assumptions. I want to say something about the place where they, respectively, start looking (as it were) for expression.

that expression is a means for transmitting information about mental states from one organism to another. The instrumental perspective takes as its primary question *how* expression can carry information about mental states to other people.

The second perspective I articulate with the help of Maurice Merleau-Ponty's treatment of expression in *Phenomenology of Perception*, and a set of remarks from Ludwig Wittgenstein's *Philosophical Investigations*. I call it the 'descriptive perspective'. The descriptive perspective takes as its guiding idea that we need to understand expression in order to understand what it is to so much as have mental states and a living body.[2] It understands expression to be essential to understanding the *mode of being* of human beings and other living creatures. As we will see, I take the different perspectives to differ not only in its guiding question (in how they approach expression) but also in substance (in what they understand expression to be).[3]

It might appear odd that I'm grouping Merleau-Ponty and Wittgenstein together under a common heading. Merleau-Ponty explicitly writes himself into the phenomenological tradition, whereas Wittgenstein doesn't. And there are important differences in their respective method and style. Perhaps there are also deep philosophical disagreements between them related to those differences in method and style. I think it is striking, however, when making the attempt at establishing a conversation between the two thinkers on the notion of expression (at least when staying with a comparison between *The Phenomenology of Perception* and *The Philosophical Investigations*) how they complement each other.[4]

---

[2] I use the term 'mental states' mostly for convenience in this paper, in order to establish a conversation between the parties involved. It is not a term I am comfortable with, since it pulls us towards thinking in terms of a contrast between the mental and static one the one hand and the bodily and dynamic on the other. But it will do for the purposes of the dialectic in this paper.

[3] The contrast between the perspectives will be rough in outline. In my experience, there is value in philosophy in alternating between zooming in and zooming out. The risk when zooming out, is of course that one rides roughshod over distinctions and loses sight of specificity.

[4] The connection between Wittgenstein and various figures (most notably Heidegger) in the phenomenological tradition has been explored for a long time, but recently there appears to be a surge of interest specifically in comparing Merleau-

It might appear odd, too, with a contrast between an instrumental perspective and a descriptive perspective. The descriptive is more commonly contrasted with the explanatory. But the contrast I am after here is not happily put in terms of a focus on *what there is* on the one hand and *why it is* on the other hand. Both Green and the proponents of the descriptive perspective are interested in what expression is. I use the notion of description here to indicate an interest both in the *what* and the *is* of "What is expression?" An interest in the *is* can be translated in terms of the question, "What is it for an expression to *be*?" This question, according to the descriptive perspective, cannot be asked without at the same time asking the question "What is it for mental states to *be*?"[5]

Green's book is a philosophical theory of expression with a particular focus on understanding the continuity between non-linguistic expression and speech-acts. Green approaches expression from "evolutionary biology as informed by game-theory" (Green 2007, 16). Given this starting point, it might appear as if the two perspectives I am contrasting are simply too far apart for the comparison to be meaningful. Wittgenstein and Merleau-Ponty, each in different ways, think philosophical clarification needs to achieve some distance to the conceptual frameworks of scientific theories.

I think, however, that Green's theory of expression is an excellent articulation of a philosophical perspective on expression that has considerable

Ponty and Wittgenstein. For example, Avner Baz uses Wittgenstein and Merleau-Ponty alongside each other on the topic of perception and motivation in the article "Motivational Indeterminacy" (2017). And two anthologies have been published recently devoted partially or entirely to exploring the affinities between the thinkers: Wittgenstein and Merleau-Ponty (2017) and Wittgenstein and Phenomenology (2018).

[5] I have struggled with the terminology and I am not completely happy with the current one. I have considered re-naming the descriptive perspective the 'phenomenological perspective'. This would have had the advantage of avoiding making it seem as if the instrumental perspective has no descriptive commitments or interests. And I think the question whether it is illuminating to describe Wittgenstein as a phenomenologist is a good one. But what I am after in this article is how Merleau-Ponty and Wittgenstein both think we need to re-think certain fundamental philosophical questions if we are to understand expression. To me, there is nothing distinctively phenomenological about this thought.

appeal, and this appeal is not dependent on a particular investment in the conceptual framework of evolutionary biology. I think (and Merleau-Ponty and Wittgenstein also thought) that it is a natural enough idea that expression must fundamentally be understood as a means for making one's mental life public. Articulating why this natural idea is problematic, perhaps radically so, is—from their point of view and mine—a worthwhile project. Creating a dialogue between this contemporary exemplar and two philosophers in our recent history who took expression to be an eminently important philosophical concept will, my hope is, create an interesting dialectic.

Although my main aim is to articulate the perspectives by contrast with each other, I will also pose a challenge for the instrumental perspective, suggesting that the descriptive perspective shows us that we need a different notion of expression from the one the instrumental perspective offers. I will begin by articulating the instrumental perspective as I find it in Green's book *Self-Expression*, and then articulate the challenge against this perspective. Then I will turn to developing the descriptive perspective. In the final section I will make some clarifications and develop the challenge further by looking at how Green attempts to make space for a constitutive link between expressions and mental states.

1. Expression as signaling: Green's instrumental perspective

The opening sentences of Green's *Self-Expression* reads:

> We express ourselves in many ways: through tone of voice, posture, the face, words, and in more subtle cases, paint, music and other forms of art. Linking these disparate phenomena together is a pattern of behavior coping with a felt need. (Green 2007, 1)

The passage begins with recalling a familiar fact. We, human beings, have a vast range of ways of expressing ourselves, from facial expressions, to novels, songs, and paintings. Having called our attention to this delightful fact, Green immediately suggests a way of initially characterizing all these forms of expression, of bringing them together under a common description. They are all, Green suggests, patterns of behavior coping with a felt need. This need, he elaborates, is the need to "manifest our point of view" (Green 2007, 1).

If the first sentence of the quoted passage is theoretically innocuous, a simple reminder of an undeniable fact, the second sentence already takes up a certain theoretical perspective on this fact. This is a theoretical perspective that Green holds on to and elaborates throughout his book. There we see expression described as a means, an instrument for a specific purpose. We have a felt need, to make our point of view manifest, and we have a way (or ways) of satisfying this need, namely expression. As we shall see, Green's elaborated theory of expression does not assert that expression always has to involve a felt need. Rather, we can see these first sentences as setting up a question: What can it mean to view expression in instrumental terms? How can we elaborate the idea that expression serves the purpose of making our point of view manifest to others?

One very important step for Green in making good on this task is the idea that self-expression is a signal (Green 2007, 26). This thesis appears in a list of twenty theses, which Green calls 'dicta'. Green does not tell us much about the choice to call his theses 'dicta'. According to the dictionary, a dictum is a pronouncement, or a formulation of general principles or truths. Green's twenty dicta function in both ways—they both declare how Green is going to use his terms and what he will take for granted, and assert what he takes to be the basic truths and principles governing the area he will be investigating. This particular dictum: self-expression is a signal, is one that directly responds to the question of what, according to Green, it means to view expression in instrumental terms.

Green gets the notion of a signal from evolutionary theories of communication. A signal is a sign *designed* for the purpose of communication. Green's notion of design is capacious, and this capaciousness is also the key to how he can hold on both to the idea of expression as an instrument, and to the idea that we as individuals do not always and everywhere express ourselves because we feel a need to relay information about our mental states. In Green's theory design spans an evolutionary process, an intentional process at the level of individuals, and an intentional process at the level of a social community (Green 2007, 5, 137-151). Expressions can be naturally selected for their capacity to transmit information. Expressions can also be selected by individuals in an effort to get one's point of view across (or perhaps be selected by a sort of happenstance). And expression

can, finally, be selected by a process of conventionalization at the level of society. If an expression is naturally or conventionally designed, its instrumentality need not be a matter of an individual using it for a specific purpose.

The idea that expression is a means of coping with a felt need thus gets broken down into different more specific notions of instrumentality, with more or less clear connections to what one naturally thinks of as "coping with a felt need". To evince a signal that evolution or society has selected for its capacity to transmit information about mental states need, conceptually or experientially, have nothing to do with feeling a need to make anything manifest. (There is a legitimate worry here that that the three notions of design are so disparate in meaning that Green's theory cannot vindicate our sense that he is articulating a unity that can be captured in one definition. But I will leave that worry to the side.)[6]

What I want to focus on now is the idea that expression is a matter of using or evincing already established means of communication, or else setting up a new connection between a state of mind and a piece of behavior. I will now articulate a challenge for Green's instrumental perspective by discussing the consequences of his theory for our understanding of a particular example, an example I take from Green's book.

Green uses a range of quite lovely examples from literature, in particular great novels of the 19th century. In the specific example on which I want to focus, we find Rosamond Vincy from George Eliot's *Middlemarch.* For texture, I render the entire passage from the novel also quoted by Green:

> This was not an infrequent procedure with Mr Vincy—to be rash in jovial assent, and on becoming subsequently conscious that he had been rash, to employ others in making the offensive retractation. However, Mrs. Vincy, who never willingly opposed her

[6] The title of Green's book is not merely Expression but Self-Expression. This might lead us to think that Green's topic is the romanticist concern with expressing one's authentic self. Green does not, however, focus much on the self. Self-expression for Green has to do with expressing one's own states of mind, which is linked to the "introspectibility" of what can be self-expressed. He distinguishes self-expression from "expressiveness", which can be a property of inanimate objects, such as a when a "windswept cliff might look melancholy" (Green 2007, 38-41).

> husband, lost no time the next morning in letting Rosamond know what he had said. Rosamond, examining some muslin-work, listened in silence, and at the end gave a certain turn of her graceful neck, of which only long experience could teach you that it meant perfect obstinacy. Eliot, 2000 [1874], 214.

In discussing Rosamond's turn of the neck, Green argues on the basis of the theory he proposes that it is not clear whether she is here expressing her obstinacy or not. Green says about Rosamond that her twist of the neck might not be a signal (and hence not an expression) since "it is not clear that she twists her neck for the purpose of showing determination, or for the sake of making as if to act on that determination." One way for it to become a signal, Green continues, is if she begins to "consciously to twist her neck for the purpose of displaying this 'perfect obstinacy.' If she does so, she may also be expressing her determination (Green 2007, 143)." Perhaps there are also less conscious and deliberate ways for Rosamond's twist of the neck to become a signal. If we understand the notion of individual design as capaciously as possible but still compatible with Green's theory, we can expand the space of possibilities here and say that her neck-twist might become a signal also by a sort of non-deliberate happenstance. She simply starts doing to twist her neck regularly when she is feeling obstinate, and it thus comes to be designed to convey her obstinacy.[7]

Green uses this example to clarify certain aspects of his theory. He never considers the thought that it might be a counterexample. That is, he doesn't consider the possibility that the example might describe a form of expression that does not fit his theory. I will now describe and develop the example in a way where it sits uncomfortably with Green's core dictum that self-expression is a signal. On this elaboration of the example, Rosamond Vincy

[7] When Green initially describes what design by individuals amounts to he describes it as the "work of an intelligent agent", and as having to do with "the choice of an intelligent, conscious agent" (2007, 5). This suggested to me that his model of individual design is an individual deciding consciously to use a form of behavior to convey a particular state of mind. In conversation, Green told me that he wants to leave it open that individual design also can be a form of happenstance. One might, without conscious intention, happen upon a form of behavior that then comes by regularity to be (for others) associated with the state of mind in question.

did not wish to display her obstinacy by twisting her neck, nor has previously by a sort of happenstance come to twist her neck regularly when she is feeling obstinate. Neither is, we might suppose, this neck-twist a conventionalized mark of determination for her society or social-group. Nor a naturally selected piece of behavior. In fact, Rosamond's obstinacy was, when she twisted her neck, not quite clear even to herself, but gradually, and in the process of her expressing it in her posture, became clear to her (and to others). She embodied her obstinacy, in a way not determined either by her intention or by previously established relations of information-bearing or representation. Her obstinacy was not only made available to others when she expressed it, but took form as she expressed it. Her obstinacy was not a previously existing phenomenon, it took shape in her novel expression of it.[8]

Green takes the case of Rosamond Vincy to have the potential to show us that there is room for individual variation in his theory of expression. There is space, he wants to show, for those who "express themselves in ways that do not conform to universal or near-universal patterns" (143). What is still the case, however, is that those non-standard (in the sense of non-universal) types of behavior need to be designed to convey what they do, in order to count as expressions on Green's view. This means that Rosamond, on the reading of the example I just gave, cannot be (on Green's theory) expressing her obstinacy in her neck-twist. For how can the neck-twist have been designed to convey Rosamond's obstinacy, if the obstinacy did not exist in any robust sense before it was embodied in the novel expression of a slight turn of the neck?

Say that Green now would respond to this question by saying that he requires something quite minimal for a piece of behavior to count as designed in the relevant sense. Say he argues that all that is required is that the piece of behavior that expresses in a novel way the non-hitherto existing state of mind is that it is intelligible by (some select group of) others as carrying information about that state of mind. Then Green would be able

[8] This reading of the example might appear precluded by the fact Eliot writes that only long experience could teach you that the twist of the neck meant perfect obstinacy. I think that "long experience" could here mean simply long experience with Rosamond, i.e., knowing her well.

to accommodate my elaboration of the example as a genuine case of expression. However, this response would effectively do away with design and explain signaling in terms of a prior notion of the intelligibility of an expression. It would not be possible to cash out the intelligibility of the expression in terms of the piece of behavior being selected to convey that state of mind. The intelligibility of the expression is rather what explains why it is selected.

Hence I see the Rosamond Vincy case as posing a dilemma for Green's theory. Either he denies that the description I gave of the case (where the state of mind took shape in a new form of expression of it) captures a genuine case of expression, or he gives up on the core claim of his theory, that expression is signaling. We might wonder what would be so pernicious about denying that a state of mind can take shape in an original form of expression. As I see it, there are two problems with this.

The first is that it appears, as a matter of everyday experience, to happen all the time. My mixture of confusion and sadness in the face of one of Trump's latest tweets can take the form of a crooked smile and a humming of the tune "My way", through which it materializes that my confusion and sadness has to do with how a certain kind of gonzo self-assertiveness can be a route to actual devastating power. I might not have any such thought before I started smiling and humming, the response was directly elicited by the tweet itself. And by no stretch was the humming and smiling designed in Green's sense to convey the information that I am sad and confused about the devastating power of gonzo masculinity.

The second problem is that denying true expressive inventiveness makes it difficult to understand how new thoughts and reactions could ever be possible. It seems as if all each subject can do in terms of expression is to make use of already existing relations of information-carrying. For a subject to consciously to start to use some piece of behavior to convey a particular state of mind, the state of mind needs to be clear enough to the subject, in which case we are presupposing that the subject has access to some way of expressing it (even just to herself). If we allow selection by happenstance, this requires a regular connection, in which case we are again presupposing that there is a mental state with some determinate content that can stand in this relation of regularity. Again the expressive articulation of genuinely new mental states appears precluded. In so far as we think that the

possibility of new thoughts and responses is an aspect of subjectivity, then we have lost something quite important.[9]

One response to this argument on behalf of the instrumental perspective could be to say that the sort of case I imagined through a new reading of the Rosamond Vincy case, should in fact be called something else, and be accounted for as a separate topic. Not all manifestations of mental states are expressions, and those that aren't designed are real enough, but they aren't expressions, this response would go. What this response would concede, would then be that expression as signaling presupposes a form of embodiment of mental states that it cannot account for, but that also has the capacity to make the subjective point of view of others intelligible. If so then the theory appears to be less interesting than advertized.

I have argued that the Rosamond Vincy case can be read as a counter-example to Green's theory, if we read it as presenting us with a case of a state of mind taking shape in a new form of expression. Green's notion of expression as signaling cannot accommodate such a case, I claimed. I also argued that there is something important at stake in preserving a space for expression in which a state of mind takes shape in a new form of expression, namely the possibility of genuinely new thoughts and responses.

The possibility I have wanted to make room for, is of central importance to Merleau-Ponty, who is one of the two thinkers I am drawing on in articulating the descriptive perspective. In the quote I used as an epigraph to this paper, we see Merleau-Ponty describing speech as a gesture that breaks a primordial silence. In the passage from which this quote is taken, Merleau-Ponty discusses how expression as the formation of significance is easily neglected when we reflect on expression. Precisely because we reflect from

[9] We might wonder here whether the argument about expressive inventiveness and variation has any bearing on non-human animal expression. Although I think there is reason to distinguish expressive variability in non-human animals and expressive variability in human beings, I think expressive variability as such also characterizes non-human expression. Although I would not agree with Alice Crary that "individual animals exhibit the same sort of natural expressive variability that human beings to" (Crary 2016, 78, my emphasis), I would agree with her that animals exhibit expressive variability. See my "'Modes of a complicated form of life': Expression and Human-Animal Continuity" for a discussion of this point (Bäckström 2018).

within a world where we are already thinking and speaking, we tend to think that all expression is a mere peddling of already formed significations. Such significations, Merleau-Ponty argues, "assumes that the decisive step of expression has been accomplished". There must be expression that "breaks a primordial silence"—i.e. expression that involves the formation of significance, both for the subject who is doing the expressing and for the (possible) other who understands it (Merleau-Ponty, 2012 [1945], 189-190).

I read Merleau-Ponty as arguing that instrumental expression is there alright, but it is not the most fundamental or paradigmatic case of expression. The more fundamental case is a new expression, where there is at the same time a new significance formed for the subject and for the (potential) other. I now want to turn to explaining how this specific thesis forms a part of the descriptive perspective on expression.

2. The descriptive perspective: Merleau-Ponty and Wittgenstein

According to the descriptive perspective, we need a different notion of expression than the instrumental one precisely to describe how the world can come to take on significance for a subject; how she can come to *make* sense of the world. If we think back to the example with Rosamond Vincy, Merleau-Ponty would urge us to see her gesture as taking her from silence to noise (metaphorically speaking in this case since her gesture is not an audible one), from a situation as of yet decoded and formulated to a situation that has taken on a particular significance for her. To think of expression as something that breaks a primordial silence is to take it to create something that wasn't present before. To break a silence in this sense is not to take some aspect of one's subjective point of view and demonstrate it to others, it is for one's subjective point of view to take some particular shape.

Earlier, I characterized the descriptive perspective as interested not merely in what an expression is, but also and perhaps primarily in the mode of being of an expression. I went on to say that the descriptive perspective takes the question of what it is for an expression to be as inseparable from the question what it is for mental states to be. The idea of a new expression

as I have explained it using Merleau-Ponty begins to give content to what this might mean. In such an expression the state of mind and its embodiment or manifestation are not distinct phenomena. Indeed, Merleau-Ponty would argue that the language of "state of mind" and "manifestation" is misleading in that it suggests that the state of mind has a kind of descriptive priority, it suggests that the "state of mind" is an independently intelligible something, quite apart from its "manifestations".

In a sense, Merleau-Ponty's entire project in *Phenomenology of Perception* can be seen as an attempt to establish a different and less misleading vocabulary for describing expression, than the one of "manifestation of a mental state". In the course of this project Merleau-Ponty argues that speech (which on his view is continuous with gestures and non-verbal forms of expression) is an "originating realm" (202). Negatively, this means that speech is not the "external sign" of an "internal recognition". Language is not the "external accompaniment of thought" (205). Positively, it means that speech, or expression, more generally, *achieves or accomplishes* thought (or other mental states, such as emotions). Thus, for Merleau-Ponty, expression (in one important sense) is not a matter of manifesting inner content, but a matter of accomplishing content (which he would neither call 'inner' nor 'outer').

Merleau-Ponty does not argue that all cases of speech or non-verbal expression are originating in this sense. In a footnote he clarifies that what he says applies to "first-hand" speech, examples of which are a child saying her first words, a lover revealing her feelings, or a philosopher "who reawakens primordial experience" (208, footnote 5). Second-order expression, which he calls speech-about-speech, "makes up the general run of empirical language" (207, footnote 4). Thus, Merleau-Ponty's argument is not that all or even most we say or otherwise express is aptly described as the formation of significance or as expression in the primary sense. Rather, the point is that alongside speech that merely repeats already formed significances (second-order-speech), there must also be the phenomenon of significance accomplished or formed in expression in the primary sense.[10]

[10] Baz (2017) articulates a similar point: "Part of what Merleau-Ponty is trying to get us to see is that we are not confined to the impersonal way of seeing things:

I said above that the descriptive perspective is interested in the mode of being of an expression, and *eo ipso* in the mode of being of mental states. Now, what does this mean, and what does it have to do with my reading of the Rosamond Vincy case and Merleau-Ponty's primordial gesture? If we think back to the description that initiated Green's inquiry, expression as coping with a felt need to manifest one's mental states, we can say that Merleau-Ponty's primordial gesture (Rosamond's twist of the neck) tells us something about what it means to *have* mental states. If expression is a sort of thing in which someone's obstinacy not merely is manifested but comes to be, we need to bring in expression right in our description of what mental states are. And doing so will entail putting into relief and questioning mental states as something we simply "have". Perhaps "having" them is rather different from the "having" of other sorts of things.

I see the later Wittgenstein as sharing Merleau-Ponty's concern with questioning, through an interest in expression, what it is to have mental states. In *The Philosophical Investigations*, this concern is brought to the fore in a series of paragraphs starting at roughly §281, where Wittgenstein approaches from different angles the question of what is for pain to be. We talk about pain as something we have in parts of our bodies, but if we consider these forms of talk too much in isolation from the rest of our use of the concept of pain, we risk understanding having pain as having an object, such as having a broken bone in one's foot, or possessing an attribute, such as having black hair. In §302, he turns to the question what it is to imagine someone else's pain. Whatever this means, it can't mean, he claims, to "make a transition from one place of pain to another". What this would amount to is to imagine oneself feeling pain in some region of someone else's body. He concludes this paragraph by saying, "Pain-behavior may

though we must always rely on an inherited background of impersonal meanings[…] there is always the possibility of seeing things more or less creatively, differently, personally. […] Consider how even the most basic biological states and functions—hunger, thirst, eating, drinking, being hot or cold, needing and seeking shelter, experiencing sexual desire, satisfying it, and so on—have come to mean for us so much more than whatever they might be thought to mean 'purely biologically'; and consider also how each one of those states and functions may still be given a more or less new and personal meaning by an individual[.] (350)"

point to a painful place, but the subject of pain is the person who gives it expression." (Wittgenstein, 2001, [1953], §302) If you want to say that pain is a possession or an attribute, Wittgenstein argues here, remember that it is something a subject expresses. A subject does not express her broken bone, or her black hair. To express something isn't merely or primarily to describe or ascribe it to oneself, nor is it to show or point to something. Such language retains the idea that there is something, a primary reality, that the expression describes, points to, or shows. An expressive movement, rather, can be a movement of becoming, where that which is expressed is completed and shaped in its expression.[11]

What unites Wittgenstein and Merleau-Ponty here is the idea that our thinking about the mind and the body has a tendency to proceed in neglect of the fact that we are subjects of expression. This often takes the form of levelling important distinctions between having a living body and having other "possessions". Wittgenstein highlights that when we talk about pain located in a part of the body, the body is not here an object in the same sense as, say, a stone is an object. He engages in a thought-experiment:

> Couldn't I imagine having frightful pains and turning to stone while they lasted? Well, how do I know, if I shut my eyes, whether I have not turned into a stone? And if that happened, in what sense will *the stone* have the pains? In what sense will they be ascribable to the stone? And why need the pain have a bearer at all here?!
>
> And can one say of the stone that it has a soul and *that* is what has the pain? What has a soul, or pain, to do with a stone?
>
> Only of what behaves like a human being can one say that it has pains. (Wittgenstein 2001 [1953], §283)

[11] My understanding of expression, in particular in relation to Wittgenstein, is heavily indebted to the work of David Finkelstein. In his book Expression and the Inner (2003), Finkelstein argues that pain and its expression "make sense together in something like the way that two parts of a single sentence do" (135). This metaphor pulls us in the direction of thinking of expression as precisely something which finishes or accomplishes a unity of significance, not as something which indicates or points to something already completely determinate and fixed.

We might think that this last comment shows us that Wittgenstein is skeptical about attributing pain to other animals than human beings. But this is not what he means with the phrase "behaves like a human being". He might have said, instead: only of an expressive creature can we say that it has pains. Pain belongs in the life of an expressive animal.[12] This becomes clear in the paragraph immediately following the previous. There he compares attributing pain to a stone to attributing it to a number. But, he says, "look at a wriggling fly and at once these difficulties vanish and pain seems to be able to get a foothold here, where before everything was, so to speak, too smooth for it (§284)."

A stone or a number is not an expressive body, whereas a fly is, or is at least a limiting case of one. An expressive body is one about which our concept of pain gets a foothold. This means that for Wittgenstein, while expression can be an instrument of communicating this or that specific pain, or this or that specific desire, it is also and importantly the mark of a conceptual (grammatical, categorical) difference: it indicates a qualitative break. Paragraph 284 continues:

> And so, too, a corpse seems to us quite inaccessible to pain.---Our attitude to what is alive and what is dead, is not the same, All our reactions are different.---If anyone says: "That cannot simply come from the fact that a living thing moves about in such-and-such a way and a dead one not", then I want to intimate to him that this is a case of the transition 'from quantity to quality'. (284)

The final sentence of this paragraph is, I think, Wittgenstein's attempt to find a language for what Merleau-Ponty describes as the "gesture that breaks the primordial silence". In the previous section I explained Merleau-Ponty's idea in terms of a new expression, a mental state coming to take shape in a novel expression. But we might understand the idea of a primordial silence as also indicating how expression is descriptively distinct from other forms of movement. A living thing does move about in such-and-such ways, where a non-living thing might move about in different ways (a stone

[12] Here again I am indebted to Finkelstein who says that, "a pain and its expression hang together in the logical space of animate life" (2003: 135).

might fall or roll, for instance). But "move about" then is a phrase covering up a qualitative difference, a radical break. An expressive movement is preceded by silence, in the sense that its mode of being is different from the mode of being of something dead, non-expressive.

In the introduction I mentioned that Green also wants to articulate a constitutive connection between expressions and mental states. In distinction with Wittgenstein and Merleau-Ponty, however, he does not think articulating this constitutive connection requires us to be suspicious of using language to describe this relation taken from the domain of mere physical objects. If we look at the analogies Green uses to describe how expressions are bound up with mental states, it will be easier to see how his perspective differs from the descriptive. This will also allow me to re-state the challenge I posed in the previous section, but from a different angle.

3. Apples, galaxies, and emotions

I have argued that we need space for cases where mental states take shape in novel forms of expression, and that doing so requires us to think about what it is for expressions and mental states to be. Now, Green also thinks that there are constitutive links between mental states and their expressions. This is not a required aspect of the instrumental perspective as I have articulated it, but it is a thought Green wants to make space for. This in turn has to do with the fact that Green wants room for the possibility of directly perceiving other people's mental states rather than inferring them from their behavior. (He shares this ambition of making space for perceiving rather than inferring mental states with both Merleau-Ponty and Wittgenstein.) In Green's theory, expressions are described as (at least sometimes) characteristic components of mental states (Green 2007, 88-93). Green's primary example in this regard is emotions. Emotions are such as to include (at least some of) their signals as parts, he argues. When one perceives an expression of emotion that is a characteristic component of the emotion, one *eo ipso* perceives the emotion itself.

Now, at first blush this is a difficult thought to wrap one's head around. The idea of a signal starts to look oddly self-referential: a signal is designed to convey information about a mental state of which itself is a part. For

this idea to be less mind-boggling, we need to think of the signal-part or aspect of the emotion as having some relatively robust criteria of identification that sets it apart from "the rest" of the emotion (which the signal-part is designed to convey). It is, after all, not particularly mind-boggling to think about a sweater as a composite whole that includes the tag designed to convey information about its properties.

When we try to think about what this part/whole relation could mean in the case of emotions and their expressions, however, we encounter difficulties. There is the part of the emotion we perceive, but there are also the unperceived parts of the emotion. These unperceived parts of the emotion now look out of reach in a worrisome way. The label on the sweater mentions properties (such as what material it is made of) that we are familiar with and understand in other ways than reading about them on clothing tags. But the properties the signal makes us privy to are not, or at least not obviously, such that we could be familiar with them in other ways than through the signals. We are left with *only* the tags, as it were, and this strikes us as problematic and disappointing situation.

Green wants to articulate what sort of part/whole relation he envisages by using two analogies, first with apples and then with galaxies.

> Someone who presents to me an apple from one angle has thereby shown me an apple even if I do not inspect its interior or its other side. The reason is that a sufficiently large portion of a side of an apple is, for normal human observers, not only itself perceptible but also a characteristic component of the apple. (Green 2007, 86)

Now, an apple can of course be turned around, in which case its other side would appear to perception. Apples are not such that their back sides are perpetually hidden from view. But what about the other components of states of mind? Green responds to this worry with a new analogy: galaxies. He concedes that we cannot see all parts of an emotion, but that this is in fact not peculiar to the case of emotions. We can see galaxies, Green argues, even when we cannot in principle perceive the black hole in their middle (Green 2007, 89). Hence, there are perceptible objects with parts in principle hidden from view. Emotions are akin to, Green, suggests, such perceptible objects.

From the perspectives of Merleau-Ponty and Wittgenstein, these analogies raise more descriptive and epistemological questions than they answer. The first idea, that of an expression as the "facing side" of an emotion, pushes us in the direction of thinking of that there must be another perceptual vantage point on the emotion, or some way of making the non-facing side appear. When we then realize that there doesn't seem to be any such vantage-point (apart from, perhaps, the one of the subject herself) we are thrown upon worries about how we could know anything about the properties of the non-facing side. The second analogy, where emotions are galaxy-like complex objects with invisible centers, leaves us wondering where we would get the idea of such an emotional center. Galaxies and black-holes are objects of natural-scientific understanding and theorizing. But it is less clear by virtue of what theory-like construction we would get any robust idea of the invisible emotional center of a complex mental state.

For Wittgenstein and Merleau-Ponty both, expression requires us to look with suspicion on an attempt to assimilate mental states to physical objects. But a notion of expression that is antecedently conceived of in instrumental terms will not be of any use in this respect, they would argue. When Green includes signals as characteristic components of emotions, he then (they would say) effectively tacks signals onto something we have been given scarce resources to understand. From the perspective of Wittgenstein and Merleau-Ponty, Green misses the chance to let the notion of expression open for an understanding of the mode of being of expressions and, thereby, of mental states.

4. Concluding remarks

My aim in this paper was to formulate a challenge to Green's thesis that an expression is a signal, by developing a contrast between two perspectives on expression, the instrumental and the descriptive. It might seem as if I in the end articulated two distinct challenges. The first challenge would be how the instrumental perspective can make space for mental states to take shape in new forms of expression. The second would be whether the instrumental perspective has the resources to help us understand the mode of being of mental states and expressions. In fact, I take these challenges to

be one and the same. For making space for the possibility of the formation of significance through new expressions means acknowledging that at least in such a case we need to understand the relation between the mental state and the expression in terms, to borrow Merleau-Ponty's words, of the expression accomplishing or achieving the mental state. This way of putting the link is in sharp tension with the idea that the expression is one part of some complex whole whose other part (the mental state itself one is tempted to say) has some intelligibility apart from its expression. Hence, if we are to understand Merleau-Ponty's primordial gesture, we need a constitutive link that cannot be captured in the terms offered to us by the instrumental perspective.

Acknowledgments

I owe thanks to Mitchell S. Green, Martin Gustafsson, Eva Schwarz, Fredrik Svenaeus, and an anonymous reviewer for comments on previous versions of this paper.

References

Baz, Avner. 2017. "Motivational Indeterminacy." *European Journal of Philosophy* 25 (2): 336–57. https://doi.org/10.1111/ejop.12163

Bäckström, Stina. 2013a. "Our Relation to Our Own Expressions: Comment on Bar-On, Green, and Finkelstein." *Analytic Philosophy* 4/54: 466–76. https://doi.org/10.1111/phib.12029

Bäckström, Stina. 2013b. "Seeing People and Knowing You: Perception, Shared Knowledge, and Acknowledgment." *European Journal for Philosophy of Religion* 5/4: 55-73. https://doi.org/10.24204/ejpr.v5i4.205

Bäckström, Stina. 2016. "Expression and Self-Consciousness." *Philosophical Topics* 44 (1): 161–80. https://doi.org/10.5840/philtopics201644110

Bäckström, Stina. 2018. "'Modes of a Complicated Form of Life': Expression and Human-Animal Continuity". In *Wittgenstein and Naturalism*, edited by Kevin Cahill and Thomas Raleigh, 223–40. London and New York: Routledge. https://doi.org/10.4324/9781315301594

Crary, Alice. 2016. Inside Ethics: *On the Demands of Moral Thought*. Cambridge, Mass. and London: Harvard University Press. https://doi.org/10.4159/9780674089075

Eliot, George. 2000 [1871-72]. *Middlemarch*. New York and London: W.W. Norton and Company.

Finkelstein, David H. 2003. *Expression and the Inner*. Cambridge, Mass. and London: Harvard University Press.

Green, Mitchell. 2007. *Self-Expression*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283781.001.0001

Kuusela, Oskari, Mihan Ometiță, and Timur Uçan, eds. 2018. *Wittgenstein and Phenomenology*. London and New York: Routledge. https://doi.org/10.4324/9781315626307

Merleau-Ponty, Maurice. 2012 [1945]. *Phenomenology of Perception*, translated by D.A. Landes. London and New York: Routledge. https://doi.org/10.4324/9780203981139

Romdenh-Romluc, Komarine, ed. 2017. *Wittgenstein and Merleau-Ponty*. London and New York: Routledge.

Wittgenstein, Ludwig. 2001 [1953]. *Philosophical Investigations*, third ed., translated by G. E. M. Anscombe. Oxford: Blackwell Publishing.

Organon F 28 (2) 2021: 303–325
https://doi.org/10.31577/orgf.2021.28203

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

The Epistemic Value of Music

Marina Bakalova*

Received: 16 April 2019 / Revised: 9 September 2019 / Accepted: 15 December 2019

Abstract: Assuming that music can be expressive, I try to answer the question whether musical expressiveness has epistemic value. The article has six parts. In the first part, I provide examples of what music can express. I suggest that it can express inner states with phenomenal character. In the second part, I build up an argument in favor of the claim that, granted its expressiveness, music can convey conceptual content which is not verbal, and which cannot be expressed verbally. This conclusion is limited to concepts like lyrical, nostalgic, melancholy, joyful, distressful etc. In the third part, I explain what musical expressive content is, in contrast and by analogy to, propositional content. In the fourth part, I apply Mitchell Green's multi-space model of artistic expression to music. I argue that Green's theory of expression provides a powerful explanation of how a musical sequence can express states with phenomenal character. In the fifth part, I use that model to define adequacy conditions for musical expressive ascriptions. In the last part, I attempt to explain musical knowledge by combining Green's multi-space model with Sosa-style virtue epistemology.

Keywords: Cross-modal congruence; knowledge as success from ability; musical expressiveness; phenomenal concepts.

* Bulgarian Academy of Sciences
https://orcid.org/0000-0002-2847-1037
Institute for Philosophy and Sociology, Bulgarian Academy of Sciences, 13A Moskovska Str., Sofia 1000, Bulgaria
marina.bakalova@gmail.com

0. Introduction

Does listening to music deliver valuable knowledge to the listener: knowledge that is not trivial, knowledge that is causally efficient, and knowledge that one cannot easily gain otherwise? To many this question would sound unrealistically demanding *apropos* the cognitive potential of music. There is a simple reason why. Normally, I cannot let you know that mice are mammals or that *World War I* started on July 28, 1914 by playing you a piece of music. This holds for a long list of propositions about the external world that we paradigmatically take to constitute our knowledge.

In this article, I will try to show that music has epistemic value beyond that straightforward consideration, and that it gives us non-trivial knowledge which is causally efficient, and which we cannot easily gain otherwise. I claim that, granted its expressive potential, music gives us knowledge of the phenomenal character of real or imaginary inner states and episodes. In particular, it does so by displaying typical qualitative characteristics of these states and episodes in sound sequences. How can a sound sequence display phenomenal characteristics? The obvious answer is: by expressing them. When I use the term "expressing" I will mean expressing a state as opposed to expressing a particular person's state.

The presence of phenomenal characteristics, of course, essentially depends on our capability of grasping them. I shall argue, together with Mitchell Green (2007, 178-182), that we are capable of creating and grasping artistic expressions due to systematic congruencies between certain expressive stimuli and certain mental states. Such congruencies are due to a general cognitive phenomenon, known as *cross-modal (*or *inter-modal) congruence* (to be explained in Section 4). Cross-modal congruence is experimentally proven by Lawrence Marks (1978, 1982, 1987, 1995). My task in this paper is, first of all, to show that the phenomenon of cross-modal congruence is enough to explain musical expression. Because of that, I argue, Green's original theory of artistic expression stands on its own right, and does not need to be backed by the contour and convention theory.[1] Then I

[1] The contour and convention theory presented in (Kivy 1989) and (Kivy 2002) states that music is expressive as a result of a structural analogy with human behavior or, when this is not the case, it is expressive by convention.

claim that music can supply some of our concepts with conceptual content which, in long term, enhances self-knowledge and enables empathy.

Let me start with few preliminary remarks. I take it for granted that music can be expressive. I will try to give an account of what an average, even unarmed listener can get, epistemically speaking, from listening to music. In order to make things simpler, I will reduce my analysis to two main components: an attentive listener and a piece of expressive music. I will not touch upon other related topics. First, I will not mess with knowledge resulting from the ability to play an instrument or to understand scores. Our listener does not need to have a musical expertise. Secondly, our analysis is independent of three other, often taken as important, factors when knowledge is associated to music: listener's acquaintance with the cultural context of the musical work, her ability to grasp composer's or performer's intentions, and her actually feeling the expressed emotion as opposed to just hearing the expression. These three aspects would certainly enhance listener's epistemic grasp of the expression, but are not necessary for it.

Relatedly, the statement that a particular piece of music expresses something can mean two things. First, it can mean that it expresses something that a composer or a performer intended it to express. Secondly, it can mean that music itself displays emotions. Being expressive of something in this second sense is not strictly tied to anybody intending it to be so expressive. For instance, a piece of music can express an attitude inadvertently.[2] As I have already mentioned, I will use the verb "express" in the second sense, independently of the composer's intention. Consequentially, I take it that a naïve listener can gain knowledge of the qualitative characteristics of a state simply because these characteristics are present in a piece of music. Thus, I take the following simple condition as basic for my further analysis: if a piece of music is *de facto* expressive of x one can gain knowledge of x by listening to that piece.

[2] For inadvertent creations see (Zvolensky 2016).

1. What music can express

Before addressing the main issue, let me deliberate upon the question of what music is capable of expressing. Music can express states with phenomenal character. State is "a condition or way of being that exists at a particular time."[3] Arguably, music can express any state with phenomenal character.[4] Since music develops through time, the musically expressed states are usually unified in episodes that can have an overall character of their own. Inner episode is a state which develops through time, or a series of states that flow into each other.

In what follows, I will talk about music expressing emotions, feelings, moods, and sense impressions. What is the difference between them? Emotions are automatic reactions in response to our interpretation of specific triggers. They bear clear relation to whatever elicited them. An emotion is typically elicited by evaluating an event as relevant to a goal. It is positive when the goal is advanced and negative when the goal is impeded (Wilson and Keil eds. 1999, 273-275). Emotions are temporary and considered independent of logical reasoning. However, as shown by Antonio Damasio (1994), they play a crucial role in our decision making and rational choice. Feelings can be characterized as emotions processed by thinking and are usually longer-lasting than emotions. They involve awareness of an emotion and awareness of the environment. Moods have similar basis to emotions but lasts longer. Whereas emotion tends to change the course of action, mood tends to resist disruption (Wilson and Keil eds. 1999, 273-275). Next, sense impressions involve awareness due to stimulation of sense organs. The awareness is turned into and memorized as a bodily feeling associated to the presence of these stimulations. Finally, an expressive piece of music is supposed to evoke something in the listener. "To evoke" means to make someone remember or recognize something, i.e. the qualitative character of a state in a musical fragment. Hearing a state in a musical fragment, as I argue further, is enough for recognizing that state. Feeling the state (in the

[3] Cambridge dictionary: https://dictionary.cambridge.org/dictionary/english/state.

[4] States can be mental or physiological. Mental state is a psychological state as opposed to physiological state such as the state of being pregnant or the state of being thirsty. Most often music expresses mental states.

sense of emotional contagion) as opposed to just hearing it is a way of listening to music with greater depth and understanding. Yet, it is not necessary for one's cognitive grasp of the expression.

Music is most often taken to express emotions. An outstanding example of that is the 3rd part of Bartok's Concert for Orchestra, "Elegia" expressing a rich palette of grimness and emotional distress, going in many directions typical of that state.[5] In contrast to that, the end of the third part of Tchaikovsky's Sixth Symphony, "Pathetic" is massively life-approving, and joyful. Most often music expresses emotional episodes where one emotion is predominant but it is interwoven with a rich palette of other emotions.

Besides emotions music can express also sensations and it can deliver conceptual meaning, or so I would like to argue. Here are a few examples.

The first movement of Beethoven's Six "Pastoral" Symphony starts with an indication on the score by the composer "Awakening of cheerful feelings on arrival at the countryside" and indeed it sounds like being in the middle of the nature enjoying its sights. In the 2nd movement, "Scene by the brook", we can recognize the typical impressions of being around a brook. Not to speak about the 4th movement, "Thunder, Storm", where the threatening impressions of an arriving storm and thunders are ingeniously recreated by the great composer.

Next, music can express concepts or categories like elegance, lyricism, drama, nostalgia, and melancholy in a way that is illuminating in regard to obtaining (a better) understanding of these concepts. More precisely, it can express conceptual content in a non-verbal way. Mozart's Piano Sonata No. 15 (Facile), for instance, is a quite uncontaminated illustration of elegance. Played to a child it could be the child's first encounter with the conceptual meaning of elegance—a concept that is in any case learned ostensively. That

[5] The statements that I make about the examples given in this section are based on my personal introspection during listening to these pieces. They were also consulted with professor of conducting and doctor honoris causa of Bulgarian National Conservatory Ivan Bakalov to whom I am very grateful. During a presentation at the Philosophy Department of the University of Liege, Belgium, I played some of these pieces to the audience, and they agreed with the general characterizations provided here.

music will automatically supply the child's concept of elegance-to-be-mastered with reference to particular qualitative characteristics.

Here are a few other examples of how music can deliver conceptual content. Some musical pieces express concepts by previously established stereotypes. For instance, Schumann's Kinderszenen Pieces, Mendelssohn's Lieder ohne Worte or Chopin's Nocturnes are stereotypical examples of lyricism. They are conventionally described as examples of "lyricism" (e.g. in musical lexica). This does not mean, however, that they express lyricism by convention. Rather, they possess typical features which make them lyrical, and which are experienced as lyrical in a motivated manner.[6] Again, we can explain to someone what lyrical means by playing any of these works precisely because they embody typical qualities of the lyrical experience. The same holds for other concepts like "nostalgia" and "melancholy". Melancholy music can give us an idea of what melancholy is. Imagine an exercise in which you and I have to find typical musical expressions of nostalgia and melancholy that we agree upon. Such exercise would likely be illuminating for purposes of distinguishing between these two concepts, and gaining deeper understanding of each of them.

Other pieces of music can evoke conceptual content in lack of corresponding stereotypes. For example, Dan Hartman's album New Green Clear Blue is supposed to be a journey in the subconscious. In Hartman's own words:

> The basis of this album was to use tones and shades in certain patterns so the door to a listener's subconscious would be opened…There's nothing placed in the music to tell you things. It's merely my trying to tap the subconscious feelings of people who listen to it….I was unlocking my own subconscious. And the first two or three pieces I listened to I felt were planets away from fulfilling my concept. Little by little it began to flow… I sat at the keyboard and created some sounds that seemed interesting. When I listened back, I was amazed it came out completely as one piece…. I played it to my neighbors. Some said it put them in a state where they felt very intimate with themselves. (Campbell 1989, 22)

[6] I am grateful to anonymous reviewer for helping me to clarify this idea.

In short, Hartman's album enabled the listeners to explore their subconscious without a previously settled stereotype. At first, the music he wrote did not correspond to his concept, but then it came out surprisingly as he wanted it because he managed to recreate in sound the phenomenal properties that he felt were right.

Our examples are based on both absolute and program music.[7] Absolute music refers to purely abstract organization of sounds—to those musical works that contain no titles or words, such as Mozart's *Piano concerto No. 21.* Program music usually refers to songs, to music complemented with words. When philosophers discuss musical expression they prefer to limit their talk to absolute music. But program music has its own typically musical way of expression. Titles or words of a musical work can trigger one's perceptual imagination to take a certain path; yet music, on its own, does a lot, epistemically speaking, to supply these words with phenomenal content. For example, the title of Hartman's album "New Green Clear Blue" could enable one to experience a blue color as having the characteristics of a new green: freshness, serenity etc. by listening to that music (not just by digesting the verbal meaning). In other words, a musical title is just too schematic and uninformative without the proper musical part. It cannot be considered as entirely determining the musical meaning. Instead, it is just a trigger for listening in a certain way. Hence, program music on its own can convey adequate, clear, and rich information about some of the aspects of our states that we cannot so informatively talk about.

In sum, some words of our languages refer to our states: lyrical, cool, dramatic, stressful, etc. Understanding the full-fledged meaning of these words entails having been in touch with certain phenomenology. Generally speaking, that phenomenology is more efficiently communicated with music than with words. Music enables a first-hand experience of the phenomenology of our states, and thus it contributes to our conceptual knowledge. By doing so, it can serve as fuller-fledged answer to questions like "What does melancholy mean?" than a verbal definition: "a feeling of pensive sadness, typically with no obvious cause". It adds flesh and bones to the verbal meaning.

[7] Even though some of the compositions we have been discussing have titles, I have been focusing on what music can convey independently of these titles, as pure abstract organization of sounds.

2. The master argument

Musical examples given in the previous section are meant to support my main argument together with another important premise. Here is the premise. It seems to me uncontroversial that, as far as concepts about inner experiences are concerned, experiences themselves supply us with richer conceptual content than the schematic content encoded in the verbal symbols standing for these concepts. If that is true, we can draw a bolder conclusion: possession of conceptual content is possible without the existence of a corresponding verbal symbol.[8] Here is an illustration. The phrase "sexual harassment" was introduced in USA in a 1973 report about discrimination called *Saturn's Rings* by Mary Rowe (*Kamberi and Gollopeni 2015).* For sure many people had experienced sexual harassment before the phrase was introduced in English language. We can plausibly assume that, since they were acquainted with the phenomenon first-hand, they possessed raw conceptual content related to the verbal label introduced later. In other words, since they knew what it feels to be sexually harassed, they knew what sexual harassment means better than people who had never experienced it. Based on this example we can conclude that possession of conceptual content is possible without the existence of corresponding word or phrase in a language standing for the concept in question.[9] This statement is crucial for my argument.

Here is the argument itself:

1. We possess concepts of inner experiences, the content of which relates to how these experiences feel like. Examples of such concepts are: lyrical, nostalgic, melancholy, joyful, distressful etc. (assumption)
2. We can possess conceptual content related to such concepts without the existence of corresponding verbal symbols. (from our example of "sexual harassment")

[8] I admit that this step of the argument is bold given the debate over the representational nature of the mind during the last sixty years, provoked by the Language of Thought Hypothesis. For an updated version of the hypothesis see (Fodor 2008). However, I think that the statement is justified in the context of my argument.

[9] I am grateful to Zsofia Zvolensky for suggesting me to use this example.

Therefore,

3. A verbal phrase does not necessarily exhaust the conceptual content it stands for. (from 1, 2)
   Therefore,
4. In some cases, it is impossible to exhaustively communicate a concept just by using the corresponding verbal phrase. (from 1, 2, 3)
5. Music can convey to the listener conceptual content that words cannot. (from 4 and the musical examples)
6. By listening to music the listener can find in herself emotions or phenomenal experiences, more generally, which she has never experienced before. (from the example of Dan Hartman's album New Green Clear Blue)

This argument clearly approves of an unpopular position, namely that phenomenal experiences have cognitive value which is worth epistemological attention. Accordingly, the putative knowledge that we gain through music is valuable. I think that phenomenology of our senses matters cognitively for at least two reasons. First, without it, we would miss conceptual content that determines our everyday communication. Secondly, how a state feels can have significant causal consequences for our actions. For instance, there are implicit states that we cannot verbally articulate or introspect. Since we cannot introspect them we cannot monitor them either. But they can affect our behaviors. Good examples are unconscious fear and unconscious jealousy. Getting to know the phenomenal character of such states enables one to recognize them within oneself and within others when empathy is needed.

3. Expressive Content of Music

One specific difference between musical expression and verbal reports is that listening to music involves first-person perspective of the listener. Music puts the listener in intimate contact with her own states by offering a match for these states. The possibility of matching a musical theme with one's own experiential states is crucial for the capacity of music to convey knowledge to an attentive listener. The match is due to musical theme

having expressive content which corresponds to the phenomenal character of a state.

I will call articulation of phenomenal character "expressive content". I borrow the notion of content from the contemporary discussions of content of perception to stand for something *like* "the content of a newspaper" as opposed to "the content of a bucket".[10] In particular, I will use "expressive content" to stand for musical *articulation* of the expressed states.[11] Musical expressive content is very different from propositional content. However, I will argue that it has adequacy conditions by analogy to the way in which propositions have truth conditions. Adequacy conditions of musical experience are crucial for the capacity of music to convey knowledge of our inner states.

In the next two sections, I will address the following two questions: how musical expressive content can match phenomenal character of mental episodes and what adequacy conditions of musical articulation are.

4. The multi-space model of musical expression

My argument outlined above is inspired by Mitchell Green's book "Self-Expression" (Green 2007). In the last chapter of the book, Green develops a theory of artistic expression which can serve as a basis for analyzing knowledge that we typically gain through art. He points out that the main role of artistic expression is to show how our feelings feel. This is possible, Green (2007, 178-182) claims, due to the phenomenon of inter-modal congruence (also known as cross-modal congruence). According to this phenomenon, some sensations within one sensory modality seem to bear more of an affinity to some sensations within another sensory modality than to

[10] The distinction is drawn by Susanna Siegel (2016).

[11] This notion of "phenomenal content" comes close to what Paul Noordhof (2018) calls "phenomenal content": "Talk of content at this point is not meant to convey some notion of propositional content. It is simply to capture the thought that our mental lives seem made up from various elements which we cite to characterize what it is like to be undergoing them at a certain point in time. Some of these elements may be no more easy to specify than as a feeling of yearning." (Noordhof 2018, 97)

others. Examples of the phenomenon include: people systematically relating high pitch with bright light and low pitch with gloomy light. The same holds between sensations, moods, emotions, and all states with phenomenal character. For example, the major chord sounds systematically cheerful to us, whereas the minor chord sounds sad. Other examples that Green cites concern people thinking that yellow is more like the sound of a piccolo than it is like the sound of an oboe; that the smell of sulfur is more like rough than it is like smooth; that the taste of lemon is more like the minor chord C–E flat–G than it is like the major chord C–E–G, etc. (Green 2007, 179) The congruence, Green notices, holds in an irreversible way:

> (S)ome kinds of inter- as well as intrapersonal inversions do not seem possible. It is difficult, for instance, to see how there could be an interpersonal inversion as between pain and pleasure. This would require that the experience that I feel upon cutting my hand with a knife is like the experience you feel upon stroking velvet. Likewise, we can rule out the possibility that a minor chord sounds sad to me but happy to you. (Green 2007, 184)

Green provides a methodologically insightful explanation of why and how the congruence occurs. He offers a multi-dimensional explanatory model in order to explain the congruence. Green suggests that our sensations, emotions and moods may be described along a number of dimensions, including the following three:

- intense/mild
- pleasant/unpleasant
- dynamic/static.

Green limits his talk to these three basic dimensions for simplicity having in mind that the theory allows for introduction of unlimited number of such dimensions. Hence, he introduces the idea of a three-dimensional space. Green (2007, 179) asks us to imagine the above mentioned dimensions as placed in a coordinate system creating a three-space. He further assumes that we perceive phenomenal characteristics as regions in the three-space. There are few possibilities to locate states in that space. First, imagine a joy which stands phenomenally in a particular intersection between the three dimensions (say rather intense, rather pleasant and relatively

dynamic). It will be represented with one point in the three-space, an intersection between the three parameters. We will call these points "regions". Alternatively, a state might have phenomenal characteristics only along the intense-mild dimension, but not along the other two dimensions. That state will be mapped as a single point on one dimension only. Finally, a state might be two-dimensional, for instance pleasant and mild but neither dynamic nor static—such as the taste of a smooth cream. This state will be mapped with two dots on two regions along the intense-mild dimension and the pleasant-unpleasant dimension.

The idea of the three-space provides a powerful explanation of how a stimulus can express states with phenomenal character. Compare our example of a joy which stands phenomenally in a particular intersection between the three dimensions (intense, pleasant and dynamic) with an auditory experience that occupies the same or very close particular intersection of these three dimensions (i.e. an intense, pleasant and dynamic sequence of sounds). The two experiences feel the same because they are congruent in the three-space. Granted their congruence, that sequence of sounds can express that joy; a listener can hear it as that joy.

To put it more formally, let us distinguish between an expressive stimulus "ES" and an expressed state, "E". We can plausibly assume that when ES overlaps with a region or a set of regions in the three-space which E occupies, it expresses E. In the case of precise mapping between the regional spaces of E and ES, ES will be quite informative about the qualitative character of E. If E is too complex, presumably ES can give us an idea of how E feels even if it does not map all the regions of E. Certainly, there is an issue of how much overlapping is enough to make a state recognizable, but we will not deal with it here.

Let us turn back to music. Music is an art which develops through time. So, if we want to use the three-space theory to explain musical expressiveness we will have to assume that musical expression draws a trajectory of expressive stimuli or regional activations in a multidimensional space and in a multidimensional time. Musical trajectory (MT) is the musical counterpart of what I call ES together with its progression in time. Imagine a state E developing in time from point A to point D and activating different regions in the three-space at different moments: ABC and D. MT expresses

E by activating the same three dimensional regions as E. Thus MT enables a listener to phenomenally experience a mental episode developing in time. Table 1 below illustrates the idea.

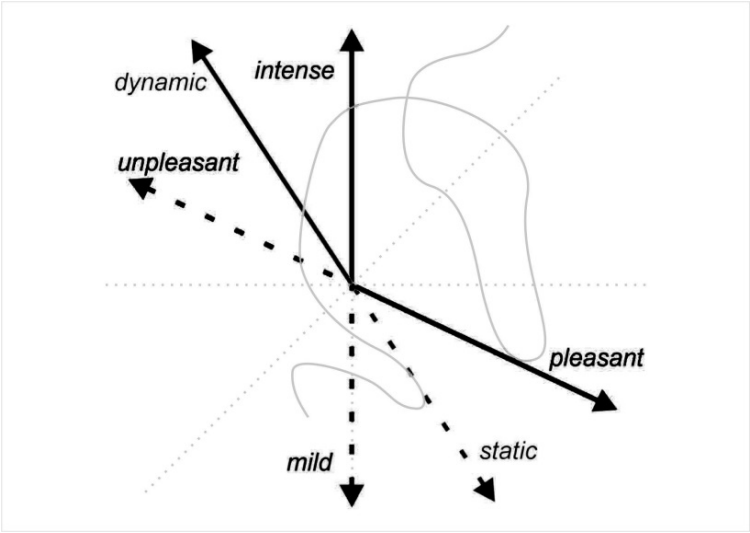

Table 1: Musical trajectory in the three-space

I would like to provide some examples of how the Western musical theory enables composers to "describe" mental attitudes using the three-space. The examples are based on tempo, dynamics and harmony.

Musical tempo allows a composer to place a musical fragment along static-dynamic dimension. The standard expressive tempo forms vary form *grave* (very, very slow) to *largo* (very slow) to *adagio* (quite slow) to *andante* (a walking pace) to *moderato* (moderate) to *allegro* (fast and cheerful) to *vivace* (lively) to *presto* (very fast) and finally to *prestissimo* (fastest). The overall shape of the melody going upward, downward, or remaining static also contributes to that.

Another chapter from the musical theory, dynamics, allows composers to express states along intense-mild dimension ranging from *pianissimo* to *fortissimo.* Directions to change dynamics either suddenly or gradually, on the other hand, enable ranging of intensity through time. The standard

possibilities here are *crescendo, decrescendo* or *diminuendo*, *sforzando*, etc.

What about the pleasant/unpleasant dimension? Music can express different pleasant states such as joy, liberation, triumph, romance etc. On the other hand, it can express sadness, drama, tragedy, horror, anxiety, and other unpleasant states. To start from the unpleasant, there are various ways to suggest an unpleasant state through a piece of music. One way is to accelerate tension, typical of emotions like anxiety, distress, drama, and horror. This can be done by increasing harmonic density, i.e. by harmonic accumulation. Similar effect can be achieved through a denser rhythm as well as by gradually increasing the density of the facture to *tutti*. We encounter such a way of building tension in "The Death of Tybalt" from Prokofiev's ballet *Romeo and Juliet* where the timpani beat and the full capacity of the orchestral *tutti* create a feeling of a strong tension and tragedy. Another example is the 4th part of Beethoven's *Pastoral Symphony* "Thunder Storm" where use of *crescendo* and the increase of density of sound and rhythm represent a growing anxiety in the face of upcoming storm. Sadness falls under phenomenally different group of unpleasant states. It is silent and slow. Sadness can be represented by using minor harmony, timbre in the middle to low register, and articulation in *legato*.

Likewise, there are various ways to express pleasant emotions. Here the choice of a theme is quite important: a well-connected, uninterrupted melody, predominantly in *legato*, consonant in terms of intervals and harmony creates a feeling of pleasure. For achieving such effect, melody should display certain diversity which helps the development of the phrase. To put it simpler—it should not be boring. Appropriate articulation and diversity of tempo also contribute to expressing pleasant emotions.

These are some basic possibilities. Composer, of course, can use various expressive means simultaneously to place a musical fragment in a single/multidimensional region, and thus to express and articulate phenomenal character of a state developing in time. Western musical system alone is a well-developed and structured palette of such expressive means.

I believe the explanatory potential of the multi-space theory, proposed by Mitch Green, is visible to the reader. One can ask what the nature of this theory is and what its advantages over other theories of musical

expressiveness are.[12] I take Green's multi-space theory to be a useful explanatory model of artistic expression. It is used to codify the ways in which we experience affinities among experiences from different sensory modalities. Also, it yields predictions concerning what sorts of congruencies people will experience. For instance, it will predict that people will find yellow more similar to the sound of piccolo than to the sound of oboe; that compositions in minor chord would, more likely, sound sad than joyful to us, etc.

Applied to music, Green's multi-space model has a significant advantage over other theories of musical expressiveness[13]. Unlike any of them, it explains musical expressiveness as property of the music itself. There are two main alternative theories of musical expression: the expressive theory and the arousal theory. The expressive theory explains musical expressiveness in terms of composer's or performer's intentions to express emotion. But it faces the inconvenience of having to explain inadvertent expressions. Green's multi-space theory accounts for that possibility: expressiveness supervenes on the quality of the sound sequence, not on anyone's intention. The arousal theory, on the other hand, states that musical expressiveness is the quality of music to trigger emotions in the listener. That theory faces a circularity problem. It affirms that musical expressiveness depends on the listener's response given that the response itself is supposed to depend upon the expressed emotion (in the *explanandum*). The multi-space theory avoids, to a certain extent the circularity of the arousal accounts. It suggests that listener's response depends upon the precise location of the musical passage in the assumed multi-space which can fail to be adequate. These overall advantages of Green's theory of expressiveness will help us to explain, in what follows, the epistemic potential of music.

5. Adequacy conditions of musical expressive ascriptions

Next, I want to address the issue of adequacy conditions of musical expression. But first, let me draw your attention to the notion of fictional

[12] I am grateful to my blind reviewer for raising this question.

[13] For an outline of the main theories of musical expressiveness see section "Emotions in the Music" of (Kania 2017).

truth in literature. There are things true in a literary fiction although, of course, they are not actually true. For instance, "Anna Karenina was driven to suicide by her failed affair" is true in the famous novel by Tolstoy. The debate about fictional truth is mostly limited to literature. However it is useful for our purposes to try to find an analogue in the domain of music. Despite the obvious differences, I claim that there is an analogy between fictional truth and musical expression. This is roughly how it goes:

> When a novel x asserts that y (and does not deny it), y is true in x, where y is a fictional event or state of affairs.

is analogous to

> When a musical sequence x stands within certain parameters in the multi space-time, it is expressive of any state with the same parameters in the multi space-time.

One may think, and rightly so, that the term "truth" is unfortunate when it comes to musical expression. In many cases, x can be expressive of p, q, y, or z depending on the associations of the listener. Nevertheless, on my view, all these ascriptions can be *adequate* to x if they are congruent with x. "Being adequate to x", according to the multi-space theory means occupying the same region in the multi-space that x occupies. So, the listener can adequately associate x with any possible state that occupies the same region in the three-space that x occupies. Adequacy is the right term here and, of course, it differs from truth. Yet, in some cases, the musical context enables us to choose the most plausible referent of the expression.

Imagine again a musical sequence developing in the multi space-time from point A to point D activating the same regions: a, b, c and d in a listener's mind. I claim that in case of adequate matching, the listener can literary perceive the expressed phenomenal properties. Green is actually skeptical about such possibility:

> Another source of resistance to the thesis that some emotions are perceptible might focus on their qualitative dimension. We've acknowledged that many emotions, including the basic ones, often characteristically have a qualitative feel: there is a certain way that rage feels to the person undergoing it; likewise for

> disgust, and perhaps also for happiness and sadness. Furthermore, how these emotions feel is not something that a third party can perceive with her senses: it is far from clear how I might go about observing the qualitative character of your emotion. (Green 2007, 91)

It seems to me that, given the explanatory potential of Green's own multi-space model, such skepticism is unjustified. Music is not only capable of showing how certain states feel but it can do so by making the qualitative characteristics of such states perceptible to the audience. I want to argue that the strength and the originality of Green's theory comes vastly from the fact that it accounts for such a possibility. This gives it a significant advantage over Peter Kivy's contour and convention theory which Green thinks is a necessary complement of his multi-space model. Although Green draws inspiration from Kivy's valuable legacy in philosophy of music, I do not think that he needs to consider the contour and convention theory as an indispensable ground for his own theory.

Let me explain why. Peter Kivy (1980) claims that music expresses emotion by being structurally analogous to human expressive behavior or, where that is not the case, it expresses emotions by convention. The theory fits to a great majority of musical examples.[14] However, it fails to account for the intuition that music can express contents of inner states independently of purely behavior characteristics, i.e. of purely motor behavior. For example, music can express complex attitudes that have no analogue in human behavior and, at the same time, without previously established conventions.

In her review of Kivy's book *The Corded Shell*, Anne Hall has detected a similar problem in the contour and convention theory:

> Compared to music, physical motion can be quite ambiguous in its expression of emotion. We hear music as feeling when we hear the kinds of processes and events that characterize both music and our feelings: tension, suspension, deception, growth, subsidence, resolution, disintegration, triumph. In order to hear music as feeling we must hear not just its surface characteristics but the

[14] For instance, Mussorgski's "Two Polish Jews: Rich and Poor" from Pictures at an Exhibition.

> way it works, its course of events, which is why this seems a more important part of musical experience than the hearing of music simply as analogue of motor behavior. (Hall 1984, 107)

Green's multi-space theory, in contrast, connects musical expressiveness to our inner states. If music is well equipped to recreate the phenomenal path of our mental lives, one might wonder why, at all, considering the analogy with motor behavior? One reason is because music sometimes characterizes events in the external world as opposed to inner states. Such events have contour which the multi-space theory apparently cannot account for. How would Green face the challenge? It is plausible to assume that music "speaks" of external events by sonically recreating our *sense impressions* of such events in the multi-space. Hence the multi-space theorist can assume that when a musical sequence occupies the same region(s) that our sense-impressions of an external event occupy, it expresses that event. For instance, sunsets feel mild, pleasant and static to us. So, musically, we can represent a sunset by producing a sound sequence that is mild, pleasant and static in a sunset-ish way.[15] As far as the convention part of Kivy's theory is concerned, it was offered to account for the expressive difference of major and minor chord. Again, I think that the multi-space theory provides a better explanation. Instead of saying that a composition in minor chord sounds sad by convention, the theory says that these compositions have characteristics that occupy the same regions as our sad states. If all that is correct, the multi-space theory has the explanatory resources of the contour-and-convention theory and a sufficient generality to stand on its own ground.

6. Music and knowledge. The listener's perspective

In this section, I provide an account of musical knowledge and I try to explain further why listening to music is epistemically valuable. According to Peter Kivy (2002, 135-160 and 251-264) music is not about enriching our

[15] For recreating typical characteristics of an external event one may need to use a fine-grained scale of nuances. In this case—to make the sunsetish features more salient—the three parameters maybe insufficient.

knowledge; it is about pleasure and enjoyment of musical form with its development in time. From a certain, perspective Kivy is justified to claim that. As we suggested in the beginning of this article, it is clear that music is not meant to deliver propositional knowledge, at least not directly. Secondly, non-propositional knowledge is not considered to be proper knowledge by the majority of epistemologists. Epistemologists working in the analytic tradition have distanced themselves from non-propositional knowledge. There are two strategies to do that. The first one is to restrict the epistemological analysis of knowledge to propositional knowledge only. The second one is to claim that non-propositional knowledge is reducible to propositional knowledge. Also, knowledge about the phenomenology of our states has been vastly underestimated by contemporary epistemologists. One reason for that is because it is considered irrelevant to knowledge of the external world as shown by the inverted spectrum thought experiment.[16]

By "non-propositional knowledge" I mean knowledge-how. I have argued in (Bakalova 2019) that non-propositional knowledge is basic and that it is not reducible to propositional knowledge. I have also argued that both kinds of knowledge can be subsumed under a common account of knowledge offered by Ernest Sosa (2007) and John Greco (2010). The two authors claim that knowledge is success from ability, i.e. that it is an apt performance: successful, because competently formed. In this section, I will use the ability-based account of Sosa and Greco to define musical knowledge.

Ernest Sosa (2007, 22-44) clarifies the above definition of knowledge in the following way. He maintains that each performance has a goal, and can be assessed according to three main criteria: *accuracy*, *adroitness*, and *aptness* (AAA structure). Accuracy is a criterion of successfully achieving the goal, adroitness is a requirement of having the ability for successfully achieving the goal, and finally and crucially, aptness is achieving the goal as a result of applying the ability. The crucial epistemic relation of "aptness" is a causal relation between competence and reaching the goal, where competence is a cause and reaching the goal is an effect. To clarify the notion of aptness, Sosa draws an analogy between a cognizer and an archer. When an archer hits the target as a result of her skill her shot is apt.

[16] For a discussion of this topic, see (Chalmers 2004).

Likewise, when a cognizer reaches cognitive success more generally as a result of her competence, her success is apt and therefore it amounts to knowledge. Even when it comes to propositional knowledge, the epistemic normativity is a kind of performance normativity: beliefs fill in the slot of performances;[17] truth fills in the slot of success.

The target of our analysis is phenomenal knowledge obtained by listening to music. I take phenomenal knowledge to be a kind of knowledge-how: knowledge how certain mental states feel. It also entails a procedural and more straightforwardly performative kind of knowledge-how where adroitness is salient: having good orientation in the multi-space and eventually - being able to use that orientation in practice.

Let me clarify the main components of my analysis which correspond to the structural components of Sosa's account: performance, success, and justifying ability. Our journey starts from attentive listening to expressive music. The first stop is listener's *phenomenal grasp* of musical expression. The grasp is presumably based on one's sensitivity guided by cross modal congruence. It corresponds to Sosa's *performance.* Secondly, I take the ability responsible for that grasp to be *perceptual ability.* In concrete, it is the ability of *hearing* a musical sequence *as* a kind of state. Hearing something as something else is a perceptual recognitional ability. I can recognize Emma by the sound of her hills in the corridor. One way to do so is by detecting the phenomenal character of the way she walks. Likewise, musical recognitional ability reliably brings to mind matching or adequate sensations (in the sense described in the previous section). Such recognition can happen with or without emotional contagion.

Let us move to the success component. Exercise of ability in question can amount to hearing or seeming to hear a musical fragment as expression of certain state. In successful cases, one hears musical expressions adequately. If the listener only seems to hear an expression without actually hearing it, it would not count as success. In such cases, the listener might be biased by her own subjective attitudes or distracted by her momentary moods or she may have lost the ability.

[17] For criticism, see Engel (2013)

In sum, when a listener is able to follow perceptually the phenomenal character of a musical expression and she does so as a result of attentively exercising her ability, she gains musical knowledge. We are now ready to propose our definition of musical knowledge which combines Green's multi-space model of musical expressiveness with Sosa-style virtue epistemology:[18]

> *MK*: A listener knows that a piece of music is expressive of certain state *x iff* the listener's grasp of *x* is adequate because adroitly placed on the multi-space occupied by *x*.

I think that this definition of musical knowledge is not *prima facie* controversial. What is more controversial perhaps is the value of musical knowledge. So, at the end of this article, I would like to say few words about the value of musical knowledge as defined here—knowledge that is exclusively a product of listening to music. I think that such knowledge enables us to go deeper into how we experience our lives and how other people do so. It enhances our orientation in the multi-space of sensual information, and enables us not only to imagine but also to have vivid awareness of certain states when needed. For instance, the capability of evoking vivid awareness helps us to understand and empathize with others when they are in need. It matters also in situations where we have to take rational decisions based on our estimation of emotional consequences. Together with that, not only one's emotional intelligence grows, but one's expressive intelligence grows too. In sum, awareness of the qualitative side of our experiences could impact our values and decisions that we make personally and as a society: our capability of practical wisdom, interacting with each other, giving pleasure to each other, enhancing and raising the quality of our lives. Finally, I hope that this article makes a modest contribution towards appreciating the epistemic value of music and also—that it encourages the analysis of non-propositional knowledge about which we know so little.

[18] For instance the basic principles outlined in (Sosa 2007).

References

Bakalova, Marina. 2019. "Why Practical Knowledge is Basic?" In *Traditional Learning Theories, Process Philosophy and AI*, edited by Katie Anderson, and Vesselin Petrov, 51–63. Louvain-la-Neuve: Chromatika.

Block, Ned. 2003. "Mental Paint." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by Martin Hahn, and Bjørn Ramberg, 165–200. Cambridge, Mass.: MIT Press.

Campbell, Mary. 1989. "Dan Hartman Manages to turn a Career", *Mohave Daily Miner* 6, Mar. 7: 22.

Davies, Stephen. 2011. *Musical Understandings and Other Essays on the Philosophy of Music*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199608775.001.0001

Engel, Pascal. 2013. "Sosa on the Normativity of Belief." *Philosophical Studies* 166 (3): 617–24. https://doi.org/10.1007/s11098-013-0200-0

Green, Mitchell. 2007. *Self-Expression*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283781.001.0001

Hall, Anne C. 1984. Review of *The Corded Shell: Reflections on Musical Expression*, by Peter Kivy. *Music Theory Spectrum* 6: 103–10. https://doi.org/10.2307/745807

Chalmers, David. 2004. "The Representational Character of Experience." In *The Future for Philosophy*, edited by Brian Leiter, 153–82. Oxford: Oxford University Press.

Damasio, Antonio. 1994. Descartes' Error: Emotion, Reason, and the Human Brain. New York: Grosset/Putnam Book.

Fodor, Jerry. 2008. *Lot 2: The Language of Thought Revisited*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199548774.001.0001

Greco, John. 2010. *Achieving knowledge: A virtue-theoretic account of epistemic normativity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511844645

Kamberi, Ferdi, and Besim Gollopeni. 2015. "The Phenomenon of Sexual Harassment at the Workplace in Republic of Kosovo." *International Review of Social Sciences* 3 (12): 580–92.

Kania, Andrew. 2017. "The Philosophy of Music." *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), edited by Edward N. Zalta. Last updated July 11, 2017. https://plato.stanford.edu/archives/fall2017/entries/music/

Kivy, Peter. 1989. *Sound Sentiment: An Essay on the Musical Emotions, Including the Complete Text of "The Corded Shell"*. Philadelphia: Temple University Press.

Kivy, Peter. 2002. *Introduction to a Philosophy of Music*. Oxford: Clarendon Press.

Marks, Lawrence. 1978. *The Unity of the Senses: Interrelations Among the Modalities*. New York: Academic Press.

Marks, Lawrence. 1982. "Bright Sneezes and Dark Coughs, Loud Sunlight and Soft Moonlight." *Journal of Experimental Psychology: Human Perception and Performance* 8: 177–93. https://psycnet.apa.org/doi/10.1037/0096-1523.8.2.177

Marks, Lawrence. 1987. "On Cross-Modal Similarity: Auditory-visual Interactions In Speeded Discrimination." *Journal of Experimental Psychology: Human Perception and Performance* 13: 384–94. https://psycnet.apa.org/doi/10.1037/0096-1523.13.3.384

Marks, Lawrence. 1995. "Intermodal Similarity and Cross-modality Matching: Coding perceptual dimensions." In *Geometric Representations of Perceptual Phenomena*, edited by R. Duncan Luce, Michael D'Zmura, Donald Hoffman, Geoffrey J. Iverson, and A. Kimball Romney, 207–33. Hillsdale, NJ: Erlbaum. https://doi.org/10.4324/9780203773949

Noordhof, Paul. 2018. "Imaginative Content." In *Perceptual Imagination and Perceptual Memory*, edited by Fiona Macpherson, and Fabian Dorsch, 96–133. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198717881.003.0006

Siegel, Susanna. 2016. "The Contents of Perception." *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. Last updated October 4, 2016. https://plato.stanford.edu/archives/win2016/entries/perception-contents/

Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge. Volume One*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199297023.001.0001

Wilson, Robert A., and Frank C. Keil, eds. 1999. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge Mass. and London, England: MIT Press.

Zvolenszky, Zsófia. 2016. "Fictional Characters, Mythical Objects and the Phenomenon of Inadvertent Creation." *Res Philosophica* 2 (93): 311–33. https://doi.org/10.11612/resphil.2016.93.2.2

Organon F 28 (2) 2021: 326–359
https://doi.org/10.31577/orgf.2021.28204

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Self-Expression in Speech Acts

Maciej Witek*

Received: 23 July 2019 / Revised: 9 September 2020 / Accepted: 12 December 2020

Abstract: My aim in this paper is to examine Mitchell S. Green's notion of self-expression and the role it plays in his model of illocutionary communication. The paper is organized into three parts. In Section 2, after discussing Green's notions of illocutionary speaker meaning and self-expression, I consider the contribution that self-expression makes to the mechanisms of intentional communication; in particular, I introduce the notion of proto-illocutionary speaker meaning and argue that it is necessary to account for acts overtly showing general commitments that are not 'marked' as being specific to one or another illocutionary force. In Section 3, I focus on Green's account of expressive norms and argue that their function is to stabilize rather than constitute the structure of illocutionary signalling systems; moreover, I examine critically Green's idea according to which expressive norms enable us to indicate the force of our speech acts and suggest that they play a key role in the mechanisms for epistemic vigilance. Finally, in Section 4, I elaborate on the idea of discourse-constituted thoughts—or, in other words, thoughts that exist in virtue of being expressed in making certain conversation-bound speech acts—and use it to develop a more comprehensive model of the expressive dimension of speech acts.

* University of Szczecin
https://orcid.org/0000-0002-0462-1678
Institute of Philosophy and Cognitive Science, University of Szczecin ul. Krakowska 71-79, 71-017 Szczecin, Poland.
maciej.witek@usz.edu.pl

Keywords: Expressive norms; illocutionary communication; self-expression; speaker meaning; speech acts.

1. Introduction

In this paper I examine Mitchell S. Green's (2007; 2009; 2016) notion of self-expression and the role it plays in his theorizing about illocutionary communication. In particular, my focus is on expressive speech acts: assertions, requests, promises, apologies, congratulations, and other illocutions "for which Moorean absurdity is possible" (Green 2009, 155). According to Green, a characteristic feature of expressive speech acts is that they "are designed to provide propositional-knowledge-enabling evidence of psychological states" (*Ibid.*, 146); in other words, expressive speech acts are designed to signal and show introspectively available states that are specified in their sincerity conditions: assertions are designed to express beliefs, requests desires, promises intentions, apologies regrets, and so on. Green also claims that the mechanism which vouchsafes the stability of the above-mentioned signaling systems—*i.e.*, systems comprised of illocutionary acts and the psychological states they express—makes an essential use of the so-called sincerity or expressive norms; roughly speaking, the norms in question constitute the structure of the social environment in which expressive speech acts function as handicaps: signals difficult to fake in virtue of being costly to produce (*Ibid.*, 150-151).

The paper is organized into three parts. In Section 2, I begin with discussing Green's notions of illocutionary speaker meaning (Green 2007, 74) and self-expression (*Ibid.*, 43) and, next, move to considering the various roles of self-expression in intentional communication. In Section 3, after discussing Green's model of expressive speech acts as handicaps, I argue that the function of what Green calls expressive norms is to *stabilize* the structure of illocutionary signalling systems rather to enable us to indicate the force of our speech acts; in other words, I cast doubt on the idea that expressive norms enable us to indicate "how what is said is to be taken and what would count as an appropriate reply" (2009, 160), and suggest that they form an essential part of the mechanisms for epistemic vigilance. Finally, in Section 4, I distinguish between two types of self-expression: (*i*)

expressing with the thinking-to-speaking direction of influence and (*ii*) expressing with the speaking-to-thinking direction of influence or, in other words, between (*i*) expressing discourse-independent thoughts and (*ii*) expressing discourse-dependent thoughts. I argue that the notion of discourse-constituted or discourse-dependent thoughts—together with the corresponding idea of expressing with the speaking-to-thinking direction of influence—enables us to arrive at a more comprehensive model of self-expression in illocutionary communication.

2. Meaning, expressing, and intentional communication

2.1. Speaker meaning as a species of signalling

In "Self-Expression" Mitchell Green (2007) offers a redefinition of the Gricean notion of speaker meaning. The basic idea underlying his model of intentional communication is that "speaker meaning [is] a species of signalling" (Green 2007, 75). More specifically, he uses the conceptual framework of his signalling model of communication to define three notions: *factual speaker meaning* (*Ibid.*, 67), *objectual speaker meaning* (*Ibid.*, 68), and *illocutionary speaker meaning* (*Ibid.*, 74). A central idea behind these concepts is that to speaker-mean an item—a fact, an object, or a commitment, respectively—is to signal and overtly show it: to perform an action intending that in performing it, first, one enables knowledge about the signalled item in an appropriately endowed receiver and, second, makes it manifest that one has *this* intention. In the remaining part of this paper I use the term 'intentional communication' to refer to the practice that consists in performing acts of objectual, factual, and illocutionary speaker meaning; in Subsection 2.3 below, however, I introduce the notion of *proto-illocutionary speaker meaning* and argue that it is necessary to account for those forms of intentional communication that cannot be explained as cases of objectual, factual or illocutionary speaker meaning.

According to Green, a *signal* is "any feature of an entity that conveys information (including misinformation) and that was designed for its ability to convey that information." (Green 2007, 49) For instance, bright coloration in a poisonous tree frog signals potential predators that it is noxious if

eaten; an alarm call produced by a vervet monkey signals to its conspecifics the presence of a predator in their vicinity; my scowl signals to my interlocutors that I'm anxious or angry, and my smile signals that I'm happy.

Veracious or sincere signals constitute special cases of *showing*. Generally speaking, to show an item is to provide evidence for it and thereby enable knowledge of what is shown in appropriately endowed observers. I can use calculations to show that there is a black hole in the center of the Milky Way; I can roll my shoulders up to show my mosquito bites and, by the same token, to make them perceptible; and I can draw a picture of my new house to enable my friends to know how it looks. In sum—Green (2007, 47-49; 2009, 141-142) argues—there are three forms of showing that enable, respectively, *propositional*, *perceptual* and *qualitative* knowledge in appropriately endowed observers.

Consider, for instance, a frog that belongs to a population of frogs whose bright coloration is associated to a high enough degree with their being noxious; according to Green, if the frog does not use Batesian mimicry to escape predation, its bright coloration not only signals, but also shows that the frog is noxious. Consider, by analogy, a population of vervet monkeys producing alarm calls that in a sufficient number of cases correlate with occurrences of predators; a veracious alarm call produced by one of these monkeys, then, both signals and shows the presence of a predator in its vicinity. Finally, taking into account behavioral regularities to be found among humans, my sincere scowl not only signals, but also shows my anger, and my sincere smile both signals and shows that I am happy.

Let us go back to discussing Green's notion of speaker meaning. To speaker mean an item, he claims, is to deliberately and overtly show it. In other words, it is to produce or use a signal with an intention, first, to enable knowledge about the signalled item in an appropriately endowed receiver and, second, to make *this intention* manifest. Like Grice (1975) and Bach and Harnish (1979), then, Green defines the notion of speaker meaning in terms of overt and self-referential intentions.[1] Unlike Grice and his fol-

[1] A characteristic feature of a self-referential intention is that its content involves reference to the intention whose content it is (Green 2007, 66n); some self-referential intentions are reflexive in that their fulfilment consists in their recognition (Bach

lowers, however, he claims that speaker meaning involves no audience-directed intentions. In particular, it does not require intentions to produce effects on others, e.g., an intention to get the audience to form certain beliefs or other propositional attitudes. Manifesting or showing an item—an external fact or object, a psychological state, a piece of information, and so on—consists in making it publicly accessible; it can, but does *not* necessarily have to, involve bringing about a corresponding change in the cognitive state of the receiver. According to Green "it's necessary and sufficient for speaker meaning that one overtly show something, or overtly show that something is so, or overtly show one's commitment to a content in a certain way." (Green 2007, 46)

In short, what constitutes the force and content of an act of speaker meaning construed as a species of signalling is not the Gricean audience-directed intention with which it is made—if there is any—but what can be called the communicating agent's *signalling intention*: her overt and self-referential intention to manifest or show an actual fact, a real object, or her commitment to a propositional content. In putting things this way, Green rejects what Lepore and Stone call prospective intentionalism: "the view that the meaning of an utterance derives from the changes that the speaker plans for the utterance to bring about in the conversation." (Lepore and Stone 2015, 200)[2] At the same time, he integrates intentional communication—that consists in producing and interpreting acts of speaker meaning—into a broad and comprehensive picture of communication provided by the signalling model (Green 2007; 2009) recently enriched with his conception of organic meaning (Green 2017; 2019b). According to this model, "[s]ignals constitute what [Green calls] organic meaning, which overlaps with natural

and Harnish 1979). As Green (2007, 68) notes, some authors cast doubt on the possibility of self-referential propositional attitudes. For instance, Mark Siebel (2003) argues that self-referential intentions in general and reflexive intentions in particular are unthinkable mental representations. For an attempt to resist Siebel's scepticism about self-referential intentions, see (Witek 2009); for a critical discussion of Witek's defence of the psychological reality of reflexive intentions, see (Siebel 2020, 106-107).

[2] For a discussion of Lepore and Stone's criticism of prospective intentionalism, see (Witek 2016).

meaning while including speaker meaning as a special case." (Green 2017, 317)

To speaker mean an object is to overtly and deliberately show it or, in other words, to make it *perceptible to* or at least *accessible to the attention of* appropriately situated observers; to factually speaker mean that *P* is to overtly and deliberately show that *P* or, in other words, to show the fact that *P*; finally, to speaker mean that *P* φ-ly—i.e., to perform an act of illocutionary speaker meaning—is to overtly show one's commitment to the proposition *P* under force φ or, in other words, overtly show one's φ-specific commitment to the proposition that *P*. It is instructive to stress that showing, unlike signalling, is a 'success' notion: one can only show actual facts and real objects (Green 2007, 49); a faked signal only *appears* to show what it signals. For this reason, both objectual and factual speaker meaning are factive. Illocutionary speaker meaning, by contrast, is not factive: one can illocutionary speaker mean that *P*—for example, one can mean that *P* as an assertion—even though one lies or is mistaken in believing that *P*. The point is that what the speaker shows in asserting that *P* is not the fact that *P*, but her force-specific commitment to the proposition that *P*.[3] Of course in issuing an utterance that takes effect as an assertion that *P* one can, provided it is true that *P*, not only show one's commitment to the proposition that *P*, but also show the fact that *P*. It is important to keep in mind, however, that the utterance under discussion constitutes two different though closely related acts: as an act of *factual* speaker meaning, it shows the fact that *P*; as an act of *illocutionary* speaker meaning, it shows the speaker's force-specific commitment to the proposition that *P*.

2.2. Self-expression as a species of signalling

Green uses the conceptual framework of his signalling model of communication to define the following notion of *self-expression*:

> (S-E) "Where A is an agent and B a cognitive, affective, or experiential state of a sort to which A can have introspective access, A expresses her B if and only if A is in state B, and some action

[3] For Green (2007, 102-103), one cannot lie that *P* without asserting that *P*, i.e., without showing one's commitment to the proposition that *P*.

> or behaviour of A's both shows and signals her B." (Green 2007, 43)[4]

A blush on my face betrays my embarrassment. It shows my affective state or, in other words, makes it cognitively accessible or even perceptible to appropriately competent observers: it is a cue that can be used by others to recognize my current psychological state. Nevertheless, it does not signal my embarrassment: the proper function of a blush on one's face—i.e., its proper purpose understood as the effect for the production of which it has been selected for (Millikan 1984 and 2004)—is not to convey information about one's affective state; rather, it seems to be a by-product of one's "flight or fight" response to an embarrassing situation, which results from dilating one's blood vessels to increase the quantity of oxygen delivered to one's muscle.[5] In short, my blushing shows but do not express my embarrassment. By contrast, a characteristic smile on my face, if sincere, both shows and signals my friendly attitude towards my interlocutor; it is natural to assume, namely, that friendly smiles have been selected or designed for their ability to convey friendly attitudes. For this reason, one's sincere smile can be said to express one's friendly attitude. It is instructive to note, however, that insincere friendly smiles fail to express what they signal. Like showing, expressing is a 'success' term. My sincere scowl express my anger. My insincere scowl, by contrast, only *appears* or *purports* to express it.

2.3. Self-expression in intentional communication

Self-expression construed as sincere signalling of one's own introspectible mental states plays at least three roles in intentional communication. First, (*i*) in some cases what the communicating agent expresses is identical to

[4] It is worth noting—following one of the reviewers—that in one of his recent papers Green (2019a) offers a revised definition of self-expression in which he drops the "of a sort to which A can have introspective access" requirement. However, a detailed discussion of this revision goes beyond the scope of the present paper.

[5] As Green notes, however, "[a]dvances in the evolutionary biology of facial expression might show (…) that blushing did evolve to signal embarrassment" (2007, 27) and, by the same token, show that a blush on one's face is a signal; for references to recent works on blushing, see (Green 2019a).

what she speaker-means. Second, (*ii*) no matter whether it is speaker-meant or not, what is expressed may contribute to the determination of what is said. Third, (*iii*) many speech acts have a characteristic expressive dimension—e.g., sincere assertions express beliefs, sincere promises intentions, and sincere apologies regrets—which plays a role in illocutionary communication.

(*i*) Let us focus, first, on cases in which what one speaker-means coincides with what one expresses (Green 2007, 84). For the sake of illustration, let us consider what Tim Wharton (2003, 456n) calls *deliberately shown natural behaviours*: openly displayed natural manifestations—signs or signals—of what is within. For instance, my involuntary shiver means naturally that I'm cold. It is instructive to stress, however, that it is a *cue* or *sign* rather than a *signal*, since the proper function of one's shiver is to "generate heat by rapid muscle movement" (Wharton 2003, 469) rather than to convey that one is cold. By contrast, my scowl is a *signal* whose function—*i.e.*, the effect for the production of which scowls *qua* signals were selected for—is to convey that I'm angry. However, even if produced in a normal way—that is to say, if produced spontaneously as direct manifestations of what one feels—shivers, scowls, and other natural behaviours can be overtly used to communicate what they naturally signify or signal. More specifically, I can openly display my shiver to make it manifest that I'm cold; by analogy, I can deliberately demonstrate my scowl to overtly show my anger. In general, one can openly and deliberately show one's feelings by—as Wharton puts it—"making no attempt to conceal a spontaneously-produced natural sign or signal in circumstances where it is obvious to both communicator and audience that she could have taken steps to conceal them." (Wharton 2012, 575) In a similar vein, Green argues that automatically produced natural manifestations of our affective states can be regarded as intentional signals—i.e., as instances of intentional communication—provided they "are not inhibited but, (a) at the time they are manifested, could have been, and (b) we refrain from inhabiting them for a reason". (Green 2007, 96) Consider, for instance, my friend who deliberately maintains her spontaneous scowl or, in other words, purposefully makes no attempt to inhibit it. In doing this, she openly and intentionally shows her anger. In my view, she can also be regarded as *overtly* showing her anger, provided she makes clear the intention to show it.

In producing a sincere overt scowl, then, I speaker-mean my anger or the fact that I'm angry; in other words, I perform an act of objectual or factual speaker meaning respectively. At the same time I express my anger. Therefore, in deliberately and overtly displaying my scowl I speaker-mean and express one and the same thing or state. What is more, my act of intentional communication *builds* and *draws on* my act of self-expression and, as a result, what I speaker-mean is directly determined by what I express: to express one's own introspectible state is to *signal and show* it;[6] to speaker-mean what one expresses it suffices to signal and show it *overtly*. "Overt self-expression (…) is a form of speaker meaning." (Green 2007, 82)

In short, sincere and overtly demonstrated natural manifestations of what is within take effect as acts of speaker meaning. Let us have a closer look, however, at *non-sincere* overt signals of introspectible states which, according to Green (2007, 97-98), should be also regarded as cases of speaker meaning. In particular, let us consider whether sincere and non-sincere overt scowls can be accounted for in a uniform way. Recall that in producing a sincere overt scowl, I perform an act of factual or objectual speaker meaning; in other words, what I overtly show is either the fact that I am angry or my being angry, respectively. My non-sincere, overt scowl, by contrast, only *appears* or *purports* to show my anger and, for this reason, cannot be regarded as an act of objectual or factual speaker meaning (by the same token, it cannot be regarded as expressing my anger). Green (2007, 97-98) claims, however, that in making a non-sincere, though overt scowl I speaker-mean my anger. How is it possible? To answer this question, it suffices to assume that what my non-sincere scowl overtly shows is *not* my anger, *but* my commitment to my being angry. In this respect, my non-sincere scowl can be likened to an act of illocutionary speaker meaning. In the case of the latter, however, what we show is a force-specific commitment, e.g., the commitment characteristic of assertions, the commitment characteristic of presumptions, and so on. In the case of my non-sincere, overt scowl, by contrast, what I show is a general commitment that is not 'marked' as being specific to one or another illocutionary force. For this reason, I propose to describe it as a case of *proto*-illocutionary speaker meaning: like illocutionary

[6] See Green's characterisation of self-expression discussed in Subsection 2.2 above.

speaker meaning, it shows my commitment to a certain object or fact; the commitment thereby shown, however, is not force-specific.[7]

It remains to be considered whether (*a*) sincere and non-sincere overt scowls represent two distinct subcategories of speaker meaning—that is, factual and proto-illocutionary, respectively—or, rather, (*b*) should be accounted for in a uniform way as instances of proto-illocutionary communication. Although I am more sympathetic to hypothesis (*b*), I think that its justification requires empirical studies on overt self-expression. It should be examined, for instance, whether (*a*) only non-sincere manifestations of affective states have a normative dimension, whereas their sincere counterparts are cases of merely factual speaker meaning, or, rather, (*b*) both sincere and non-sincere overt manifestations give rise to normative expectations on the part of the audience. [8] My hypothesis is that the latter is the case: in my view, there are no independent reasons for assuming that the difference between sincere and non-sincere overt manifestations corresponds to the distinction between acts of *merely* factual speaker meaning (i.e., cases of showing facts but not commitments) and acts of *merely* commitment-incurring speaker meaning (i.e., cases of showing commitments but not facts).

(*ii*) Let us focus on the role of self-expression in determining the intuitive and occasion-sensitive truth-conditional content of an utterance, which can be called, following the Gricean tradition in pragmatics, its *impliciture* (Bach 1994), *explicature* (Wilson and Sperber 2012) or, simply, *what is said* (Recanati 2004).

As Wharton observes, our speech acts *are accompanied by* or *involve* the use of a number of paralinguistic indicators—such as facial expressions, intonation patterns, gestures, and so on—that function as signals conveying information about what we feel while speaking; they make up natural codes

[7] It is instructive to note that acts of proto-illocutionary speaker meaning are akin to what Green (2007, 100-101) calls "non-conventional, non-conversational implicatures."

[8] As one of the reviewers rightly notes, one can take the sincere and overt cases to be both factual and commitment-incurring acts of speaker meaning and at the same time describe non-sincere overt cases as commitment-incurring only. In my view, however, this option is in line with hypothesis (*b*).

or natural signalling systems that enable us to show and express our affective states. At the same time they may contribute to the determination of what we say or, in other words, play a role in constructing the occasion-specific truth-conditional contents of our utterances. Consider—following Wharton (2003, 463)—a situation in which Jack utters the following words in a furious tone:

(1) I am angry.

and frowns sternly in an overt manner. The angry tone of his voice and his frown show and express his affective state. What is more, the furiousness of his voice—and the same can be said about the sternness of his frown—*corresponds to* and *indicates* the intensity of his anger. Independently of whether it is speaker-meant or not, the expressed content contributes to the determination of what Jack says or, in other words, to the occasion-specific truth-conditional content of his utterance. It is instructive to note, namely, that sentence (1) contains a token of 'angry', whose interpretation requires establishing the type and degree of Jack's anger. To arrive at such an interpretation, Jack's interlocutor makes use of "her ability to discriminate among tiny variations in his facial expression and tone of voice" (Wharton 2003, 464), that is to say, her ability to read natural manifestations of Jack's anger.[9] As Wharton puts it:

[9] One of the reviewers is critical of the above-presented discussion of the 'I am angry' example and claims that the predicate 'angry' does not behave in the way described by Wharton. To support his or her opinion, the reviewer points out that, first, we can refer to two people—one moderately angry and the other very angry—and say:

(2) Both of those two people are angry.

Second, he or she claims that language interpretation is on the whole digital, while reading facial expression is analogue. Even though I agree with the two points made by the reviewer, I do not think that they speak against Wharton's analysis of the case under discussion. First, viewed from the perspective of the post-Gricean truth-conditional pragmatics (Carston 2002; Recanati 2004; Sperber and Wilson 2012; for a critical overview see Chapter 1 of Jaszczolt 2016), the two different tokens of 'angry'—one occurring in the utterance of (2) and the other occurring in Jack's utterance of (1)—encode the same concept ANGRY, which provides an input to a pragmatic process of modulation. In the two cases under discussion, the process gives

> the paralinguistic indicators Jack uses overtly will play a role in Lily's determining not only what—to return to Grice's terminology—he has meant$_{NN}$, but also the proposition she takes Jack to be expressing (or what he has said). 'Angry' is a degree term, and the truth conditions of Jack's utterance of 'I am angry' will vary according to the type or degree of anger he intends to communicate (and hence reflects in his paralinguistic behaviour). (Wharton 2003, 463)

(*iii*) Illocutionary acts "for which Moorean absurdity is possible" (Green 2009, 155) have a characteristic expressive dimension: assertions are naturally taken to express beliefs, requests desires, promises intentions, apologies regrets, and so on. According to Green, the verb 'express' used in the preceding sentence should be read as 'show and signal one's own introspectible state'; in other words, the common-sense idea of expressing thoughts with words can be adequately elaborated and theoretically refined along the lines of his model of self-expression.

---

rise to the construction of two different occasion-specific or *ad hoc* concepts ANGRY* and ANGRY**, respectively. According to the relevance-theoretic perspective adopted by Wharton (2003; 2012; see also Wilson and Wharton 2006), the concept ANGRY* can be characterised by its inferential role (Wilson and Sperber 2012: 22), i.e., by reference to the logical and conversational implications derivable from the utterance of (2) that make it a relevant ostensive stimulus (see *Ibid.*, 107-115). It can also be argued that the information carried by natural signals produced by Jack—e.g., a certain prosodic pattern and facial expression—plays a role in the context-sensitive process that takes the linguistically encoded concept ANGRY as its input and gives the modulated concept ANGRY** as its output. In other words, it can be argued that the information plays a role in constructing the explicature of Jack's speech act: the occasion-sensitive proposition that results from enriching the logical form of his utterance and interprets the thought he intends to communicate. Second, one can reconcile the idea of language interpretation as a digital process with the post-Gricean idea of pragmatically modulated concepts. Linguistic interpretation is digital in that it delivers encoded concepts that function as linguistically determined and stable meanings of lexical items; the concepts so delivered, however, undergoes the pragmatic process of modulation that gives rise to a plethora or even continuum of occasion-specific meanings. It is instructive to add that the modulating processes under discussion can make use of the information carried by natural signals that constitute analogue codes.

Generally speaking, sincere expressive speech acts both signal and show speakers' current psychological states. In this respect, then, they are akin to sincere and overt natural manifestations of psychological states: my sincere assertion that *P* and my sincere scowl express my belief that *P* and my anger, respectively; what is more, my insincere assertion that *P* only *appears* to express my belief that *P*, and my non-sincere, overt scowl only *appears* to express my anger. Recall, however, that in deliberately and overtly displaying my scowl, I speaker-mean what my scowl expresses; in other words, what is speaker-meant is directly determined by what is expressed. In the case of expressive speech acts, by contrast, speakers do not mean what their acts express. In asserting that *P*, for instance, I speaker-mean that *P* assertorically and express my belief that *P* (more specifically, I express my belief that *P as* justified in a way appropriate for knowledge; for a discussion of this idea, see Section 3 below). In short, it is relatively easy to identify and describe the role of self-expression in acts of what I dubbed proto-illocutionary communication: what is expressed directly determines what is meant. It is not so easy, however, to identify the role of self-expression in illocutionary communication. I consider this topic in the next section.

3. Self-expression in illocutionary signalling systems

A central idea behind Green's account is that expressive speech acts are designed to show the states that are specified in their sincerity conditions: assertions are designed to show beliefs, directives desires, and promises intentions. Recall, however, that the verb 'express' that occurs in constructions of the form '*X* expresses mental state *MS*' is factive: according to Green, an asserter "who is not sincere provides good evidence for a belief that she lacks, and thus *only appears* to show her belief" (Green 2009, 146; the italic is mine—MW).[10] It remains to be examined, therefore, what stabilizes the *illocutionary signalling system* that consists of our expressive

[10] It is instructive to note that in his more recent definition of expression presented in (Green 2019a), Green drops the requirement that the verb 'to express' is factive. In the present paper, however, I stick to the definition that comes from (Green 2007) and (Green 2009).

speech acts—e.g., assertions—and the mental states they express. In short, what vouchsafes the reliability of speech acts understood as expressing signals?

To answer this question, Green argues that expressive speech acts in general and assertions in particular are handicaps: signals "that can only be faked with great difficulty as a result of being costly to produce." (Green 2009, 150-151) For instance, a male peacock's tails, when produced and maintained in a predatory environment, functions as a reliable signal of its owner's fitness. What vouchsafes its reliability is the fact that it is a handicap: growing and carrying a long tail costs extra calories; what is more, long and colourful feathers makes their owner slower and easier for predators to spot and catch. The crucial point, however, is that in producing and maintaining a long tail, a peacock who lives in a predatory environment incurs the cost of closing off the option of not being sufficiently fit without exposure to the risk of being easily hunted and killed; in other words, peacocks whose trains are faked signals of their fitness incur the cost of being exposed to an unusually high risk of losing life. By analogy, a speaker who asserts that P incurs the cost of closing off the option of not having the belief that P without exposure to the risk of being spotted by a social 'predator': an epistemically vigilant (Sperber *et al.* 2010) agent who, after recognising the speaker's insincerity, will reproach or even punish her and, as a result, compromise her reputation as a credible social partner. As speakers we put a lot of effort into creating and maintaining our image of ourselves as reliable and credible persons.[11] In performing insincere assertion, then, we *expose ourselves to* or at least *increase* the risk of compromising the opinion others have on us.[12]

[11] In (Witek 2019d) I have argued that in theorising about the normative aspect of illocutionary dynamics we should distinguish between perlocutionary and illocutionary credibility: one's power to "produce certain consequential effects upon the feelings, thoughts, or actions of the audience" (Austin 1975, 101) and one's collectively agreed authority to bring about changes in the domain of normative facts "such as attributions of rights, obligations, entitlements, commitments." (Sbisà 2002, 434) For the present purposes, however, it suffices to use the general notion of credibility.

[12] As one the reviewers has noted, we also incur the risk of a loss of credibility when we make sincere assertions based on insufficient evidence.

In short, in virtue of being handicaps our expressive speech acts are reliable indicators of our current psychological states: beliefs, desires, intentions, regrets, and so on. It is instructive to note, however, that signals function as handicaps only if produced and interpreted in a certain environment. For instance, a peacock's train is a handicap only in a predatory environment. By analogy, expressive speech acts are handicaps because they are issued and interpreted in a normative environment: a social context the structure of which is constituted by the so-called sincerity or expressive norms and the normative expectations they give rise to. According to Green (2009, 154), the general form of expressive norms is given by the following schema:

(EN) One who produces S is to be in condition C; otherwise she is subject to a loss of credibility.

For example, the practice of making assertions is governed by the following expressive norm:

(AN) One who asserts that P is to believe that P; otherwise she is subject to a loss of credibility.

When performed in an appropriate normative environment, then, expressive illocutionary acts can be regarded as reliable indicators of what is within. It is difficult to fake them because of the limitations put on by expressive norms of the (EN) form.

Green's idea of illocutionary acts as veracious signals of what is within gives rise to a number of questions. One can ask, first, (Q_1) how is it possible for an illocutionary act—or, in other words, for the use of illocutionary force—to show and express what is within. Second, it is worth considering (Q_2) what determines the correlation between types of illocutionary acts and types of expressed psychological states; in other words, in virtue of what assertions are associated with beliefs, requests with desires, promises with intentions, and so on. Third, one can ask (Q_3) what is the role that self-expression plays in illocutionary communication construed as a variant of intentional communication.

To answer question (Q_1), Green (2009) uses his model of expressive speech acts as handicaps, according to which illocutionary acts are reliable indicators of what is within. A speaker who performs an expressive illocutionary act,

then, produces strong enough evidence that she has an appropriate psychological state. For instance, in asserting that *P*, she produces a signal that provides strong enough evidence that she believes that *P*. If, in addition, her assertion is sincere, it enables propositional knowledge about her belief in her interlocutors; in short, it shows and thereby express her belief that *P*.

Let us focus on question (Q_2), which, as far as I know, is not directly addressed by Green. At first sight, it seems to have an obvious answer. One might think, namely, that the signification relation between illocutionary acts and what they express is determined or fixed by expressive norms of the (EN) form; in other words, expressive norms determine the structure of illocutionary signalling system. For instance, assertions correlate with beliefs—or, more specifically, one's assertion that *P* signals one's belief that *P*—in virtue of the constraints imposed by the norm (AS). In my view, however, expressive norms *presuppose* rather than *determine* the signification relations that hold between expressive speech acts and expressed psychological states; that is to say, their function is to *stabilize* rather than *constitute* the structure of illocutionary signalling systems.

One way to answer question (Q_2), I think, is to adopt Millikan's (1984; 2004; 2005; cf. Witek 2015a; 2015b) teleosemantic theory.[13] Elsewhere (Witek 2019b), I have argued that the signification relations that hold between speech acts and psychological states—e.g., between assertions and beliefs, requests and desires, and so on—are to be analysed in terms of Normal conditions for proper functioning of expressing speech acts construed as cooperative intentional signs in Millikan's (2004) sense; cooperative intentional signs, in turn, can be likened to signals in Green's sense:

[13] One of the reviewers suggests that Millikan's teleosemantic framework might be used to address question (Q_1), too. I am very sympathetic to this suggestion. As a matter of fact, elsewhere I have argued that the "capacity of speech acts to express psychological states (…) can be accounted for in terms of [Millikanian] speaker-hearer conventional patterns and the role they play in coordinating joint actions. The normativity of sincerity [or expressive] rules, in turn, can be explained in terms of Normal conditions for proper functioning of illocutionary acts *qua* cooperative intentional signs." (Witek 2019b: 91) A detailed discussion of this topic, however, goes beyond the scope of the present paper.

items that convey information (including misinformation) and have been designed for their ability to convey that information. Roughly speaking, the fact that one's assertion is sincere in accordance with the relevant signification relation—that can be called, following David Lewis (2002), the signaller's *contingency plan*—is a Normal condition for its proper functioning as an expressive signal, where the proper function of one's assertion *qua* assertion is to contribute to the achievement of what I call *mental coordination*: a preferred correspondence between the interacting agents' individual representations of their shared mental states. A detailed discussion of this proposal goes beyond the scope of the present paper.[14] For the current purposes it suffices to note that the signification relations holding between sincere expressive illocutionary acts and what they express constitute the Normal condition under which the acts can contribute to the achievement of mental coordination between the conversing agents, where 'Normal' is to be read along the Millikanian lines as standing for conditions to which a given device or trait has been adapted rather than for *statistically normal* conditions in which it functions. In short, the sincerity of a speech act constitutes the Normal condition for its functioning as a devise for establishing mental coordination.

Finally, let us address question (Q_3) and consider the role that self-expression plays in illocutionary communication. Unlike acts of proto-illocutionary communication, speech acts do not speaker-mean what they express. Therefore, their expressive dimension does not play a direct role in the determination of their communicated content. Green's proposal is that the function of self-expression in illocutionary acts qua illocutionary acts is to help indicate the force with which the speaker says what she does. He claims:

> Holding fixed what is said, expressive norms enable us to indicate how what is said is to be taken and what would count as an appropriate reply. Such norms enable us to do that by enabling us to show the psychological state (belief, acceptance, belief as justified, etc.) from which the conversational contribution flows. One could also report the psychological state from which the

[14] For a discussion of the idea of mental coordination, see Witek 2019b.

> contribution flows ('I believe that p' , etc.), but for most purposes such explicitness is otiose as compared to the stunning power of speech acts to telegraph our states of mind. (Green 2009, 160)

Let us note, first, that from the perspective of the teleosemantic account of illocutionary signalling systems it would be better to speak of psychological states from which conversational contributions Normally flow. Second, the proposal in question pertains only to sincere speech acts; it can be extended so as to cover cases of non-sincere illocutions by saying that expressive norms enable us to indicate the force of an utterance—i.e., "how what is said is to be taken and what would count as an appropriate reply"—by enabling us to pretend to show the introspectible state that the resulting act is designed to signal. Third, Green's account of the role of self-expression in illocutionary communication seems to be circular: the use of an illocutionary force shows a certain psychological state and thereby indicates how what is said is to be taken; in other words, the force of an utterance indicates itself.[15]

To justify the last point, let us distinguish between the following five items:

(*a*) the use of force φ,
(*b*) the norm by which it is (constitutively) governed,

[15] As one of the reviewers points out, one can maintain the view that expressive norms (*i*) enable speakers to express, and thus show, their psychological states, and (*ii*) determine what subsequent conversational moves are appropriate, but reject the idea that they do the latter by doing the former. Such a clarified and modified version of Green's account of expressive norms is free from circularity. Another reviewer, by contrast, has his or her doubts whether the circularity charge against Green's account of expressive norms and their role in illocutionary communication is fair; he or she claims, namely, that expressive norms establish the correlation between types of illocutionary acts and types of psychological states, and, as the corollary of this, they enable the speaker to indicate the force of her act by enabling her to show the psychological state from which the act flows. In my view, however, the account under discussion is circular: it says that expressive norms of the form "one who uses force *F* ought to have psychological state *PS*" enable the speaker to use illocutionary force *F* to show her psychological state *PS* and, in this connection, indicate the force of her act; in short, her use of illocutionary force *F* indicates itself.

(*c*) the psychological state that is shown in making the resulting act and the *light* in which it is shown,
(*d*) the score-changing potential of the resulting act, i.e., "how what is said is to be taken and what would count as an appropriate reply" (Green 2009, 160),
(*e*) the φ-specific commitment overtly shown—i.e., illocutionary speaker-meant—by the speaker.

In my view, Green's proposal under discussion can be reconstructed as follows: (*b*) enables (*a*) to indicate (*d*) by showing (*c*). In other words, norm (*b*) plays a key role in the mechanism that enables our uses of force φ to show (*c*), and (*c*) correlates with (*d*) and (*e*); the latter two items, in turn—the score-changing potential of the resulting act and the commitment that the speaker overtly shows—are two complementary descriptions of force φ.[16] In short, the complex mechanism that involves expressive norms, it seems, enables us to use an illocutionary force to indicate itself.

In my view, there is no uniform answer to the question what indicates the force of an utterance or, in other words, what indicates how the utterance is to be taken and what would count as an appropriate reply to it. There are many types of illocutionary force indicative devices, some of which are linguistic (*e.g.*, performative prefixes, grammatical moods, illocutionary adverbs, and so on), whereas other are paralinguistic in Wharton's (2003; 2012) sense. One example of the latter can be a natural code of

[16] According to Marina Sbisà, speech acts are "context-changing social actions" (Sbisà 2002, 421) whose types are to be defined by reference to how their performance affects the context of their production; in particular, the force of an illocutionary act is to be defined by reference to how it affects the score of conversation by making certain subsequent conversational moves appropriate or out of order (for a discussion of this idea, see (Witek 2015c) and (Witek 2019c). According to Green (2007, 74), in turn, what determines the force of an act is the speaker's overt intention to show that she is committed to the propositional content she puts forth in a way defined by appropriate conversational norms. In my view, these two descriptions of illocutionary force are complementary: the commitment that is overtly shown by the speaker—*i.e.*, the commitment she speaker-means—can be spelled out in terms of the score-changing potential of her act; for a discussion of this idea, see (Green 2007, 72-73), (Green 2009, 157), and (Witek 2019b, 92).

prosodic patterns whose function is to indicate or even deliberately show certain aspects of the speaker's psychological states. For instance, a tone of my voice can be a reliable indicator of whether the state that I express in uttering sentence "*P*" is my belief that *P as* justified in a way appropriate for knowledge or, rather, my belief that *P as* backed with some justification; in the former case, my utterance is to be taken as assertion, whereas in the latter case—as a conjecture.[17] Therefore, at least in some cases self-expression—that, for instance, can employ natural codes—plays a role in indicating the force of an utterance. To acknowledge this, however, is not to say that it plays such a role every time expressive illocutionary acts are made.

What is, then, the relation between (*c*) the psychological state that is shown in making the resulting act and the *light* in which it is shown and (*e*) the φ-specific commitment overtly shown by the speaker? State (*c*) is expressed, but not speaker-meant. Commitment (*e*), by contrast, is speaker-meant, but not expressed. There is, however, a *norm-based* or, provided expressing rules of the (EN) form are constitutive in character, even *definitional* or *constitutive* association between (*c*) and (*e*). What is the role, therefore, that the expressive dimension of expressive speech acts plays in illocutionary communication?

My tentative answer is that the expressive dimension of speech acts plays a key role in the mechanisms for epistemic vigilance (Sperber *et al.* 2010): expressive norms enable our interlocutors to recognize sincerity conditions of our speech acts—whose force can be indicated or marked by means of different linguistic and paralinguistic devices, some of which involve forms of self-expression—and thereby to recognize whether we are benevolent and trustworthy speakers. A detailed discussion of this idea, however, goes beyond the scope of the present paper.

4. Discourse-independent and discourse-constituted thoughts

Commenting on his characterisation of self-expression (see formula (ES) discussed in Subsection 2.2 above), Green remarks that "A need not have

[17] For a discussion of the difference between assertions and conjectures, see (Green 2007, 73) and (Green 2009, 157-158).

been in state B before her act of self-expression; it is consistent with the account given thus far that what is expressed is somehow constituted by the expressive act" (Green 2007, 43). In other words, the signalling model of communication—as well as the account of self-expression developed within its framework—seems to allow for the existence and expression of what can be called *discourse-constituted thoughts*. My aim in this section is to elaborate on the idea of mental states that exist in virtue of being expressed—that is to say, to motivate and define the notion of discourse-constituted thoughts—and propose it as a refinement of the conceptual framework of Green's account of the expressive dimension of speech acts.[18]

In particular, I want to draw a distinction between expressing with the thinking-to-speaking direction of influence and expressing with the speaking-to-thinking direction of influence and, next, use it to elaborate on the idea of psychological states from which conversational contributions *Normally* flow. Roughly speaking, in expressing with the thinking-to-speaking direction of influence we signal and show our discourse-independent thoughts; in expressing with the speaking-to-thinking direction of influence, by contrast, we show and signal our discourse-constituted thoughts. The distinction between these two types of expressing shed light on the issue of sincerity in illocutionary communication. It is difficult to perform insincere speech acts with the thinking-to-speaking direction of influence because of the limits put on by expressive norms; it seems even more difficult to perform insincere speech acts with the speaking-to-thinking direction of influence, since they express discourse-constituted thoughts: mental state whose key aspects exists in virtue of being expressed and as such exemplify what Slobin (1996) calls 'thinking for speaking'.

An example of a *discourse-independent* or *ready-made* thought is my belief that Marina is an expert in speech act theory; it is part of my belief box, waiting for an appropriate conversational occasion on which it can be activated and expressed by my telling that Marina is an expert in speech

[18] The idea of discourse-constituted thoughts comes from (Jaszczolt and Witek 2018), where it is used to develop a speech act-based model of *de se* utterances; for a detailed discussion, see (Jaszczolt and Witek 2018, 198-205). A similar idea is discussed by Stina Bäckström (2021), who argues, following Merleau-Ponty, that there are thoughts achieved or accomplished "in expression in the primary sense".

act theory. One example of a *discourse-constituted thought*, in turn, is the Stalnakerian informative presupposition (4) of the speech act made by Peter in uttering sentence (3):

(3) I have to pick up my sister at the airport.
(4) Peter has a sister.

According to Stalnaker, presupposition is "a propositional attitude of the speaker" (Stalnaker 2002, 701): to presuppose that P is to *believe* (*Ibid.*, 717) or *accept* (Stalnaker 2014, 25) that the proposition that P is part of the common ground of the ongoing conversation.[19] Therefore, in uttering sentence (3) to make a certain speech act, Peter entertains a thought—a belief or acceptance state—that the proposition expressed by sentence (4) is part of the common ground between him and his audience. Let us assume, however, that prior to the time of this utterance Peter's audience had no idea whether he had a sister and that Peter was aware of their ignorance. Therefore, prior to the time of the utterance under discussion, he had no thought—no belief and no acceptance state—to the effect that the proposition

[19] One of the reviewers points out that Stalnakerian presuppositions are not beliefs but acceptance states. However, in "Common Ground", a paper published in 2002, Stalnaker takes presuppositions to be "the speaker's beliefs about the common ground" (Stalnaker 2002, 717) and uses the concept of acceptance to define the notion of common ground only. He stipulates that "[a]cceptance (…) is a category of propositional attitudes and methodological stances toward a proposition, a category that includes belief, but also some attitudes (presumption, assumption, acceptance for the purposes of an argument or an inquiry) that contrast with belief, and with each other. To accept a proposition is to treat it as true for some reason." (*Ibid.*, 716) Next, Stalnaker identifies the common ground with "common belief about what is accepted" (*Ibid.*, 716) and argues that this definition is consonant with his original idea of presuppositions as *beliefs* about the common ground. It is instructive to note, however, that in one of his more recent works Stalnaker claims that "[a]n agent A presupposes that ϕ if and only if A accepts (for purposes of the conversation) that it is common ground that ϕ." (Stalnaker 2014, 25) A detailed discussion of the evolution of Stalnaker's views on presuppositions, however, goes beyond the scope of this paper. For the present purposes, it suffices to assume that Stalnakerian presuppositions are propositional attitudes or thoughts of the speaker that can take the form of beliefs or acceptance states.

expressed by sentence (4) is part of the common ground. Nevertheless, in uttering sentence (3) to make a certain speech act, he signals and shows that he believes or accepts that it is common ground that he has a sister. This presupposition is formed *in the course* and *for the sake* of speaking. In other words, it exists in virtue of being expressed. What is more, at the very moment he utters sentence (3), Peter is justified in taking the proposition that he has a sister to be part of the common ground. Viewed from the perspective of Stalnaker's sequential-update model of accommodation, namely, the speech act under discussion is a manifest event—"an event that, when it occurs, is mutually recognized to have occurred" (Stalnaker 2002, 708)—and as such automatically updates the common ground both at the level of what is directly illocuted and, if nobody objects, at the level of what is presupposed. More specifically, if nobody says "But you have no sister!",[20] the presupposition of Peter's speech act is accommodated, i.e., the proposition expressed by sentence (4) becomes part of the common ground relative to which Peter's direct assertion is interpreted.

I do not want here to discuss the details of Stalnaker's sequential-update model of accommodation and examine its adequacy.[21] The reason I mention it is that it makes an essential use of the idea of discourse-constituted thought: Peter's presupposition that he has a sister—i.e., his belief or acceptance that it is common ground that he has a sister—is formed *in the course* and *for the sake* of speaking; in other words, it exists in virtue of being expressed. As Stalnaker puts it:

> [t]here is nothing wrong, in general, with (…) expressing a belief that one would not have if one did not express it. (Stalnaker 2002, 711)

Roughly speaking, then, discourse-constituted thoughts are thoughts whose key aspects are constituted within the progressing discourse (Jaszczolt and Witek 2018). Our discourse-independent thoughts, by

[20] Of course more conditions have to be met in order for accommodation to work. For a discussion of this topic, see (Witek 2019a) and (Witek 2019c). For a discussion of blocking the accommodation of faulty presuppositions, see (Langton 2018).

[21] For a critical discussion of Stalnaker's model of presuppositions, see (Witek 2019a).

contrast, are in a sense ready-made elements of our mental life, that wait for being activated and expressed. Consistently, we can distinguish two types of expressing: expressing with the thinking-to-speaking direction of influence and expressing with the speaking-to-thinking direction of influence. The former takes place when we reveal our discourse-independent mental states: their semantic and possibly normative properties are in a sense inherited by our words. That is to say, discourse-independent thoughts can be identified with what I call—following Green (2009, 160)—psychological states from which our conversational contributions flow. Expressing with the speaking-to-thinking direction of influence, in turn, is akin to what Slobin (1996) calls *thinking for speaking*: it takes place when we form certain thoughts for the sake of our current conversational moves and their underlying goals.

Stalnakerian presuppositions are good, but not the only, examples of discourse-constituted thoughts. In my view, many speech acts that necessarily function as conversational moves—let us call them conversation-bound acts—are best understood as expressing mental states formed for the sake of current conversational purposes. Before I get to discussing their examples, let me make a general point about the structure of discourse-constituted thoughts and the mechanism whereby they are formed.

The formation of a discourse-constituted thought can be likened to what Lev Vygotsky (1978) called internalization. According to Tomasello (1999)—who in this respect follows Lev Vygotsky—some forms of our thinking result from internalizing socially-constituted patterns of communicative practices; for instance, a child's skills for critical thinking result from internalizing patterns characteristic for those forms of problem-solving dialogues in which participants present alternative perspectives on a certain problem. By analogy, let us consider a situation in which a speaker who is engaged in a certain type of dialogue—e.g., an inquiry, deliberation, persuasion, or negotiation (Walton 2010)—makes a certain illocutionary act that takes effect as a valid or appropriate conversational move. Let us further assume that the move is appropriate only if (c_1) it has a *point*, i.e., its performance is justified by the current conversational situation the speaker and her audience find themselves in, and (c_2) the speaker, in performing the act, *exercises* or *exploits* the *role* she currently plays in the dialogue: her powers,

rights, or authority. Conditions (c_1) and (c_2) are necessary, but by no means sufficient for the appropriate or felicitous performance of a conversation-bound act. Following John R. Searle (1969, 59-60), we can call them preparatory conditions: they are *sine quibus non* of the felicitous performance of a conversation-bound act, but do not determine its essential effect. Nevertheless, they offer us an insight into the structure of discourse-constituted thoughts. In my view, in performing a conversation-bound act, the speaker forms and expresses—*expresses* with the speaking-to-thinking direction of influence—a complex thought whose key aspects correspond to conditions (c_1) and (c_2); in other words, the speaker internalizes and foregrounds the (*i*) conversation-dependent *point* behind her act[22] and the (*ii*) *authority* with which it is made.

Let us have a closer look at four illocutionary act types that belong to what Green (2009) calls *assertive family*: assertions, conjectures, presumptions, and expert pronouncements. In particular, let us focus on their expressive function and consider the nature and structure of the mental states they show. The preliminary results of the analysis are presented below in Table 1.[23] Following Green (2009), I assume that the structure of the psychological states expressed by illocutionary acts involves aspectual shapes represented by means of the relevant as-clauses.

[22] What I call the *conversation-dependent point* behind making a conversation-bound act should be distinguished from the act's illocutionary point. The conversation-dependent point behind the performance of a given act is to respond to the current conversational situation the speaker and her audience find themselves in. For instance, depending on the conversational context in which it occurs, the point behind saying that Paul pushed John can be to answer a question, explain a previously reported fact, provide evidence or counterevidence for a disputed claim, and so on. The illocutionary point of an act (Searle 1979, 2), in turn, corresponds to the essential condition in Searle's analysis of illocutionary acts (Searle 1969, 60): the illocutionary point of a promise is an undertaking of an obligation by the speaker to do something, whereas the illocutionary point of a request is to attempt to get the hearer to do something.

[23] The same pattern can be applied to analysing the expressive dimension of non-assertoric illocutions, e.g. to what Marcin Lewiński (2021) calls action-inducing speech acts whose job is to express conclusions of practical reasoning or, more accurately, conclusions of practical argumentation.

Speech act type	**Expressed mental state**
S's assertion that *P*	S's belief that *P as* justified in a way appropriate for knowledge
S's conjecture that *P*	S's belief that *P as* backed with some justification
S's presumption that *P*	S's tentative acceptance of the proposition that *P as* (*i*) formed and expressed to enable the ongoing argument to proceed despite the absence of conclusive or sufficient evidence *and* (*ii*) warranted by virtue of *S*' current epistemic and conversational situation.
S's expert pronouncement that *P*	S's belief or acceptance that *P as* (*i*) formed and expressed to address and settle the current question under discussion and (*ii*) warranted by virtue of *S*'s authority as an expert in the topic under discussion.

Table 1. Assertive speech acts and what they express

The analysis of the first two illocutionary act types comes from Green (2007). For the sake of the present discussion let us assume that assertions and conjectures are not necessarily conversation-bound acts and, in this connection, express discourse-independent thoughts. Presumptions and expert pronouncements, by contrast, are best understood as conversation-bound illocutions that express discourse-constituted thoughts. That is to say, one can specify their preparatory conditions (c_1) and (c_2) and, next, argue that the structure of the mental states the acts express involve two corresponding aspects or aspectual shapes introduced by clauses (*i*) and (*ii*), respectively.

For instance, according to preparatory condition (c_1) for presumptions, the conversation-dependent point behind an act of presuming that *P*—a move made in the course of an argumentative exchange—is to enable the ongoing dialogue to proceed towards its goal despite the absence of sufficient evidence (Lewiński 2017; cf. Witek 2019c); in other words, what motivates or even entitles one of the speakers to presume that *P* at a certain stage of the argumentative dialogue they are engaged in is the mutually recognized absence of evidence sufficient to settle the current issue under

discussion.[24] According to preparatory condition (c_2), in turn, the speaker exploits her *accessibility* to "the grounding that supports [her] presumption" (Corredor 2017, 282) or, in other words, she *exercises* her authority to make a felicitous presumption that P.[25] The speech act of presuming has its mental counterpart, which can be called *mental presumption* (Witek 2019c, 3). It is a discourse-constituted thought whose essential aspects result from internalizing and foregrounding the conversation-dependent point behind the act and the authority with which it is made. More specifically, in performing the act of presuming that *P*, the speaker *expresses* her tentative acceptance of the proposition that *P as* (*i*) formed and expressed to enable the ongoing argument to proceed despite the absence of conclusive or sufficient evidence, and *as* (*ii*) warranted by virtue of her current epistemic and conversational situation. By analogy, in making an expert pronouncement in the course of an argumentative dialogue, the speaker *expresses* her belief or acceptance that *P as* (*i*) formed and expressed to settle the current question under discussion (Roberts 1996) and *as* (*ii*) warranted by virtue of her authority as an expert in the topic under discussion. Generally speaking, the speaker who performs a conversation-bound act for which preparatory conditions (c_1) and (c_2) can be given forms and expresses a discourse-constituted thought—i.e., a mental state that exists in virtue of being expressed—whose structure involves internal counterparts of the (*i*) conversation-dependent point behind the act and the (*ii*) authority with which it is made. In other words, in forming and expressing the thought, the speaker internalizes and foregrounds the (*i*) conversational situation her speech act is designed to address and the (*ii*) conversation-bound perspective on herself that corresponds to her conversational role or authority (Jaszczolt and Witek 2018). The mental state thereby formed is a discourse-constituted

[24] The *illocutionary* point of the act of presuming, in turn, is to transfer the burden of proof from the speaker to the hearer; for a discussion of this topic, see (Walton 2010), (Corredor 2017), (Lewiński 2017), and (Witek 2019c).

[25] Elsewhere I have argued that the type of authority required to perform a felicitous presumption is procedure-based, i.e., what makes a speaker entitled to presume that *P* at a certain stage of an argumentative exchange is the fact that in issuing the act the speaker follows certain procedures of rational inquiry; for a detailed discussion of this topic, see (Witek 2019c, 23-28).

thought the expression of which involves the speaking-to-thinking direction of influence.

If the thoughts expressed by conversation-bound acts—presumptions, expert pronouncements, and so on—are discourse-constituted, then the idea of *acts-flowing-from-mental-states* fails to provide a sufficient basis for a general account of the expressive dimension of speech acts. Only discourse-independent thoughts can be regarded as states from which conversational contributions flow. Discourse-constituted thoughts, by contrast, are formed in the course and for the sake of speaking; more specifically, aspects (*i*) and (*ii*) of the speaker's discourse-constituted thought result from internalizing, respectively, the conversation-dependent point behind her act *and* the authority with which it is made.

Now let us return to discussing the stability of illocutionary signalling systems. Recall that the idea of speech acts as handicaps enables us to account for the expressive dimension of illocutionary communication. When it comes to acts expressing discourse-constituted thoughts, however, the situation becomes more complex. Expressing discourse-independent thought involves a mechanism with the thinking-to-speaking direction of influence; it begins with activating a ready-made thought and moves to signalling it with words. By contrast, expressing discourse-constituted thoughts is a process with the speaking-to-thinking direction of influence; more specifically, it consists in forming a mental state the key aspects of which are constituted *in the course* and *for the sake* of speaking (Slobin 1996; cf. Jaszczolt and Witek 2018). In other words, our ability to form discourse-constituted thoughts results from internalising rules, procedures and patterns governing our illocutionary practice.

Recall that in issuing an expert pronouncement, the speaker exercises her current conversational role or illocutionary power of an expert, which is internalised and foregrounded as an aspect of the thought she thereby expresses; what she shows in saying that P, namely, is her belief that P as formed by an expert. Of course, in doing this she can be insincere; in other words, she can fail to believe that P; but even if she does not have this belief, she still presents the proposition that P as the content of her expert opinion. What is more, this aspect of her thought is constituted in the progressing discourse and results from internalising her agreed status of an expert.

Consider, by analogy, making a presumption that *P*. In performing this speech act, the speaker shows her tentative acceptance that *P as* (*i*) formed and expressed to enable the ongoing argument to proceed despite the absence of conclusive or sufficient evidence, and *as* (*ii*) warranted by virtue of her current epistemic and conversational situation. Consider, for instance, an act of presuming A's honesty made in the course of an informal, non-legal argumentative exchange. The act is felicitous—felicitous in that it succeeds in putting the onus of proof on the addressee who would like to question A's honesty (see Corredor 2017; cf. Lewiński 2017)—provided the speaker is adequately related to A, i.e., she has an appropriate acquaintance with him, knows him personally and meets him more or less regularly or at least has heard a lot of good things about him, and so on. In short, the speaker is supposed to have the required illocutionary power or authority[26] that enables her to make felicitous presumptions about A's honesty; it is instructive to note, namely, that this authority can be attacked by an *Ad Hominem*, i.e., by saying to the speaker "You know nothing about A". To cut the long story short: in issuing a presumption that *P*, the speaker signals or expresses—expresses with the speaking-to-thinking direction of influence—her current discourse-constituted thought; what is more, it is difficult for her to be insincere in doing this[27], since showing her tentative acceptance that *P as* (*i*) formed and expressed to enable the ongoing argument to proceed, and *as* (*ii*) warranted by virtue of her current epistemic situation, is not the same mental state as believing that *P*: one can entertain the former without having the latter.

[26] For a more detailed discussion of this idea, see (Witek 2019c, 23-28).

[27] As one of the reviewers rightly points out, this claim needs more support. He or she asks, for instance, what if a juror says, "Let's presume that A is innocent," and then proceeds to act in such a way as to suggest that he intended no such thing. In my view, the juror's act is an abuse of condition Γ.2 in John L. Austin's (1975, 16-18) sense, according to which the speaker who invokes an accepted procedure to perform a certain speech act is supposed to "conduct [herself] subsequently" in accordance with what the procedure specifies. Nevertheless, it takes effect as a binding presumption in that it succeeds in putting the onus of proof on those who would like to reject the opinion that A is innocent. What is more, it expresses the juror's *tentative acceptance* that A is innocent *as* formed for the sake of the current conversational purposes no matter whether she believes that A is innocent or not.

In sum, it is difficult to perform insincere speech acts that express discourse-independent thoughts because of the limits put on by expressive norms. Nevertheless, it seems even more difficult to perform insincere speech acts that express discourse-constituted thoughts, since some of their key aspects exist in virtue of being expressed and are formed *in the course* and *for the sake* of speaking.

4. Conclusions

My aim in this paper has been to examine Green's notion of self-expression and the role it plays in his model of intentional communication. In particular, my focus has been on the expressive dimension of illocutionary acts for which Moorean absurdity is possible. I have proposed a number of ideas and categories and argued that they constitute a natural refinement and extension of the conceptual framework of Green's signalling model of communication.

In Section 2 I have proposed the notion of proto-illocutionary speaker meaning which, together with the corresponding idea of proto-illocutionary communication, seems to be necessary to account for acts that overtly show general rather than force-specific commitments. I have argued, for instance, that in making an overt scowl—no matter whether it is sincere or not—I overtly show my commitment to the proposition that I'm angry, but the commitment I show is not specific to a particular illocutionary force.

In Section 3 I have focused on Green's model of the expressive dimension of speech acts and the role expressive norms play in illocutionary communication. In particular, I have offered a critical discussion of his idea according to which expressive norms enable us to indicate illocutionary forces of our utterances and suggested that the norms in question should be rather viewed as playing a key role in the mechanisms for epistemic vigilance (Sperber *et al.* 2010).

Finally, in Section 4, I have distinguished between (*i*) expressing discourse-independent thoughts and (*ii*) expressing discourse-constituted thoughts or, in other words, between (*i*) expressing with the thinking-to-speaking direction of influence and (*ii*) expressing with the speaking-to-

thinking direction of influence. Following Jaszczolt and Witek (2018), I have defined discourse-constituted thoughts as mental states whose key aspects are formed *in the course* and *for the sake* of speech acts that express them; in other words, discourse-constituted thoughts exist in virtue of being expressed and result from *internalising* certain discourse parameters—*e.g.*, the role of the speaker, the current question under discussion she addresses, and so on—that characterise the speaker's act. I have also argued that the notion of discourse-constituted thoughts, together with the corresponding category of expressing with the speaking-to-thinking direction of influence, enable us to arrive at a more comprehensive account of the expressive dimension of illocutionary acts.

Acknowledgements

I am very grateful to Mitchell S. Green for the stimulating discussions we had on the role of self-expression in speech acts. I would also like to thank Marcin Lewiński and two anonymous reviewers for their insightful comments and suggestions that helped me to improve the first draft of this paper.

Funding

The preparation of this work was supported by the National Science Centre, Poland, through research grant No. 2015/19/B/HS1/03306.

References

Austin, John L. 1975. *How to Do Things with Words*. Oxford: The Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198245537.001.0001

Bach, Kent, and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts*. Cambridge, Mass.: MIT Press.

Bach, Kent. 1994. "Conversational Impliciture." *Mind & Language* 9: 124–62. https://doi.org/10.1111/j.1468-0017.1994.tb00220.x

Bäckström, Stina. 2021. "Must Expression Be Instrumental?" *Organon F* 28 (2): 282–302. https://doi.org/10.31577/orgf.2021.28202

Corredor, Cristina. 2017. "Presumptions in Speech Acts." *Argumentation* 31: 573–89. https://doi.org/10.1007/s10503-017-9425-y

Green, Mitchell. 2007. *Self-Expression*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283781.001.0001

Green, Michell S. 2009. "Speech Acts, the Handicap Principle and the Expression of Psychological States." *Mind & Language* 24 (2): 139–63. https://doi.org/10.1111/j.1468-0017.2008.01357.x

Green, Mitchell S. 2017. "How much Mentality is Needed for Meaning?" In: *The Routledge Handbook of Philosophy of Animal Minds*, edited by Kristin Andrews, and Jacob Beck, 313–23. London: Routledge. https://doi.org/10.4324/9781315742250

Green, Mitchell S. 2019a. "From Signaling and Expression to Conversation and Fiction." *Grazer Philosophische Studien* 96(3): 295–315. https://doi.org/10.1163/18756735-09603002

Green, Mitchell S. 2019b. "Organic Meaning: An Approach to Communication with Minimal Appeal to Minds." In: Further *Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*, edited by Alessandro Capone, Marco Carapezza, and Franco Lo Piparo, 211-228. Springer (Perspectives in Pragmatics). https://doi.org/10.1007/978-3-030-00973-1_12

Jaszczolt, Kasia M. 2016. *Meaning in Linguistic Interaction*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199602469.001.0001

Jaszczolt, Kasia M., and Maciej Witek. 2018. "Expressing the Self. From Types of De Se to Speech-act Types." In: *Expressing the Self. Cultural Diversity & Cognitive Universals*, edited by Minyao Huang & Kasia M. Jaszczolt, 187–222. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198786658.003.0010

Langton, Rae. 2018. "Blocking as Counter-Speech." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 144–64. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0006

Lepore, Ernie, and Matthew Stone. 2015. *Imagination and Convention. Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198717188.001.0001

Lewiński, Marcin. 2017. "Argumentation Theory Without Presumptions." *Argumentation* 31: 591–613. https://doi.org/10.1007/s10503-017-9421-2

Lewiński, Marcin, 2021. "Conclusions of Practical Argument: A Speech Act Analysis". *Organon F* 28 (2): 420–57. https://doi.org/10.31577/orgf.2021.28207

Lewis, David. 2002. *Convention: A Philosophical Study*. Oxford: Blackwell Publishers. https://doi.org/10.1002/9780470693711

Millikan, Ruth G. 1984. *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press.

Millikan, Ruth G. 2004. *Varieties of Meaning*. Cambridge, Mass.: MIT Press. https://doi.org/10.7551/mitpress/7072.001.0001

Millikan, Ruth G. 2005. *Language: A Biological Model*. Oxford: Oxford University Press. https://doi.org/10.1093/0199284768.001.0001

Recanati, Francois. 2004. *Literal Meaning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511615382

Roberts, Craige. 1996. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *OSU Working Papers in Linguistics* 49: 91–136.

Sbisà, Marina. 2002. "Speech Acts in Context." *Language & Communication* 22: 421–436. https://doi.org/10.1016/S0271-5309(02)00018-6

Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Searle, John R. 1979. *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213

Siebel, Mark. 2003. "Illocutionary Acts and Attitude Expression." *Linguistics and Philosophy* 26(3): 351–66. https://doi.org/10.1023/A:1024110814662

Siebel, Mark. 2020. "The Belief View of Assertion." In: *The Oxford Handbook of Assertion*, edited by Sanford Goldberg, 99–118. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190675233.013.4

Slobin, Dan. 1996. "From "thought and language" to "thinking for speaking". In: *Rethinking Linguistic Relativity*, edited by John J. Gumperz, and Stephen C. Levinson, 70–96. Cambridge: Cambridge University Press.

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. 2010. "Epistemic Vigilance." *Mind & Language* 25(4): 359–93. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* 25: 701–21. https://doi.org/10.1023/A:1020867916902

Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199645169.001.0001

Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. Cambridge, Mass., London, England: Harvard University Press. https://doi.org/10.2307/j.ctvjsf4jc

Vygotsky, Lev. 1978. *Mind in Society. The Development of Higher Psychological Processes*. Edited by Michael Cole. Cambridge, Mass.: Harvard University Press.

Walton, Douglas. 2010. "Types of Dialogue and Burden of Proof." In: *Computational Models of Argument*, edited by Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo R. Simari, 13–24. Amsterdam: IOS Press (Frontiers in Artificial Intelligence and Applications, Vol. 216). https://doi.org/10.3233/978-1-60750-619-5-13

Wharton, Tim. 2003. "Natural Pragmatics and Natural Codes." *Mind & Language* 18(5): 447–77. https://doi.org/10.1111/1468-0017.00237

Wharton, Tim. 2012. "Pragmatics and Prosody." In: *The Cambridge Handbook of Pragmatics*, edited by Keith Allan, and Kasia M. Jaszczolt, 567–84. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139022453.031

Wilson, Deirdre, and Dan Sperber. 2012. *Meaning and Relevance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139028370

Wilson, Deirdre, and Tim Wharton. 2006. "Relevance and prosody." *Journal of Pragmatics* 38: 1559–579. https://doi.org/10.1016/j.pragma.2005.04.012

Witek, Maciej. 2009. "Scepticism about Reflexive Intentions Refuted." *Lodz Papers in Pragmatics* 5(1): 69–83. https://doi.org/10.2478/v10016-009-0005-y

Witek, Maciej. 2015a. "An Interactional Account of Illocutionary Practice." *Language Sciences* 47: 43–55. https://doi.org/10.1016/j.langsci.2014.08.003

Witek, Maciej. 2015b. "Linguistic Underdeterminacy: A View from Speech Act Theory." *Journal of Pragmatics* 76: 15–29. https://doi.org/10.1016/j.pragma.2014.11.003

Witek, Maciej. 2015c. "Mechanisms of Illocutionary Games." *Language & Communication* 42:11–22. https://doi.org/10.1016/j.langcom.2015.01.007

Witek, Maciej. 2016. "Accommodation and Convention." *Polish Journal of Philosophy* 10(1): 99–115. https://doi.org/10.5840/pjphil20161016

Witek, Maciej. 2019a. "Accommodation in Linguistic Interaction. On the so-called Triggering Problem." In: *In* Philosophical Insights into Pragmatics, edited by Piotr Stalmaszczyk, 163–92. Berlin & Boston: De Gruyter (Philosophical Analysis, Vol. 79). https://doi.org/10.1515/9783110628937-009

Witek, Maciej. 2019b. "Coordination and Norms in Illocutionary Interaction." In: *Normativity and Variety of Speech Actions*, edited by Maciej Witek and Iwona Witczak-Plisiecka, 66–98. Leiden: Brill (Poznań Studies in the Philosophy of the Sciences and the Humanities, Vol. 112). https://doi.org/10.1163/9789004366527_005

Witek, Maciej. 2019c. "Illocution and Accommodation in the Functioning of Presumptions." *Synthese* https://doi.org/10.1007/s11229-019-02459-4

Witek, Maciej. 2019d. "The Expressive Dimension and Score-changing Function of Speech Acts from the Evolutionist Point of View." *Grazer Philosophische Studien* 96(3): 381–98. https://doi.org/10.1163/18756735-09603008

Organon F 28 (2) 2021: 360–398
https://doi.org/10.31577/orgf.2021.28205
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Limitations of Non-Gricean Approaches to the Evolution of Human Communicative Abilities

Mateusz Włodarczyk*

Received: 28 August 2019 / Revised: 20 September 2020 / Accepted: 12 December 2020

Abstract: In this paper, I examine two non-Gricean approaches to the evolution of human communicative abilities: Mitchell S. Green's account of organic meaning and Dorit Bar-On's account of expressive communication. I argue that the non-Gricean approaches in question face certain problems: i) they focus on the adaptive function of communicative behaviours and ignore questions about their mechanisms, ontogeny and phylogeny; ii) the notion of organic meaning does not constitute an intermediate form between natural and non-natural meaning but should rather be understood as a special case of natural meaning; iii) the non-Gricean approaches under scrutiny cannot explain the transition from dyadic to triadic communication. I also outline the differences between Gricean and non-Gricean approaches and argue against the usefulness of the non-Gricean approaches discussed in this paper in explaining the evolution of human communicative abilities.

Keywords: Evolution of communication; evolution of language; expressive communication; Gricean communication; organic meaning.

* University of Szczecin
https://orcid.org/0000-0002-3658-1842
Institute of Philosophy and Cognitive Science, University of Szczecin ul. Krakowska 71-79, 71-017 Szczecin, Poland.
mateusz.wlodarczyk@usz.edu.pl

1. Introduction

Paul Grice's theory of meaning and his theory of conversational implicatures were designed to provide a rational reconstruction of communicative processes rather than to describe the psychological processes that underlie our ability to communicate (Wilson 2017; Saul 2002). Still, many researchers concerned with both the ontogeny and phylogeny of human communication describe their theories as either Gricean or non-Gricean. One example of the former category is Relevance Theory (Sperber and Wilson 1986), which postulates the Cognitive Principle of Relevance (Wilson 2017) not only to explain human forms of communication, but also to explicate mechanisms that govern cognition *in general*. And while Relevance Theory, as a post-Gricean approach, distances itself quite significantly from Grice's original proposal, there are other approaches (especially in the neo-Gricean camp) which retain more of the Gricean spirit and still try to create plausible explanations of the ontogeny and phylogeny of language (Tomasello 2010, 2014; Thompson 2014; Moore 2016, Fitch 2010; Hurford 2003, 2014). As a consequence of the fact that "being Gricean" is quite a nebulous concept, with many Gricean researchers rejecting certain tenets of Grice's theory, it is useful to at least roughly determine what those approaches have in common. For the purposes of this paper, we can point to two main aspects: firstly, on the original Gricean account, the meaning of a speaker's communicative act is constituted by her intention, and successful communication consists in the speaker's expressing and the hearer's recognizing the meaning-constituting intention behind the act; secondly, the recognition of the speaker's communicative intention is an inferential process guided by the cooperative principle, Gricean maxims, a modified set of maxims in neo-Gricean accounts, or by other general pragmatic principles posited in post-Gricean accounts.[1]

[1] One could argue that the cooperative nature of communication is another core tenet of Gricean accounts. However, according to Relevance Theory, for instance, cooperation on the part of the speaker does not have to be assumed by the hearer in order for him to infer the meaning of a communicative act. This suggests that not all Gricean accounts share the assumption that communication is cooperative in Grice's sense.

Nevertheless, despite its prevalence in both linguistics and cognitive science, the intentional-inferential model that forms the core of Gricean accounts seems to be, at least at first glance, ill-suited to explain how human communication has emerged during evolution or how children learn to communicate. This is due to the fact that it places high cognitive requirements on Gricean communicators: it is assumed that a Gricean Being, which I understand as a being capable of expressing and recovering communicative intentions, must be equipped with the following cognitive abilities (Moore 2016):

1. the ability to understand that others have beliefs and false beliefs,
2. the ability to make complex inferences about others' goal-directed behaviour,
3. the ability to form fourth-level meta-representations.

The assumption that the three abilities in question necessarily underlie Gricean communication leads to what is known as the *cognitive load problem*. We observe that children from one to three years old have the ability to participate in linguistic communication, often with great efficiency; but psychological research reported in Breheny (2006), Joseph and Tager-Fluchberg (1999), and Welmann et al. (2001) suggests that they are not equipped with the above-mentioned abilities. One then has to answer the question of how it is possible for them to participate in linguistic communication without being Gricean Beings. Moreover, Gricean approaches face the challenge of explaining how the ability to form complex communicative intentions emerged in the course of evolution. In light of the observed gap between the communicative abilities of humans and those of our extant evolutionary relatives, the explanation of the emergence of speaker meaning (in Grice's sense) should involve some sort of plausible intermediate stages between non-intentional and intentional communication. However, it is still far from clear how those stages might look.

There are different ways in which we can answer this question, and some of them remain Gricean in spirit, allowing us to maintain the basic tenets of the Gricean intentional-inferential programme.[2] However, some researchers suggest that to solve this problem we have to allow for non-intentional

[2] Attempts to resolve the problem of cognitive load within the Gricean framework go in two general directions: Firstly, by rejecting certain aspects of Grice's theory,

and non-inferential forms of information transfer which constitute a stepping stone from non-Gricean to fully-fledged Gricean communication. It is important to note that in calling these approaches non-Gricean, I do not want to say that they reject the claim that the linguistic communication of adult humans is intentional and inferential. Rather, what I have in mind is that they postulate three general claims: i) at least some forms of communication are observed in other animals (and plants) and ii) at least some forms of human communication may be adequately described and explained without reference to the intentional-inferential model; and iii) an adequate account of non-inferential and non-intentional communication—both human and non-human—may be used as a basis to develop a plausible ontogenetic and phylogenetic explanation of the emergence of intentional-inferential communication.

In this paper, I offer a critical discussion of two theories that can be regarded as non-Gricean in the sense explicated above: Mitchell Green's (2017b; 2019) conception of organic meaning and Dorit Bar-On's (2013, 2017) account of expressive communication. They are both designed to describe and account for certain forms of non-Gricean communication that may be regarded as evolutionary and developmental prerequisites for uniquely human communication. I argue that they fail to provide a satisfactory explanation of the evolution of human communicative systems; specifically, I claim that they face the following three problems: firstly, they conflate questions about functions and questions about causes; secondly, contrary to what Green claims, the examples of organic meaning he describes do not seem to be really distinct from what Grice discussed as cases of natural meaning; thirdly, the two models in question fail to provide a sufficient basis for explaining the transition from dyadic to triadic communication.

The paper is organized into four parts. Section 2 discusses Green's idea of organic meaning as an intermediate form between natural and non-natural meaning. Section 3 offers a brief presentation of Bar-On's model of

we can lower the demands for Gricean Beings (Moore 2016, Bach 1987); secondly, we can argue that children do in fact possess the abilities necessary to be Gricean Beings (Thompson 2014).

expressive communication. Section 4 argues that the two non-Gricean accounts under discussion faces serious problems. Finally, Section 5 outlines and discusses the differences between Gricean and non-Gricean approaches, and argues against the usefulness of the two non-Gricean approaches described in this paper in explaining the evolution of human communicative abilities.

2. Organic meaning

A central idea behind Green's notion of organic meaning is clear from the outset:

> Natural meaning seems too austere, while non-natural meaning seems too psychologically demanding a basis on which to build an account of the development of communication, either in ontogeny or phylogeny; it also makes it difficult to understand how communication is possible among adults who are developmentally compromised. Might there be middle ground? (Green 2019, 213).

Green's goal is then to find forms of meaning that go beyond the original Gricean contrast between natural and non-natural meaning, and, more importantly, that can be used to explain the evolution and development of human communicative abilities without being susceptible to the problem of cognitive load. It is worth noting that if the notion of organic meaning is to be used to explain the phylogeny and ontogeny of communication, it takes us beyond Grice's rational reconstruction and leads us to the realm of psychological reality—tentatively, we can say that it explains how communication systems emerge among organisms that do not reach a certain cognitive threshold (and maybe even among organisms that reach that threshold—more on that later). But how to define a notion of meaning that would constitute an intermediate form between natural and non-natural meaning? As we know, Grice (1957) defined five conditions that must hold for the particular use of "mean" to fall into the natural category:

1. One cannot consistently say, "Those spots mean measles, but he hasn't got measles." That is, "mean" in its natural usage is factive.

2. One cannot argue from "Those spots mean measles," to any conclusion about what is or was meant by those spots.
3. One cannot argue from "Those spots mean measles" to any conclusion about what anyone meant by those spots.
4. One cannot restate the above example in terms that involve direct discourse. That is, one cannot rephrase "Those spots mean measles," by saying, "Those spots meant, 'measles'," or "Those spots meant, 'He has measles'."
5. One can restate "Those spots mean measles," as "The fact that he has spots means that he has measles."

While Grice acknowledged that the division between natural and non-natural meaning may not be exhaustive (Grice 1957, 379), he nevertheless focused his analysis only on those two aspects. But if our task is to find some sort of intermediate form of meaning—as Green (2019) suggests—then we can now define the task as finding forms of meaning that satisfy only some of the conditions specified by Grice.

Green's starting point is the notion of a *cue* defined as follows:

> [a] feature F of the environment is a cue for organism C if C is able to use the information conveyed by F in a way that tends to improve its chances of survival or reproduction. (Green 2019, 214)

As Green points out, cues are not yet cases of communication; however, communication can emerge on the basis of cues; i.e., in situations when "organisms design, manipulate, or otherwise exploit cues in a way that benefits them due to the response that is engendered among the organisms that use those cues" (Green 2019, 214). When both the sender and the receiver benefit from the exchange of information, we can speak of communication. Using the notion of a cue we can now define the concept of a signal:

> S is a signal iff it is a behavioral, physiological, or morphological characteristic fashioned or maintained by natural selection because it serves as a cue to other organisms (Green 2019, 215).

The following two points should be emphasized here. Firstly, it is crucial that in this story there is no mention of intentions or inferences; therefore, communication involving sending and reading signals so defined can be

potentially referred to in an explanation of the emergence of communication among organisms that are clearly unable to form intentions or to understand that others have any kind of mental states (e.g., bacteria). Secondly, the explanation in question would be based solely on the adaptive function of signals understood as characteristic properties of signalling organisms.

Recall that Green's goal is to find forms of meaning that satisfy only some of Grice's five conditions. How can we use the notions of cues and signals defined above to achieve this goal? Let's consider, following Green (2019, 218), a population of colourful frogs from the family *Dendrobatidae*. Frogs belonging to this family are aposematic: the bright coloration serves as a warning for potential predators that the frog wearing it is extremely noxious. Since the coloration of a frog can be treated by a potential predator as a warning of the frog's toxicity, it can be classified as a cue: if a predator, as a receiver, produces behaviour appropriate to the presented cue, it will increase its chances of survival. Moreover, in this case, both the signaller and the receiver benefit from the exchange of information—the frog, of course, in virtue of not being eaten—so the coloration can also be understood as a signal in accordance with the definition presented above.

Green considers the following hypothetical scenario. Let's assume that in a population of colourful toxic frogs, a mutant is born: a colourful but *non-poisonous* individual. Such an individual would benefit from his coloration for free—because producing the toxin is costly in terms of calories—and in consequence would gain a fitness advantage over his non-mutant conspecifics. Following Green (2019, 218), if we analyse this situation in terms of Grice's five conditions, we can see that:

1. One can consistently say, "That bright coloration means that the frog is noxious, but it isn't noxious." Accordingly, 'mean' as used here is not factive.
2. One can argue from, 'That bright coloration means that the frog is noxious,' to a conclusion about what is or was meant by that bright coloration.
3. One cannot argue from, 'That bright coloration means that he is noxious' to any conclusion about what anyone meant by that bright coloration.

4. One cannot restate the above example in terms that involve direct discourse. That is, one cannot rephrase 'That bright coloration means that he is noxious,' by saying, 'That bright coloration means, "noxious"'.
5. One cannot restate 'That bright coloration means that he is noxious,' as 'The fact that he is brightly coloured means that he is noxious.'

Therefore, only two of the five conditions (3 and 4) for natural meaning hold in the mutant frog case under discussion. Green concludes that the form of meaning that arises in such communicative systems, which he dubs organic meaning, is an intermediate stage between natural and non-natural meaning.[3] Since the emergence of communicative systems based on organic meaning does not depend on intentions and inferences, we can observe them among many different species, from plants to primates. As a corollary of this, the examples of organic meaning in use that Green provides are quite numerous: alarm calls of birds and mammals, intonational patterns in speech, facial expressions and pre-Theory of Mind utterances.

3. Expressive communication

While Green's notion of organic meaning enables us to describe a very broad class of communicative systems, Dorit Bar-On (2013, 2017) concentrates on a narrower category of expressive behaviours, examples of which are growls, lip-smacks, facial expressions of anger, fear, and pain; and alarm calls. The hallmark of these behaviours is that they *show* the signaller's state of mind to suitably attuned observers, and those observers directly recognize the expressed state without the need of inferences (Bar-On 2013, 356). They can be understood as signals, as defined in Maynard-Smith and Harper 2003, but are distinct from other types of animal signals due to certain characteristics: they can guide the attention of the receiver to the

[3] One important aspect of communicative systems that are based on organic meaning is that they are stable; that is, they are resistant to cheats. While stability is crucial when we want to explain how communicative systems are established in the course of evolution, it is not connected to the problems that non-Gricean approaches face, as discussed in section 4; therefore, it is not mentioned here.

object of that state (e.g., the predator that caused the emission of the alarm call); they show the quality of the state (e.g., the degree of fear or happiness); and they show that the signaller is disposed to act in a certain way (Bar-On 2013, 356). Expressive signals are units of expressive communication, which is defined as a:

> (…) form of social, intersubjective, world-directed and overt communicative behavior that is naturally designed to enable expressers to show their intentional states of mind to suitably endowed observers, so as to move them to act in certain ways (toward the expresser or the object of her expressed state), in part by foretelling the expresser's impending behavior (Bar-On 2013, 360).[4]

Bar-On also points out that expressive signals can be produced with a certain level of flexibility;[5] however, they are not produced with communicative intentions, nor does their interpretation depend on any kind of inference on the part of the receiver:

> On the expresser's side there's no need for any active desire to cooperate or to share information, or any belief about what it would take to fulfill it. Even where the production of the behavior or some of its aspects are under the voluntary control of the producer, the behavior is not produced *with the intention of affecting others' states of mind*. (…) At the same time, appropriate, active responses to producers' expressive performances can be entirely spontaneous and grounded in simple contagion or other forms of 'resonance'; they needn't be calculated or dependent upon rational assessment of available evidence or inference involving attributions of mental states to others. (Bar-On 2013, 359; italics in the original)

Expressive communication can be understood, then, as a form of communication that is naturally designed; it enables signallers to show their psychological states overtly to receivers and thereby to produce certain

[4] Note that 'intentional' in this definition pertains to the aboutness of the signals (intentionality in Brentano's sense), not to the fact that they are produced with communicative intentions.

[5] On empirical evidence of the flexible nature of alarm calls, see (Crockford et al. 2012).

response on their part. It is also non-intentional and non-inferential; therefore, it can be classified as a form of non-Gricean communication.[6] In fact, Bar-On does not specify any sort of cognitive abilities necessary for participating in expressive communication. As a corollary of this, expressive communication, like communicative transactions employing cases of organic meaning, can be viewed as involving a wide variety of signals: alarm calls of birds and primates, growls and howls of canines, felid hisses, and sounds produced by cetaceans. However, due to their differences from other animal signals, expressive signals are believed to foreshadow human linguistic communication, as they allow animals to "openly share information about their current states of minds and impending behavior, as well as about their environment to suitably responsive others" (Bar-On 2017, 306); i.e., they specify forms of communication that have certain characteristics of human communication without the need of intentional-inferential cognitive architecture.

4. Problems for non-Gricean approaches

In this section, I argue that Green's conception of organic meaning and Bar-On's theory of expressive communication face certain problems that can be quite challenging if we regard the two models under discussion as attempts to explain the phylogeny and ontogeny of human communicative abilities. It is worth pointing out that the source of those problems is what was supposed to be the biggest strength of the proposals under scrutiny: namely, that they promise to explain the emergence of communication systems without appealing to any sorts of processes involving communicative intentions or inferences about those intentions. In the following sections, I develop three challenges to the above-presented non-Gricean approaches, which I briefly mentioned before. Specifically, I argue that: i) the non-Gricean approaches discussed in Sections 2 and 3 above concentrate on the adaptive function of certain forms of behaviour while neglecting the issue of their underlying mechanisms; that is to say, the models offered by Green

[6] Expressive signals according to Bar-On (2017, 306) can be classified as psychologically involved but only in the sense that they reflect and affect the producer's and recipient's current psychological states.

and Bar-On address the question about the function of a communicative behaviour and largely ignore the question about its causes and mechanisms; ii) appearances to the contrary, the examples of organic meaning discussed by Green seem to constitute special cases of Grice's natural meaning and, for this reason, cannot be regarded as constituting an intermediate form of meaning; iii) both approaches, due to their non-mentalistic nature, fail to explain the transition from dyadic to triadic communication.

To avoid any misunderstandings, it is worth stressing that I do not want to argue that the non-Gricean approaches under discussion fail to explain the adaptive functions of certain communicative systems or how those systems might have emerged in the process of evolution. On the contrary, I think that both approaches are very convincing in that matter and explain certain aspects of non-human communication. However, the questions that the Gricean approaches to the evolution of human communication are designed to answer go far beyond the scope of the non-Gricean approaches: the former not only try to explain the adaptive function of our communicative abilities, but also answer the question how, given the intentional-inferential nature of human communication, human agents gained the cognitive architecture that underlies this form of communication, and how much of that architecture we share with our closest evolutionary relatives. If we conceptualize the goal of the Gricean approaches in this way, it becomes clear that non-Gricean and Gricean theories of the evolution of human communicative abilities overlap to a minimal extent. I would also argue that the limited scope of the non-Gricean approaches prevents them from giving substantial insight into either the phylogeny or ontogeny of human communication.

4.1 Tinbergen's four questions

In his classic work, *On aims and methods of Ethology*, published in 1963, Niko Tinbergen introduced four questions that one has to answer in order to fully understand the behaviour of an animal. The questions pertain to:

1. Causation: what immediate effects the external and internal factors have on the occurrence of behaviour (Hogan 2009)?; what are the mechanism of control of the behaviour (Bateson and Laland 2013)?

2. Function: What are the effects of particular behaviour on the organism's fitness (Cuthill 2009)? What is the survival value of behaviour (Shettleworth 2009), in terms of its current utility (Bateson and Laland 2013)?
3. Ontogeny: How does the trait develop in individuals (Nesse 2013)? How do experience and genetic makeup combine to cause the animal to behave as it does (Shettleworth 2009)?
4. Phylogeny: how did a particular behaviour evolve (Shettleworth 2009)? What is the phylogenetic history of the trait (Nesse 2013)?

While these questions do, in fact, inform each other, it is crucial not to conflate them. As Jerry Hogan pointed out, commenting on the independence of the question about causation from the question about function, 'the outcome of behavior can never determine its occurrence' (Hogan 1994, 9). In other words, even if we determine the adaptive value of a certain behaviour, we cannot, solely on that basis, say anything about the mechanism that causes that behaviour. So, natural selection can only be understood as a casual factor in the *historical process of evolution*, not as a proximate cause of cognitive phenotypes (Bolhuis 2009, 173).

It is important to note the relation between Tinbergen's four questions and another distinction made by Baker (1938) and popularized by Mayr (1961), i.e., the distinction between the ultimate and proximate causes of behaviour. According to Mayr (1961, 1503) proximate causes are understood as immediate set of causes of a particular behaviour, for example bird migration; those causes in the case of migration can include the physiological condition of the bird interacting with photoperiodicity and drop in temperature. On the other hand, ultimate causes are "causes that have history and that have been incorporated into the system through many thousands of generations of natural selection" (Mayr 1961, 1503). Hogan and Bolhuis (2009) state that although Tinbergen's causation corresponds quite precisely to the Mayr's proximate causes, the same cannot be said about the correspondence between Tinbergen's function and Mayr's ultimate causes. As Cuthill (2009) says, by function Tinbergen understood the survival value of the behaviour: its *effects* on the fitness of the animal, in term of the current utility of that behaviour (Bateson and Laland 2013); Mayr on the other hand, wanted to use his distinction between ultimate and proximate

causes to draw a line between causes of behaviour that preceded the life of an animal, and those that occur within an animal's lifetime (Dewsbury 1999, 190) Therefore, Mayr's ultimate causes and Tinbergen's functions constitute different categories: since the function of an item or behavioural trait describes its current utility in terms of the effects it has on the fitness of the organism, it cannot be understood as its cause. This is especially clear when we look at the definition of the term 'ultimate causes' offered in Mayr (1993, 94), where ultimate causes are characterized as laws "which cause changes in the DNA of genotypes". Nevertheless, there is a tendency to treat the notions of *ultimate causes* and *functions* as equivalent or even identical to each other (Hogan and Bolhuis 2009). The relationship between the notions introduced by Tinbergen and Mayr is significant from the point of view of the theses presented in the papers by Macphail and Bolhuis (2001) and de Waal (2008) that will be discussed below. The reason is that while Macphail and Bolhuis use the notions of function and cause in their discussion, de Waal on the other hand is referring to ultimate and proximate causes. However, it is worth stressing how de Waal understands those terms. He defines the proximate cause of a piece of behaviour as a "situation that triggers behavior and the mechanism (psychological, neural, physiological) that enables it"; next, he characterizes the ultimate causes of a behaviour as "the benefits an organism or its close kin derive from [it], hence the probable reason why the behavior was favored by natural selection" (de Waal 2008, p. 280). Therefore, his definition of ultimate causes can, in my opinion, be likened to Tinbergen's characterization of functions rather than to Mayr's understanding of the term 'ultimate'.

It is also worth noting that originally none of Tinbergen's questions directly pertained to cognition; de Waal (2008) and Hogan (2009) note that was Tinbergen's goal: to avoid muddling the scientific discussion about behaviour by appealing to poorly understood cognitive processes. However, as Hogan (2009) points out, the study of cognition has come a long way, and nowadays it is noncontroversial to ask questions about cognitive structures that underlie behaviour. Therefore, the study of causes of behaviour could be now understood—among other things—as involving questions about possible cognitive structures that produce observed animal behaviour. As Shettleworth says (2009, 10): "perceptions, representations, decisions as well as

the neural events that accompany them, are all possible proximate causes of behavior"; and again: "In terms of Tinbergen's four questions, cognition is one of the proximate causes of behavior" (Shettleworth 2009, 11) The fact that the questions about cognitive structures are now understood as questions about the causes or mechanism of behaviour is clearly evident in papers by Macphail and Bolhuis (2001) and de Waal (2008) presented below.

Lastly, I would like to point out that there is currently a debate about the status of Tinbergen's four question as well as that of Mayr's proximate-ultimate distinction[7] in ethology and evolutionary biology. For example, Calcott (2013) suggests that concentrating on the proximate-ultimate distinction, one can obscure another perspective of lineage explanations. Hogan (1988) analyses relations between Tinbergen's questions in the context of development, showing that when studying animal behaviour we have to account for complex interactions between phylogenetic, ontogenetic and causal processes. Bateson and Laland (2013) stress: i) the importance of the influence of parental phenotype in development; ii) analysis of different levels of organization at which natural selection can act; iii) and the need for an update of terminology. Additionally, they point out the need for integrative solutions in regards to four questions. Laland et al. (2011) argue that Mayr's distinction rests on incorrect views on development, which in turn has a consequence for our understanding of the role of culture in the evolution of uniquely human traits. However, despite those issues, there is no doubt about the usefulness of both Tinbergen's four question and Mayr's distinction. Laland et al. (2011, 1514) write:

> Mayr's concern that proximate and ultimate explanations should not be regarded as alternatives remains entirely valid today and is an important and useful heuristic that applies broadly across biological disciplines.

A similar opinion can be found in the works of Laland and Bateson (2013) and of Verhulst and Bolhuis (2009). The consensus among the scholars seems to be that while there is still room for discussion and disagreement about certain details, the general approach to the study of animal behaviour

[7] I would like to thank an anonymous reviewer for pointing out this issue.

proposed by Tinbergen is adequate. And certainly, one cannot deny the importance of the distinction between the cause of a behaviour—that is, its underlying mechanism—and its function.

As evident in Macphail and Bolhuis's (2001) critique of the neuroecological approach to cognition, ignoring the differences between questions about functions and questions about causes can lead to severe problems. The neuroecological approach—which, when applied to human cognition, is called 'evolutionary psychology'—assumes that natural selection results in the emergence of optimal solutions to problems faced by a species—so that, in effect, the cognitive capacities of animals could be deduced from our knowledge of what the optimal behavioural patterns would be (Macphail and Bolhuis 2001, 343). For example, as the authors suggest, one of the consequences of such an approach would be the idea that food-storing birds have different memory mechanisms from non-storing birds. But as I have already stated, answering the question about the adaptive function of a behaviour (and the food-storing behaviour can obviously be analysed in terms of its adaptive value) is different from answering the question about the causes—in terms of cognitive structures—of the behaviour under scrutiny. The only way to establish whether different behaviours are caused by different mechanisms is to look at empirical evidence—and as Macphail and Bolhuis (2001) show, there is no compelling evidence for the hypothesis that memory mechanisms of storing and non-storing bird species are qualitatively different. As they say:

> We suggest that (i) functional and evolutionary considerations cannot explain the mechanisms of cognition, but that (ii) functional and evolutionary considerations may provide clues for the analysis of the mechanisms of brain and behaviour. Importantly, as we shall see, even the use of the latter, trimmed down version of 'cognitive ecology' can be misleading to the extent that we should seriously consider the value of such an approach. (Macphail and Bolhuis 2001, 344)

Frans de Waal in his 2008 paper *Putting altruism back into altruism* shows even more clearly the potential issues that can arise when we conflate the questions about the cause and function. Partly due to the fact that biologists have hijacked the terminology used in the discussion about

altruism, ultimate explanations that focus on return benefits are sometimes treated as an explanation on the level of proximate causes of altruistic behaviour. In other words, there is a widespread assumption that animals other than humans behave altruistically only because they expect reciprocity. But this explanation of altruistic behaviour—de Waal argues—is intrinsically flawed because it assumes that altruistic animals have the cognitive ability to fully grasp the complex consequences of their actions in a dynamic social environment, and to predict often substantially delayed return-benefits. A more parsimonious explanation would be to assume that altruistic behaviour is primarily driven—not only in humans, but in other animals as well—by the capacity to be affected by the emotional states of others, to assess the reasons for those states, and to identify with others, adopting their perspectives (de Waal 2008, 281). In short, we could put forth the hypothesis that what drives altruistic behaviour is empathy rather than reciprocity. A detailed discussion of de Waal's proposal of empathy as the proximate mechanism underlying altruistic behaviours goes beyond the scope of the present paper.[8] My main goal is to highlight the potential problems that emerge when we ignore the difference between questions about the adaptive function of a behavioural pattern and questions about its underlying mechanisms. In the case of neuroecological approaches, this leads to the unfounded claim that considerations of the adaptive function of a behaviour provide sufficiently strong evidence for claiming that there exist qualitatively different mechanisms underlying that behaviour. In the case of altruism, the situation is slightly different. Altruistic behaviour—which at first glance shouldn't be promoted by natural selection, since it involves increasing the recipient's fitness *at the cost* of the performer's fitness—can be adaptive if we consider the return benefits that the performer can experience. However, ultimate considerations play no role in the explanation of why certain animals behave altruistically in specific situations; they only show us why certain traits were favoured over the course of natural selection. In order to specify the mechanism (or the proximate cause)—i.e., the environmental, neural and psychological causes of behaviour—we need to look away from adaptive value and to the realms of neurophysiology and psychology.

[8] Interested readers are referred to de Waal (2008); de Waal and Suchak (2010).

My main thesis in this paper is that theories of organic meaning and expressive communication might be in danger of making a similar mistake to that made in the case of i) neuroecological approach to cognition, whose central tenets were scrutinized by Macphail and Bolhuis, and ii) biological approaches to altruism that conflate motivational and evolutionary terminology. Non-Gricean approaches to communication rest on their respective accounts of the adaptive function of communicative behaviours, and they provide rational reconstruction of certain forms of communication;[9] yet, while the accounts in question enable them to explain why certain traits emerged and were maintained as communicative devices, they do not describe or explain the mechanisms underlying the functioning of the traits under scrutiny. Viewed from the perspective adapted by de Waal in his critical discussion of altruistic behaviour, the models offered by Green and Bar-On fail to provide *proximate mechanisms* for the specified forms of communication. As a corollary of this, the claim that the two notions under discussion—i.e., *organic meaning* and *expressive communication*—can be used to describe communicative systems that do not rest on intentional-inferential capabilities is controversial, since even an adequate account of the adaptive value of a piece of behaviour says nothing about its underlying mechanisms and the mental states involved in its production.

Let us look more closely at the account of organic meaning presented in Green (2019). We can say that by defining this notion Green accomplishes two main goals. Firstly, he specifies the notion of a cue in terms of the adaptive value of some feature F of the environment and, next, uses it to characterize the concept of a signal and the idea of signalling communication. The resulting conceptual framework enables him to explain the adaptive value of certain communicative systems. Secondly, by using Grice's criteria for non-natural and natural meaning, Green concludes that organic meaning can be understood as an intermediate form of meaning. Therefore, what he offers is: i) an account of the adaptive value of certain communicative systems; ii) a rational reconstruction of a form of meaning. Recall that Grice's notion of non-natural meaning as well as his analysis of con-

[9] Which is not surprising since both approaches are based on the application of Evolutionary Game Theory to signals (Maynard-Smith and Harper, 2003).

versational implicatures was not intended as a psychological account of human communication (Wilson 2017, Saul 2002). Grice's goals were philosophical: his notion of meaning and his model of conversational implicatures were developed to provide a way of preserving the truth-conditional approach to language while simultaneously showing how philosophers can be attentive to the various ways language is *used* by speakers (Saul 2002; Neale 1992). Therefore, Grice's account does not provide us with cognitive model of utterance production and comprehension[10]. But since Grice's account as such offers no cognitive model of communication, the problem of cognitive load—or any other problem pertaining to the cognitive aspect of communication—simply does not arise at this stage. Similarly, if Green's account of organic meaning is to be understood as a rational reconstruction of a form of meaning, this account cannot be used to provide a *solution* to this problem. The problem of cognitive load, as well as questions about the development of communicative abilities, concern *cognitive mechanisms* that underlie human forms of communication. To answer these questions, one has to specify the plausible cognitive models that underlie the communicative behaviour observed in humans and other animals. Of course, this does not mean that philosophical considerations cannot yield any useful insight into the psychological processes underlying communication. For example, on the basis of Grice's analysis we can predict that communication involving non-natural meaning probably requires quite complex cognitive systems—certainly more complex than communication observed among bacteria. Bermudez (2005) argues that conceptual analysis provided by philosophers can be, in some ways, useful for psychology and cognitive science. Nevertheless, philosophical considerations do not lead to explanations at the, for example, computational, algorithmic or implementation levels specified by Marr (1982).

It is obvious that *some* Gricean approaches—for instance, the account offered by Relevance Theory—are constructed with an explicitly stated goal of providing a psychological model of communication. However, the insights they offer depart significantly from Grice's original analysis of

[10] See, for example, Saul (2002) for analysis of Grice goals, and how Grice's programme differs from more psychologically oriented relevance theorists' approach. Also Neal (1991) provides an excellent summary of Grice's programme.

communication.[11] This is understandable, since their goal is to develop a psychologically plausible model of communication. To achieve this goal, the proponents of the Gricean approach have to go beyond rational reconstruction and try to tell a story about what is happening—paraphrasing Marina Sbisà (2001)—"in the mind" of speakers and hearers. Basically, we have to provide an operationalization of Grice's ideas of communicative intentions and pragmatic inferences. It is important to stress that the problem of cognitive load can indeed arise at this stage, that is, when we move from the realm of philosophical analysis to the realm of modelling cognitive processes. Note, however, that we can try to model broadly Gricean[12] communication in ways that will avoid this problem. For example, as I have already mentioned, we can *either* try to lower the threshold for Gricean communication (Moore 2016) *or* search for empirical evidence suggesting that infants have the complex intention-reading skills required for Gricean communication (Thompson 2014). Considering whether Gricean approaches are successful in answering the cognitive load problem goes beyond the scope of the present paper. For current purposes, it suffices to note that, taking into account their focus on cognitive mechanisms, they seem to offer adequate conceptual resources one can use to address and solve the problem in question. By contrast, the conceptual frameworks of Green's theory of organic meaning and Bar-On's model of expressive communication do not seem to provide an adequate basis for considering the evolution of cognitive mechanisms and skills underlying human communication.

The problems persist even if we acknowledge that the notion of organic meaning gives us a way to look at the adaptive value of certain traits and

[11] Relevance theorists offer a radically different analysis of what is said and what is implicated; they reject Grice's conversational maxims; they also offer a different analysis of speaker's intentions.

[12] I say broadly Gricean here mainly because if we look at the different neo and post-Gricean theories that are available today we sometimes see a very different visions of Gricean communication. In fact, there is a need for discussion about what really constitutes a Gricean theory (see for example Jaszczolt 2019). However, most of the contemporary Gricean approaches maintain the intentional-inferential model of communication that can be understood as a 'cornerstone' of broadly Gricean communication.

forms of behaviour. Forms of communication that involve organic meaning described in Green (2019) do not involve any sort of complex communicative intentions on the part of the speaker or hearer; but from the standpoint of functional analysis there *shouldn't* be any mention of intentions: as the discussion of Tinbergen's four questions suggests, the analysis in terms of *functions* tells us nothing about *mechanisms*. In other words: to ask whether a particular communicative behaviour involves processing of certain mental states and how this processing is implemented is to ask a question about the mechanism of cognition; to ask about the adaptive value of a certain behaviour is to ask a questions about its function. In fact, the analysis of organic meaning in terms of its adaptive value presented in Green (2019) does not enable us to differentiate between systems having intentional-inferential cognitive architecture, on the one hand, and non-inferential and non-intentional systems, on the other[13]. For example, certain communicative systems which Green describes as involving cases of organic meaning can, in fact, rest on an ability to form some sort of communicative intentions; prime candidates would be pre-ToM utterances and animal alarm calls. As Crockford et al. (2012) show, the alarm calls produced by chimpanzees constitute a very flexible form of communication that involves the ability to monitor the attentional state of the receiver; furthermore, as Tomasello's analysis shows, the gestural communication of chimpanzees can be understood as a form of *intentional* communication defined as "communicative signals that are chosen and produced by individual organisms flexibly and strategically for particular social goals, adjusted in various ways for particular circumstances" (Tomasello 2010, 14). This class of intentional signals is a starting point for a communication from a psychological point of view: communication that involves signaller attempts to influence the psychological states of the receiver and which is contrasted with non-intentional communicative displays. While it is non-controversial that, for

[13] To expand on this issue: Sherry (2009), following Tinbergen, points out that behaviour serving the same function in different species can be caused by different mechanism; therefore, without causal analysis we cannot really say anything about similarities in mechanisms involved in the production of certain behaviour in different species even if we assume that the function of the behaviour in question is the same.

example, the communicative system based 'coloration' used by noxious amphibians involves no intentions, the same cannot be said about communication among great apes or pre-linguistic infants, which can be construed as a form of intentional, though not fully Gricean, communication. Following Macphail and Bolhuis we can say that, in the end, we need empirical evidence to say something about cognitive mechanisms involved in particular forms of behaviour. In our case, to decide whether a particular form of communication involves intention-reading skills, we have to study not only the function but also the cognitive mechanisms involved. But on Green's account all of the above mentioned forms of communication—from the communication in bacteria to pre-ToM utterances of infants—are characterized as cases of non-intentional communication that require 'minimal minds', without relying on empirical data. In light of this, it is also unclear how non-Gricean approaches can help explain the phylogeny and, in particular, the ontogeny of communication, since ontogeny deals in large part with the question of how mechanisms that cause certain behavioural patterns develop during an organism's maturation.

Green's approach to studying communicative systems seems to face the same problem as the one identified by de Waal (2008) in his critical discussion of the neuroecological account of altruistic behaviour. Because of the limitations of his conceptual framework, Green does not ask the question about the proximate mechanisms of particular communicative behaviours and, as a result, is unable to address the question about the evolution of their underlying cognitive machinery. I would like to be clear about the nature of my objections here. I think that Green's proposal accomplishes the goals formulated above: it is both a rational reconstruction of a form of meaning, and an analysis of adaptive value of certain communicative systems. However, in light of my objections, I do not think that the accomplishment of these goals can provide us with an answer to the cognitive load problem and provide a basis for phylogenetic and ontogenetic research. The problem here is that Green states that his analysis of organic meaning could be used to tackle cognitive, ontogenetic and phylogenetic problems. For example, he states:

> As a result, organic meaning may be of interest as part of an evolutionarily plausible account of the phylogeny of communication

> and of a psychologically plausible account of the ontogeny of communication. It will also shed light on aspects of adult human communication that do not meet the cognitive demands of speaker meaning. (Green 2019, 212)

Also, as shown in Section 2, Green's notion of organic meaning is proposed as a possible solution to the cognitive load problem. But in order to accomplish that goal, Green's theoretical framework has to be supplemented with an additional account of the mechanisms underlying communicative behaviour as well as with an adequate model of their ontogeny and phylogeny. This is not an easy task. For example, Hogan (1988) offers extensive discussion on the development of behavioural systems, in which he considers interactions between causal and developmental processes that result in changes in those systems, as well as the relation of the functions of a behaviour to its causes. Among others, he discusses development of so-called displays: behavioural patterns adapted to serve as signals to conspecifics (Hogan 1988, 74); Hogan makes a comparison between waltzing, a type of courtship behaviour observed in chickens and 'oblique posture with long call' observed among black-headed gulls. Both types of behaviour are classified as displays. However, in the case of black-headed gulls, experimental data suggest that social experience can be understood as a causal factor in the development of the display behaviour in question, while waltzing is observed even in chickens that were raised in isolation (Hogan 1988, 75); therefore, the development of waltzing does not seem to be mediated by social experience. This shows that even when we compare behaviours that fall into the same category from the functional perspective, the developmental processes (and possible the mechanisms) that lead to the behaviours under scrutiny can be quite different in two separate species. Once again, to provide a comprehensive evolutionary explanation of some forms of behaviour, we have to answer all four of Tinbergen's questions.

The situation is slightly different when it comes to considering the expressive communication model offered by Bar-On. The notion of expressive communication is more robust than the notion of organic meaning and, therefore, it enables us to specify more precisely the forms of communication that fall within its scope. For example, not all cases of organic meaning can be classified as cases of expressive communication. Recall two main

characteristics of expressive communication: i) it expresses an intentional state of the signaller; ii) it is a non-intentional and non-inferential form of communication. Once again, the analysis offered by Bar-On can be understood as an attempt to provide a rational reconstruction of a form of communication. What is more, like Green, Bar-On aims at answering evolutionary questions. However, she does not consider Tinbergen's questions in her analysis; for example, she leaves open the issue of cognitive mechanisms underlying the communicative behaviour. As a consequence, even if we assume that some forms of information transfer can be classified as cases of expressive communication that involve no Gricean intentions, we shall have to deal with two questions: firstly, i) how to differentiate between forms of expressive communication that rest on different cognitive abilities; secondly, ii) how did those forms *transition* into communicative *acts* backed up by communicative intentions. For example, Bar-On suggests in passing that some forms of expressive communication *can* involve the shared attention mechanism. But if this is the case, the question arises as to how shared attention mechanisms evolved; additionally, the mechanisms in question are referred to in some Gricean accounts (Tomasello 2010, 2014; Hurford 2007) of intentional communication. Therefore, in order to account for more advanced forms of expressive communication, Bar-On seems to refer to the same kind of complicated cognitive machinery that is posited by Gricean approaches.

It is worth noting that Sterelny (1995) makes a similar comment on Millikan's (1989) teleological account of representations. He points out that Millikan's analysis is "wholly within the domain of evolution and function" (Sterelny 1995, 252). This approach—which in large part ignores questions about proximate causes—can cause some problems, one of which is that Millikan's model leads to a very liberal view of what constitutes a representational system. As Allen and Hauser (1992) argue, we could say that on Millikan's account even acacias can have representations. On the other hand, while developing his own account of representations, Sterelny takes into account proximate causes, which in turn allows him to distinguish between different levels of complexity among representational systems.

To conclude, Tinbergen's four questions were designed to give a full understanding of animal behaviour, and while they inform each other, they

should not be conflated. Therefore, non-Gricean approaches, which concentrate on adaptive value, cannot provide a complete explanation of communicative behaviour, including its ontogeny and phylogeny. While I understand the need to create more parsimonious accounts of communication, the simplicity of a proposed explanation cannot be the main factor that decides whether to accept it as part of an adequate evolutionary account of human communication.

4.2 Organic meaning as a case of natural meaning

Green's goal is to find a form of meaning that constitutes an intermediate step between natural meaning and non-natural meaning. One candidate for this intermediate form is organic meaning which, according to Green, satisfies only two of the five conditions for natural meaning. Arguably, the most important claim is that organic meaning is not factive. That means we can utter the sentence:

(1) That bright coloration means that the frog is noxious, but it isn't noxious.

without producing a contradiction. Let's compare that with the paradigmatic case of natural meaning taken from Grice (1957):

(2) Those spots mean measles, but he hasn't got measles.

At first glance, (2) seems to involve contradiction. What is more, the sense of contradiction can be explained by saying that natural meaning is factive. This supports the hypothesis that organic meaning constitutes a distinct category. However, we need to remember that in the case of organic meaning in (1), we have a story—a reason to treat the use of 'mean' as non-factive. The story is that the frog in question is a mutant—it lacks a certain important quality that is typical for its species—and cannot produce a toxin. The existence of such mutants is purely hypothetical, to my knowledge. In standard conditions, and for the overwhelming amount of frogs, we cannot utter (1) without a contradiction because things in the world are a certain way: frogs that have this coloration *are* noxious. However, the hypothetical nature of this scenario is not a sufficient reason to assume that it cannot give us important insight into the nature of meaning.

Consider, by analogy, what would happen if we applied the same treatment to the standard case of natural meaning presented in (2). Let us assume that John and Jacob are representatives of two different tribes which, over the course of hundreds of years of coexistence, developed a curious ritual. Once a year, there is an outbreak of measles among the members of John's tribe (let's call them Johnathanians). Johnathanians know that members of Jacob's tribe (let's call them Jacobians) are mortally terrified of measles because the disease is especially dangerous to them. In fact, they go to great lengths to avoid any contact with any individuals showing symptoms of the disease. However, Johnathanians do not experience most of the symptoms of the measles, excluding the rash, and, being clever, use this opportunity to send their measles-stricken members to steal food from the Jacobians. Jacobians also benefit from this situation, because Johnathanians, being fierce warriors, prevent other, more aggressive tribes from taking the Jacobians' territory. From the definitions given by Green, we can say that the spots, being characteristic of measles, are cues for Jacobians: they use them to enhance their chances of survival and reproduction. Moreover, since Johnathanians also benefit from this exchange of information, we may define this as a case of communication. Finally, natural immunity to the other symptoms of measles can be understood as an adaptation for Johnathanians; therefore, it is a signal. Now, let's assume that John of the Johnathanians is special—he has a certain type of allergy that is triggered by strawberries. The allergy results in a rash that is the same as the rash of a person suffering from measles. John can, therefore, exploit his allergy and steal food from the Jacobians whenever he wants. In a situation like this, we could say:

(3) John's spots mean measles, but he hasn't got measles.

without contradiction.[14] This situation is very reminiscent of the situation with the mutant frog: In a hypothetical scenario, we can present a situation when some form of meaning which in normal conditions is factive, becomes

[14] Note that it is not necessary for either Jacobians or Johnathanians to be intentional creatures (by possessing some form of communicative intentions sensu Grice, or by being able to communicate intentionally sensu Tomasello) for this example to work.

non-factive. This suggests that organic meaning is not distinct from natural meaning. Rather, we could understand what Green calls organic meaning as cases of natural meaning that are *used* in communicative systems. Therefore, contrary to what Green claims, organic meaning can hardly be regarded as an intermediate form of meaning.[15]

4.3 From dyadic to triadic communication

One of the hallmarks of human communication is that it is primarily triadic: different communicative and linguistic devices, such as gestures (pointing and pantomiming), words, and utterances, direct the attention of others to outside entities or events (Tomasello 2010, Hurford 2007). Triadic communication appears very early in human ontogeny in the form of proto-declarative gestures. Firstly, children around 9-to-12 months of age start to show adults interesting objects with an accompanying look and smile; secondly, they point to more distal objects with the same look and smile (Carpenter and Call 2013). It is assumed that the motivation for this type of gesture is not a desire to obtain the object or to elicit a positive emotional response on the part of the adult, but to share attention with the other person. Experimental results support this hypothesis (Carpenter et al. 1998; Moore and D'Entremont 2001; Bates, Camaioni and Volterra 1975), as the response that children expect from adults is to simply acknowledge the presence of the object along with a positive emotional response (e.g., a smile). It is crucial that in these communicative situations, the signaller is actively monitoring the receiver's perceptual states (gaze monitoring) and responses, with additional attempts undertaken in case of communicative failure. As Hurford (2007, 46) states, shared attention seems to be a necessary prerequisite for triadic communication.

Most of the communicative behaviour of animals is dyadic, in the sense that it does not involve entities outside the signaller and receiver. Even the remarkably complex forms of communication that we observe among primates—for example, attention-getting and intention-movements—are

[15] I'm aware that Green treats organic meaning and natural meaning as partially overlapping categories (Green 2017). What I'm trying to say is that all cases of organic meaning are "natural-meaning-in-disguise."

primarily used to regulate dyadic social interactions. However, some forms of communication systems observed among animals can be treated as candidates for triadic communication; for example, the alarm calls of birds and mammals. The question is, what sort of cognitive mechanisms underlie these communicative behaviours? Tomasello (2008) argues that most alarm calls seem to be genetically fixed adaptive specializations over which the signaller has very little voluntary control, and which are strongly tied to the emotional states of the animal. Therefore, they do not display crucial features of human triadic communication, and cannot be understood as an evolutionary basis for intentional-inferential communication.

We need to note two points here. Firstly, Tomasello does not claim that there are no cases of triadic communication among primates, but rather that alarm calls, which indeed draw the attention of the receiver to some kind of outside entity (and with that show a certain degree of triadicity), can only be described as *functionally* referential, as they are not produced with the *intention* to draw the attention of the receiver. Thus, they are distinct from the proto-declarative gestures of children, which are produced flexibly and with the intention to share attention with others. It is not hard to create a plausible story of how functionally referential alarm calls could emerge if we use the framework of Green's organic meaning and treat them as a sort of phylogenetic ritualization which does not require communicative intentions. However, Tomasello (2008, 28) points out the existence of attention-getting gestures that chimpanzees use flexibly to draw the attention of conspecifics to found food. This is a form of proto-triadic communication that is intentional and built upon the cognitive architecture of individual intentionality (Tomasello, 2014) in which the communicator expresses his 'referential intention'. Therefore, in the Gricean approach proposed by Tomasello, there is a place for an intermediate (proto-triadic *and* intentional) form of communication that falls between mere displays, on the one hand, and fully fledged Gricean communication, on the other. Secondly, one could argue that Tomasello makes a mistake classifying the alarm calls of chimpanzees as a merely fixed adaptive specialization. As Crockford et al. (2012) suggest, these alarm calls do exhibit features of flexible communication, as they can be modulated with respect to the attentional state of the recipient. However, this does not pose a significant problem for the Gricean approach

proposed by Tomasello, as one can simply argue that those communicative behaviours rest on the same cognitive architecture which underlies proto-triadic attention-getting.

Human triadic communication seems to rests on shared intentionality, which includes shared attention and shared goals of communicating agents. Utilizing Gricean approaches, we can differentiate between non-flexible alarm calls of birds and monkeys, proto-triadic alarm calls and gestures of chimpanzees, and fully fledged Gricean communication of humans by specifying the cognitive abilities that underlie those forms of communication. The story of the evolutionary transition between dyadic and triadic communication is not only a story of adaptive function or the character of the communication (whether it involves entities outside the signaller and the receiver), but also a story about the cognitive mechanisms that produce certain communicative behaviours. And once again, non-Gricean accounts, by omitting the mechanisms underlying communication, cannot paint a full picture of this transition.

5. What is the goal of Gricean approaches to the evolution of communication?

Let me preface this section by once again emphasizing that the notions of organic meaning and expressive communication are well constructed and useful theoretical categories. As I showed in section 4.1., both Green (2019) notion of organic meaning and Bar-On (2013, 2017) notion of expressive communication offer primarily functional analysis. However, I have doubts that they can be used as an alternative to Gricean theories or that they can solve problems that Gricean approaches face. In the previous sections, I have shown that non-Gricean approaches attempt to avoid these problems (e.g., the problem of cognitive load) by concentrating on forms of communication that require minimal—or even non-existent—cognitive abilities and claim that the forms in question provide a sufficient basis on which more complex, maybe even intentional-inferential forms of communication, can build. Therefore, this strategy can be characterized as bottom-up. That is to say, its proponents attempt to account for the evolution

of communication by demonstrating how evolutionarily stable patterns of communication might have emerged on the basis of adaptive values of communicative behaviour. Notice that the notions of expressive behaviour and, especially, organic meaning cover a very wide range of forms of communication. For example, organic meaning can be used to describe communication between bacteria, on the one hand, and pre-ToM utterances on the other, with no clear way to differentiate between them. I have argued that the non-Gricean approaches under discussion face certain serious problems. Proponents of Gricean approaches, by contrast, can adopt a different, top-down strategy. The starting point is an analysis of acts of fully developed human intentional-inferential communication, with a primary focus on the differences and similarities between those forms of communication and the communicative acts of our closest extant relatives. They can categorize different forms of communicative behaviour exhibited by the species in question, hypothesize on the basis of observational and experimental data about possible cognitive mechanisms that underlie those forms of communication (questions about mechanism), how those mechanisms mature in the lifetime of the organism (questions about ontogeny), and what are the evolutionary paths of those features (questions about phylogeny). Of course, Gricean-oriented researchers will also be interested in the adaptive value of different communicative behaviours (questions about function). What is important is that in doing so, one can distinguish different forms of communication underpinned by different cognitive mechanisms. Each and every one of those forms has to be analysed with respect to the four Tinbergen questions. Moreover, one has to have a way to differentiate among the various forms of communication and their underlying cognitive mechanisms.

Let's now consider Tinbergen's questions from the perspective of a more Gricean-oriented account of communication, namely the account of evolution of human communication offered by Tomasello (2010, 2014, 2019). He adopts a broadly Gricean approach to communication, heavily influenced by Relevance Theory. He assumes that, in the process of communication, speakers and receivers have to mutually rely on their abilities to form and read communicative intentions—in other words, the basis of human cognition and communication is an array of complex mind-reading skills. In order to explain how the conventional forms of human communication emerged

in the course of evolution, Tomasello's account distinguishes at least three different cognitive mechanisms responsible for different forms of cognition and communication: i) individual intentionality, responsible for cognitive and communicative abilities observed in the species of the family *Hominidae* and especially in the genus *Pongo*; ii) joint intentionality, hypothesized to appear for the first time in the species *Homo Heidelbergensis*; iii) collective intentionality, which we observe in modern humans. Levels of intentionality are distinguished by reference to different abilities to i) represent physical and mental aspects of the environment, ii) use those representations in cognitive simulations, and iii) self-monitor the agent's own behaviour, with each level of intentionality giving the agent different skills to represent, simulate and self-monitor, and allowing for different forms of communication. For example, individual intentionality allows great apes to represent the goals and attentional states of other agents and also to perform acts of intentional communication (Tomasello 2010). What is important, the analysis offered by Tomasello provides answers to Tinbergen's four questions; secondly, it identifies clear intermediate stages between intentional communication of primates that rests on individual intentionality and intentional communication of humans that rests on the more complex mind-reading skills of shared and collective intentionality, which can help to resolve the cognitive load problem. I do not want to suggest that Tomasello's approach is the right one to account for the evolution of uniquely human cognitive and communicative abilities. However, it undoubtedly offers functional (Tomasello 2008, 2014; Tomasello et al. 2012), causal (Tomasello 2010, 2014; Tomasello et al. 2005), phylogenetic (Tomasello 2014; Tomasello et al. 2012) and ontogenetic (Tomasello 2010, 2019) considerations of human—and more broadly, primate—cognition and communication, as well as providing a basis for solving the cognitive load problem by positing intermediate phylogenetic and ontogenetic stages. Theoretically, forms of communication underpinned by the different levels of intentionality described above could be classified as cases of organic meaning; but if we treat them as such, we leave open the task of explaining i) the transition between them as well as ii) the transition between non-intentional and intentional forms of communication. And, as I previously said, while the notion of expressive communication specifies a narrower class of communicative behaviours, it still does not

account for the cognitive architecture that underlies them, giving rise to the same set of unsolved problems.

To summarize, the main problem that theories of the evolution of human communicative abilities face is the problem of the transition from non-intentional, non-inferential forms of communication to Gricean intentional-inferential communication, which involves the question of the mechanisms of those forms of communication and their intermediate stages. Gricean accounts *can* offer the solution because they address and answer questions about mechanisms, ontogeny and phylogeny. Meanwhile, non-Gricean approaches that concentrate on the adaptive value of communicative behaviour cannot solve this problem, and thus cannot be seen as an alternative to Gricean approaches, simply because they do not address the central questions to be answered.

I would also like to comment on Bar-On's (2017) claims that Gricean accounts are characterized by a sharp dichotomy between code and ostensive-inferential communication, which is problematic if we want to construct plausible evolutionary explanations for human communicative abilities, and that expressive communication, as an intermediate form, can resolve those problems. Indeed, as she points out, on Tomasello's approach—which is definitely Gricean—there is a sharp distinction between animal vocal signals and acts of intentional communication. Bar-On then states:

> On the non-Gricean EC approach advocated here, however, expressive behavior, and the kind of communication it affords, form a theoretically significant category of behavior that lies somewhere between the two endpoints Tomasello describes—i.e., merely reflexive-reactive affective displays and fully reflective-creative intentional utterances (Bar-On 2017, 360).

I would argue that it is not correct to characterize Tomasello's approach as exhibiting this kind of dichotomy. On the contrary, he explicitly indicates a specific form of communication—the gestural communication of primates—as an intermediate form between non-Gricean and fully Gricean communication. However, that form of communication *is intentional*, as defined in section 4.1., and is based on the cognitive abilities of primates that allow them to understand the causal and intentional relations that occur in the world and use that knowledge as a basis for thoughtful and

controlled actions (including communication): individual intentionality. Therefore, it is well within the scope of Gricean approaches to conceptualize the forms of communication that fall between non-intentional signals and intentional-reflective utterances—in other words, Gricean approaches allow for *degrees* of intentional communication.

If we assume, that human communication is built upon complex mind-reading skills, which, I think, is not a controversial view, then we will always have the problem of explaining the transition between forms of communication that pose little or no requirements for cognition and uniquely human intentional communication. Even if we assume that the notion of organic meaning accurately describes the communication between amphibians and predators that threaten them, then we will have to explain how non-intentional organic meaning transitioned to the intentional, yet not fully Gricean, alarm calls of chimpanzees; if we assume that alarm calls are cases of non-intentional organic meaning, we will have to explain how it transitioned to the pre-ToM utterances of infants; if we assume that pre-ToM utterances are cases of organic meaning—still, we will have to explain the transition between non-intentional pre-ToM utterances and intentional communication. And every time that we have this significant transition between non-intentional and intentional communication, the problem of cognitive load arises. And one of the reasons for that is the fact that we do not specify any cognitive mechanisms responsible for those supposedly non-intentional forms of communication, thus creating a 'cognitive gap'. I suggest that to avoid this problem one can i) pose intermediate stages of intentional communication and ii) develop models of Gricean communication that posit lower cognitive demands on communicating agents[16]. As mentioned before, Bar-On claims that expressive communication foreshadows human communication in virtue of certain characteristics that go beyond simple animal signals, but do not require the capacity to form and understand

[16] Of course, there is another option: we can reject the Gricean models of communication in favour of alternative views; in this case, the intentional-inferential model will no longer be the 'end game' of communication that requires explanation in evolutionary and developmental terms. However, since the non-Gricean accounts described in this paper do not make the claim that communication of mature humans is non-Gricean, I did not consider this approach.

communicative intentions. However, those characteristics—the open nature of expressive communication in which the signaller overtly conveys information about his mental states, impending behaviour and environment—do not constitute Gricean communication. There is still a very apparent difference—a gap—between expressive communication and organic meaning on the one hand, and Gricean communication on the other. Therefore, if we want to use non-Gricean approaches to explain the emergence of human communication, we need to explain how expressive communication, or cases of organic meaning, can *transition* into the intentional Gricean communication. Without this, the question remains unanswered. One apparent solution to this problem is to postulate some sort of intermediate levels of *intentionality*, which can underlie 'minimal-Gricean' acts of communication—and this is precisely the strategy that some Gricean approaches have adopted. Bar-On seems to be aware of that: "Once appropriated, and caught up in intentional actions, expressive signals can propagate and stabilize, and come to have a semantic-pragmatic life of their own" (Bar-On 2017, 310); and, "Expressive communication is at times triadic, relying on shared attention mechanisms that allow signalers and receivers to attend together to objects or events of mutual concern" (Bar-On 2017, 310). But the only story that non-Gricean approaches offer is a rational reconstruction of some forms of non-intentional communication and a story about how certain communicative behaviours are established due to their adaptive functions. Therefore, to bridge the gap, non-Gricean approaches have to fall back on concepts that are utilized in Gricean approaches—e.g., shared intentionality—for which they do not offer any evolutionary explanation. In other words, to provide an explanation for the emergence of Gricean communication, we need Gricean answers. For example, if we accept the expressive communication hypothesis, then we have to explain how the shared attention mechanism that supposedly allows for triadic communication has evolved and how it develops. The notion of expressive communication alone does not provide an answer to this; however if we have the story in place of how shared attention evolved, what explanatory role does expressive communication play?

6. Summary

In this paper, I have tried to identify the limitations of two non-Gricean approaches to the evolution of human communicative abilities: Green's theory of organic meaning and Bar-On's model of expressive communications. I have argued that what the two approaches in question offer does not go beyond the rational reconstruction of some forms of communication and the functional analysis of verbal forms of communication. And while this theoretical proposal could be useful, it falls short of the stated goals of providing solutions to cognitive, ontogenetic and phylogenetic problems that arise when we try to construct a plausible account of the evolution of human communication. Tinbergen's four questions show that causal, functional, ontogenetic and phylogenetic considerations in regard to behaviour require different approaches. As Hogan (1994) points out, questions about functions should be clearly distinguished from questions about mechanisms; as de Waal (2008) and Macphail and Bolhuis (2001) show, conflating these two questions can lead to serious problems. Functional analyses are important, but issues like the problem of cognitive load arise when we look at the cognitive processes that underlie our communicative abilities; therefore, functional analyses cannot be treated as solution to those kinds of problems. Furthermore, neglecting the causal analyses can lead to problems in explaining certain characteristics of human (and primate) communication, such as its triadicity. Finally, I have argued that organic meaning could be better understood as a natural meaning in use rather than a distinct category of meaning.

The main problem of the two non-Gricean approaches discussed in this paper seems to be that in order to fill the evolutionary gap between non-intentional and intentional communication they try to reconstruct forms of non-intentional communication in greater detail, while still maintaining that they do not involve any sort of communicative intentions. While this approach may have its merits, it does not really help to fill this gap since we still have to explain the transition between those forms of communication and the intentional communication of humans. On the other hand, Gricean approaches try to resolve the same problem by showing that various cases of intentional communication, broadly understood, can be more

or less cognitively demanding and, as the corollary of this, create a continuum of forms of intentional communication that can provide an adequate basis for plausible ontogenetic and phylogenetic accounts. The analysis of possible cognitive mechanisms involved in communication constitutes a large part of—at least some—Gricean approaches. For this reason, they seem to offer a more adequate conceptual framework within which one can address and attempt to solve the problem of cognitive load.

Acknowledgement

I would like to thank Mitchell Green, Paolo Labinaz, Max Urchs and Maciej Witek for discussion and comments that greatly improved earlier drafts of this manuscript. I would also like to thank Martin Hinton for carefully proofreading my manuscript.

Funding

The preparation of this work is supported by the National Science Centre, Poland, through research Grant No. 2015/19/B/HS1/03306.

References

Allen, Colin, and Marc Hauser. 1992. "Communication and Cognition: Is information the connection?" *Proceedings of the Philosophy of Science Association* 2: 81–91. https://doi.org/10.1086/psaprocbienmeetp.1992.2.192826

Bach, Kent. 1987. "On Communicative Intentions: A Reply to Recanati." *Mind & Language* 2 (2): 141–154. https://doi.org/10.1111/j.1468-0017.1987.tb00112.x

Baker, John, R. 1938. "The Evolution of Breeding Seasons." In *Evolution: Essays on Aspects of Evolutionary Biology*, edited by G. R. de Beer, 161–77. Oxford: Oxford University Press.

Bar-on, Dorit. 2013. "Origins of Meaning: Must we 'Go Gricean'?" *Mind & Language* 28 (3): 342–375. https://doi.org/10.1111/mila.12021

Bar-On, Dorit. 2017. "Communicative Intentions, Expressive Communication, and Origins of Meaning." In *The Routledge Handbook of Philosophy of Animal Minds*, edited by Kristin Andrews and Jacob Beck, 301–12. London and New York: Routledge.

Bates, Elizabeth, Luigia Camaioni, and Virginia Volterra. 1975. "The Acquisition of Performatives Prior to Speech." *Merrill-Palmer Quarterly* 21: 205–224.

Bateson, Patrick, and Kevin N. Laland. 2013. "Tinbergen's Four Questions: An Appreciation and an Update." *Trends in ecology & evolution* 28 (12): 712–18. https://doi.org/10.1016/j.tree.2013.09.013

Bermúdez, José L. 2005. *Philosophy of Psychology: A Contemporary Introduction*. New York and London: Routledge.

Bolhuis, Johan. 2009. "Function and Mechanism in Neuroecology: Looking for Clues." In *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology,* edited by Simon Verhulst and Johan Bolhuis, 163–227. Cambridge University Press. https://doi.org/10.1163/157075605774840987

Bolhuis, Johan, and Simon Verhulst, eds. 2009. *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511619991

Breheny, Richard. 2006. "Communication and Folk Psychology." *Mind & Language* 21 (1): 74–107. https://doi.org/10.1111/j.1468-0017.2006.00307.x

Calcott, Brett. 2013. "Why the Proximate-Ultimate Distinction is Misleading, and Why It Matters for Understanding the Evolution of Cooperation." In *Cooperation and its evolution*, edited by Kim Sterelny, Richard Joyce, Brett Calcott and Ben Fraser, 249–64. Cambridge: MIT Press.

Carpenter, Malinda, and Josep Call. 2013. "How Joint Is the Joint Attention of Apes and Human Infants." *Agency and Joint Attention*, 49–61. https://10.1093/acprof:oso/9780199988341.003.0003

Carpenter, Malinda, Katherine Nagell, Michael Tomasello, George Butterworth and Chris Moore. 1998. "Social Cognition, Joint Attention, and Communicative Competence From 9 to 15 Months of Age." *Monographs of the Society for Research in Child Development*, 4(63). https://doi.org/10.2307/1166214

Crockford, Catherine, Roman Wittig, Roger Mundry, Klaus Zuberbühler. 2012. "Wild Chimpanzees Inform Ignorant Group Members of Danger." *Current Biology* 22 (2): 142–146. https://doi.org/10.1016/j.cub.2011.11.053

Cuthill, Ian. 2013. "The study of Function in Behavioral Ecology." In *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology,* edited by Simon Verhulst and Johan Bolhuis, 163–227. Cambridge University Press. https://doi.org/10.1163/157075605774840987

de Waal, Frans B. 2008. "Putting the Altruism Back into Altruism: The Evolution of Empathy." *Annual Revie of Psychology* 59: 279–300. https://doi.org/10.1146/annurev.psych.59.103006.093625

de Waal, Frans B., and Malini Suchak. 2010. "Prosocial Primates: Selfish and Unselfish Motivations." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365 (1553), 2711–722. https://doi.org/10.1098/rstb.2010.0119

Dewsbury, Donald A. 1999. "The Proximate and the Ultimate: Past, Present, and Future." *Behavioural Processes* 46 (3): 189–99. https://doi.org/10.1016/S0376-6357(99)00035-2

Fitch, Tecumseh. 2010. *The Evolution of Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Francis, Richard C. 1990. "Causes, Proximate and Ultimate." *Biology and Philosophy* 5: 401–415. https://doi.org/10.1007/BF02207379

Green, Mitchell. 2017. "How Much Mentality is Needed for Meaning?" In *The Routledge Handbook of Philosophy of Animal Minds*, edited by Kristin Andrews and Jacob Beck, 313–23. London and New York: Routledge.

Green, Mitchell. 2019. "Organic Meaning: An Approach to Communication with Minimal Appeal to Minds." In *Further Advances in Pragmatics and Philosophy: Part 2 Theories and Applications*, edited by Alessandro Capone, Marco Carapezza and Franco Lo Piparo, 211–28. Springer, Cham. https://doi.org/10.1007/978-3-030-00973-1_12

Grice, Paul. 1957. "Meaning." *The Philosophical Review* 66 (3): 377–88. https://doi.org/10.2307/2182440

Hogan, Jerry A. 1988. "Cause and Function in the Development of Behavior Systems." *Handbook of Behavioral Neurobiology (vol. 9). Developmental psychobiology and Behavioral Ecology*, edited by Elliott M. Blass, 63–106. Boston, Mass.: Springer.

Hogan, Jerry A. 1994. "The Concept of Cause in the Study of Behavior." In *Causal Mechanisms of Behavioral Development*, edited by J. A. Hogan and J. J. Bolhuis, 3–15. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511565120

Hogan, Jerry A. 2009. "Causation: The Study of Behavioral Mechanisms." In *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology,* edited by Simon Verhulst and Johan Bolhuis, 35–53. Cambridge: Cambridge University Press. https://doi.org/10.1163/157075605774840987

Hogan, Jerry A., Bolhuis, Johan J. (2009). Tinbergen's Four Questions and Contemporary Behavioral Biology. In *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology,* edited by Simon Verhulst and Johan Bolhuis, 25–34. Cambridge: Cambridge University Press. https://doi.org/10.1163/157075605774840987

Hurford, James R. 2007. *Language in the Light of Evolution: The Origins of Meaning*. Oxford: Oxford University Press.

Hurford, James R. 2014. *Origins of Language: A Slim Guide*. Oxford: Oxford University Press.

Jaszczolt, Katarzyna M. 2019. "Rethinking Being Gricean: New Challenges for Metapragmatics." *Journal of Pragmatics* 145: 15–24. https://doi.org/10.1016/j.pragma.2019.01.024

Joseph, Robert M, and Helen Tager-Flusberg, H. 1999. "Preschool Children's Understanding of the Desire and Knowledge Constraints on Intended Action." *British Journal of Developmental Psychology* 17 (2): 221–43. https://doi.org/10.1348/026151099165249

Laland, Kevin N., Kim Sterelny, John Odling-Smee, William Hoppitt and Tobias Uller. 2011. "Cause and Effect in Biology Revisited: is Mayr's Proximate-ultimate Dichotomy Still Useful?." *Science* 334 (6062): 1512–16. https://doi.org/1126/science.1210879

Magphail, Euan M., and Bolhuis John J. 2001. "The Evolution of Intelligence: Adaptive Specializations Versus General Process." *Biological Reviews* 76 (3): 341–64. https://doi.org/10.1017/S146479310100570X

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York: W. H. Freeman and Company.

Maynard-Smith, John M, and David Harper. 2003. *Animal Signals*. Oxford and New York: Oxford University Press.

Mayr, Ernst. 1961. "Cause and Effect in Biology." *Science* 134 (3489): 1501–06. https://doi.org/10.1126/science.134.3489.1501

Mayr, Ernst. 1993. "Proximate and Ultimate Causations." *Biology and Philosophy* 8: 93–4. https://doi.org/10.1007/BF00868508

Millikan, Ruth G. 1989. "Biosemantics." *The Journal of Philosophy* 86(6): 281–97. https://doi.org/10.2307/2027123

Moore, Chris, and Barbara D'Entremont. 2001. "Developmental Changes in Pointing as a Function of Attentional Focus." *Journal of Cognition and Development* 2 (2): 109–29. https://doi.org/10.1207/S15327647JCD0202_1

Moore, Richard. 2016. "Gricean Communication and Cognitive Development." *The Philosophical Quarterly* 67 (267): 303–26. https://doi.org/10.1093/pq/pqw049

Neale, Stephen. 1992. "Paul Grice and the Philosophy of Language." *Linguistics and Philosophy* 15(5): 509–59. https://doi.org/10.1007/BF00630629

Nesse, Randolph M. 2013. "Tinbergen's Four Questions, Organized: A Response to Bateson and Laland." *Trends in Ecology & Evolution* 28 (12): 681–82. https://doi.org/10.1016/j.tree.2013.10.008

Saul, Jennifer. M. 2002. "What Is Said and Psychological Reality; Grice's Project and Relevance Theorists' Criticisms." *Linguistics and Philosophy* 25: 347–72. https://doi.org/10.1023/A:1015221313887

Sbisà, Marina. 2001. "Intentions from the Other Side." *Paul Grice's Heritage*, edited by Giovanna Cosenza, 185–206. Turnhout: Brepols.

Sherry, David F. 2009. "Do Ideas About Function Help in the Study of Causation?" In *Tinbergen's Legacy. Function and Mechanism in Behavioral Biology,* edited by Simon Verhulst and Johan Bolhuis, 147–62. Cambridge University Press. https://doi.org/10.1163/157075605774840950

Shettleworth, Sarah J. 2009. *Cognition, Evolution, and Behavior*. New York: Oxford University Press.

Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Cambridge, Mass.: Harvard University Press.

Sterelny, Kim. 1995. "Basic Minds." *Philosophical Perspectives* 9: 251–70. https://doi.org/10.2307/2214221

Thompson, Robert J. 2014. "Meaning and Mindreading." *Mind & Language* 29 (2): 167–200. https://doi.org/10.1111/mila.12046

Tinbergen, Niko. 1963. "On Aims and Methods of Ethology." *Zeitschrift für tierpsychologie* 20 (4): 410–33. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Tomasello, Michael, Alicia P. Melis, Claudio Tennie, Emily Wyman and Esther Herrmann. 2012. "Two Key Steps in the Evolution of Human Cooperation: The Interdependence Hypothesis." *Current Anthropology* 53 (6): 673–92. https://doi.org/10.1086/668207

Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne and Henrike Moll. 2005. "Understanding and Sharing Intentions: The Origins of Cultural Cognition." *Behavioral and Brain Sciences* 28 (5): 675–91. https://doi.org/10.1017/S0140525X05000129

Tomasello, Michael. 2010. *Origins of Human Communication*. Cambridge, Mass.: Bradford Books/MIT Press.

Tomasello, Michael. 2011. "Human Culture in Evolutionary Perspective." In *Advances in Culture and Psychology: Vol. 1. Advances in Culture and Psychology,* edited by Michele J. Gelfand, Chi-yue Chiu, Ying-yi Hong, 5–51. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195380392.003.0001

Tomasello, Michael. 2014. *A Natural History of Human Thinking*. Cambridge, Mass.: Harvard University Press.

Tomasello, Michael. 2019. *Becoming Human: A Theory of Ontogeny*. Cambridge, Mass.: Belknap Press.

Wellman, Henry M, David Cross, Julanne Watson. 2001. "Meta-nalysis of theory-of-mind development: The truth about False Belief." *Child Development* 72 (3): 655–84. https://doi.org/10.1111/1467-8624.00304

Wilson, Deirdre. 2017. "Relevance Theory." In *The Oxford Handbook of Pragmatics*, edited by Yan Huang, 79–100. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199697960.013.25

Organon F 28 (2) 2021: 399–419
https://doi.org/10.31577/orgf.2021.28206

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Common Ground, Conversational Roles and Epistemic Injustice

Felix Bräuer*

Received: 17 April 2019 / Revised: October 18 / Accepted: 18 December 2019

Abstract: People partaking in a conversation can add to the common ground of said conversation by performing different speech acts. That is, they can influence which propositions are presumed to be shared among them. In this paper, I am going to apply the common ground framework to the phenomenon of epistemic injustice. In doing so, I am going to focus on two kinds of speech acts: making assertions and asking certain kinds of questions. And I am going to look at three varieties of epistemic injustice: testimonial injustice, inquiring injustice and interpretative injustice. I am going to argue that what all these varieties of epistemic injustice have in common is that they unfairly inhibit the speaker's ability to add to the common ground in the way intended by her. This in turn negatively affects which conversational roles a speaker can play in a given conversation. Based on these results, I am going to end by looking at some of the harms that epistemic injustice inflicts upon its victims.

Keywords: Common ground; conversational roles; conversation-types; epistemic injustice; Miranda Fricker; Mitchell Green.

* Universität Mannheim
https://orcid.org/0000-0002-4289-0086
Universität Mannheim, Lehrstuhl Philosophie I / L 9, 5 – Raum 007 68161 Mannheim, Germany
felix.braeuer@phil.uni-mannheim.de

1. Common ground

Mitchell Green (2017a), drawing on Robert Stalnaker (2002; 2014), defines the common ground (CG) as follows: "A proposition [p] is common ground between agents A and B just in case both A and B accept p, both accept that both accept p, both accept that both accept that both accept p, and so on […]. This definition readily generalizes to more than two agents and to multiple propositions" (Green 2017a, 1589). We can understand "accepting" p as treating p as true for some purpose—e.g. believing that p is true, supposing that p is true for the sake of argument, imagining that p is true for the sake of a story (cf. Stalnaker 2014, 39; Green 2017a, 1589).

Here is an example to make this more vivid. Let's assume my friend Laura says to me: "Let's meet at the library at 3 pm". In this situation, she assumes the following, among potentially many other things, to be common ground between the two of us: (1) We both accept that we both have the same library in mind. (2) We both accept that we are both aware where it is located. (3) We both accept that we both have the same day in mind. (4) We both accept that we are in the same time-zone. We both accept that we both accept (1)-(4), and so on.

I might respond to Laura in several ways. I might agree with her: "Yes, let's do this". In this case, I signal to her that I assume the same common ground as her and that I can make it on time.[1] Or I might respond: "Sorry, I can't make 3 pm. What about 4 pm?". In this case, I signal to her that I assume the same common ground as her, but that I won't be able to comply with her proposal. Or I might respond: "Which library do you mean?". In the last case, I deny that it is common ground between us which library she is talking about. Here, Laura would have to say something like this: "I mean the library next to the university". In doing so, Laura proposes to add a new proposition to the common ground. She wants to make it the case that I accept that the library Laura wants us to meet at is the one next to the university. If she succeeds, she thereby makes it the case that we now both accept that we are both aware of which library Laura wants to meet at,

[1] At least regarding the propositions listed above. In what follows, I will omit this complication for the sake of simplicity.

and that we both accept that we both accept this, and so on. In other words, Laura has successfully *updated* the common ground between us.[2]

For reasons that will emerge in §2, I'll focus on two ways to update the common ground—by making assertions and by asking questions of a certain kind.[3] Let's assume I assert: "Meryl Streep holds the world record for most Oscar nominations". By asserting this, I want to make it the case that the proposition that Meryl Streep holds the world record for most Oscar nominations is accepted as part of the common ground between me and my interlocutor. More specifically, I want my interlocutor to believe that it is indeed Meryl Streep holds this distinction. My interlocutor might respond by saying "Wow, I didn't know this". She thereby signals to me that she believes me and hence that my attempt to update the common ground between us has been successful.

However, my attempt to update the common ground by making an assertion might also be resisted or challenged in various ways.[4] My interlocutor might respond: "How would you know? You don't even own a tv". Here, she registers my belief but challenges my competence to update the common ground in the desired way. That is, she doesn't accept the content of my assertion. It would now be up to me to answer her challenge. If I can't do so to her satisfaction, the content of my assertion won't become part of the common ground between us. Or my interlocutor might challenge the factual correctness of my statement: "No, that's Helen Mirren." Or she might doubt my sincerity: "You do not really believe this, do you?" Again, these challenges might lead to the content of my assertion not becoming part of the common ground between us.[5]

---

[2] Also cf. Lewis (1979) on "scorekeeping".

[3] Stalnaker takes assertion to be the speech act by which we paradigmatically update the common ground (cf. Stalnaker 2014, 36; Green 2017a, 1591). Green (2017a, 1590 f.) also examines the way in which we can update the common ground by asking questions. I will say more on the latter below.

[4] On the right to challenge speakers who make assertions cf. e.g. Hinchman (2005) and Goldberg (2011).

[5] Green (2017b) identifies three dimensions of commitments a speaker undertakes in making an assertion—liability, frankness, fidelity. Because the speaker makes a claim about how things are, she is "liable to being correct or incorrect depending on

A similar picture emerges with regard to questions (cf. Green 2017a, 1590 f.). Let's assume I ask: "Which movie won best picture at the Oscars in 1973?". By raising this question, I want to make it part of the common ground between me and my interlocutors that I don't know the answer to this question and also, typically, that the issue is worth looking into. If the question isn't answered immediately or shot down as irrelevant, then this means that it is accepted into the common ground as an *open question*. That is, we both accept it as true that we don't know, and that we both accept that we both accept that we don't know which movie won best picture in 1973, and that this is an issue that is worth looking into. Alternatively, it might happen that my question isn't shot down as irrelevant, but answered immediately. Here too, in a sense, the question is accepted into the common ground. After all, it is dealt with appropriately. However, once it has been answered, it will no longer be part of the common ground as an open question. Irrespective of this, by accepting the question into the common ground, we both also accept an existential presupposition into the common ground, namely that there was (at least) one movie that won best picture in 1973, and thus also the meta-linguistic presupposition that there is (at least) one correct answer to the question raised. And we both accept that we both accept these presuppositions, and so on. [6]

My interlocutor might now look up the information on her smartphone and assert: "It was the Godfather". If the content of her assertion is accepted then it is in turn added to the common ground. The issue is now settled.

However, a question might be resisted or challenged as well. Let's assume I ask: "Is right-wing populism on the rise in Europe again?" Here, my

how things are". She is also committed to "frankness"; that is, committed to believing what she asserts. Moreover, Green uses the term "fidelity" to express the notion that a speaker who makes an assertion thereby commits herself to respond to challenges by justifying her claims. We can see these commitments in play in the above examples. Because of "liability" and "frankness", it is appropriate to challenge the speaker's assertion in the ways indicated above. Because of "fidelity", the speaker is committed to respond to such challenges. Moreover, how well she is able to respond to such challenges will (at least in part) determine whether her assertion becomes part of the common ground.

[6] Thanks to an anonymous reviewer for pressing me to be clearer on this issue.

interlocutor might respond: "Don't be ridiculous, of course it is. How can you even doubt this?" Or she might respond: "Don't be ridiculous, of course it isn't. How can you even ask this?" In both cases, my question is rejected as illegitimate, because the issue is taken to be settled already and hence not worthy of (further) investigation. It doesn't become part of the common ground (what will likely become part of the common ground is that I have asked this question).

Somewhat more generally, Green states:

> CG depends on what interlocutors accept, so if certain parties to a conversation refuse to acknowledge that one speaker has performed a speech act, then neither its force nor its content will become part of CG, which will only register that this speaker has performed an act of speech. Or perhaps they acknowledge the performance of an illocution, but put the speaker up to impossibly high standards before the content of that illocution is absorbed into CG. (Green 2017a, 1600 f.).

These remarks suggest that a speaker might be unjustly prevented from adding to the common ground. In what follows, I am going to take up this suggestion. That is, I will be using the common ground framework outlined so far to look at the debate on epistemic injustice. I am going to argue that in cases of epistemic injustice, what typically happens, among other things, is this: *A speaker is, due to a negative identity stereotype against her, unfairly*[7]

[7] It is worth stressing that I take the prevention in question to be unfair because it is based on a negative identity stereotype. Of course, this leaves open the possibility that there might be other unfair reasons for preventing a speaker from adding to the common ground in the way intended by her. Moreover, I take it that is not always unfair to prevent a speaker from adding to the common ground due to considerations connected to the speaker's identity. To see this, consider the following example: I have a friend who is an expert in economics but who is ignorant about physics. Both is known to me. Now, this friend makes two claims—one about the economical aspects of nuclear energy and the other about its hazardous nature. If I accept her first claim, but not the second (which, as a consequence, doesn't become part of the common ground between us), it seems that I am hardly behaving unfairly towards her. The reason is that my rejection of her second claim is not based on a negative identity stereotype against her. Rather, the rejection is based on a fair

prevented from adding to the common ground in the way intended by her.[8]

assessment of her respective qualifications to make statements concerning different topics. (Thanks to Maciej Witek for pressing me to be clearer on these issues and for suggesting the example).

[8] One might wonder about cases in which a speaker is unfairly prevented from adding to the common ground, not due to a negative identity stereotype against her, but due to some more specific features the speaker exhibits. Wouldn't such cases also count as examples of epistemic injustice? Let's assume I wrongly don't believe a speaker because she exhibits some nervous tick, which I mistake for a sign of dishonesty. Although such behaviour would be somewhat unfair and potentially harmful to the speaker, I wouldn't, following Fricker (2007), classify this as an instance of epistemic injustice, but rather of an instance of "epistemic bad luck". Here, it's not that I don't believe the speaker because I am harbouring prejudices against her. The only reason I'm not believing her is that I am missing important information about her. She exhibits a behaviour commonly associated with dishonesty because she is nervous, not because she in fact is dishonest. What is more, if I had this piece of information, then I would likely believe her. It's just bad luck for her that I do not. The latter is not the case with epistemic injustice. In the above examples, the speaker doesn't suffer a credibility deficit because the hearer lacks important information. Instead, the (unjust) reason for distrust is more systematic and more deeply rooted.

Of course, this is not to say that in cases where testimonial injustice is present, acquiring more information about the speaker can't change a hearer's perception of the speaker's credibility. To see this, consider the following example: A woman, who has a degree in nuclear physics, claims that nuclear energy is dangerous. Her male interlocutor, who doesn't know she has the relevant expertise, dismisses her claim as the ramblings of an anxious woman. Moreover, let's assume that (a) he would have believed her, had he known about her physics degree, and that (b) he would have simply believed a male interlocutor making the same claim, independent from whether he had any relevant background information about said interlocutor, and that (c) this difference in reaction on his part is due to a negative identity stereotype he harbours against women. Here, we have a case of testimonial injustice where having additional information would affect how the hearer perceives the speaker's credibility. That being said, it is worth stressing that this example is different from the nervous tick-example in so far as here a negative identity stereotype, and not just lack of information, is crucial for explaining the low credibility initially attributed to the speaker. (Thanks to an anonymous reviewer for suggesting the second example and for pressing me to be clearer on this issue).

2. Varieties of epistemic injustice

Miranda Fricker (2007) discusses two kinds of epistemic injustice—*testimonial injustice* and *hermeneutical injustice*.[9] The former is what I am going to focus on in this paper. One is a victim of testimonial injustice if, in making a statement, one suffers a credibility deficit due to negative identity stereotype. One of the examples Fricker provides for this phenomenon is the following: In the screenplay for Patricia Highsmith's *The Talented Mr. Ripley* Marge Sherwood wants to convince the father of her fiancé, Herbert Greenleaf, that his son has been murdered by Tom Ripley. Mr. Greenleaf responds to this by dismissing the content of her assertion—"Marge, there's female intuition, and then there are facts". Here, Marge isn't believed by Mr. Greenleaf, although she is in fact right, due to a credibility deficit based on a negative identity stereotype: As a woman she is taken to be emotional rather than rational. Hence, so Herbert Greenleaf's thinking goes, one shouldn't attach too much weight to her words in this situation. [10] The injustice Marge suffers here is epistemic in kind in so far as she is not taken seriously as a valuable informant, and hence as a provider of knowledge, on the issue in question.[11]

---

[9] For an overview over the epistemic injustice debate cf. McKinnon (2016).

[10] McGowan (2009) and Caponetto (2020) discuss a similar phenomenon under the moniker "authority silencing", which takes place, for example, "when women speak, or try to speak, as experts in male-dominated fields. […]. Often, in spite of being competent and thus satisfying the requirements for counting as an expert in a given area, a woman finds that her utterances do not count as expert speech acts. Her expert status is not recognized and hence fails to give hearers any special reason to trust what she claims" (Caponetto 2020, 7). I believe, however, that the phenomenon of testimonial injustice, as conceived by Fricker, is broader than this. Marge, in the example above, is clearly epistemically wronged. But she is not wronged as an expert in a male-dominated field. After all, she isn't a police officer and hence doesn't make her statement concerning Tom Ripley's guilt in this capacity.

[11] In this paper, I shall follow Fricker in concentrating on cases in which there is a negative identity stereotype against the speaker. However, one might also hold that giving a speaker a credibility access due to some positive identity stereotype constitutes a form of epistemic injustice. As Medina (2013) points out, credibility judgements often have a contrastive quality. That is, the credibility of a testifier is

In the wake of Fricker's influential work, philosophers have identified further varieties of epistemic injustice. For instance, Christopher Hookway (2010) draws attention to a phenomenon we can call *inquiring injustice*. Hookway notices that a person can become a victim of epistemic injustice not just when she asserts something, but also when she asks questions of a certain kind. That is, she can become a victim of epistemic injustice when she raises questions that attempt to shape an *inquiry*. When I talk of "epistemic injustice" in connection with raising questions, it's such *inquiry questions* I have in mind, although I will sometimes simply talk about "questions" for the sake of brevity.[12]

Here is an example to make more vivid what I mean by "inquiry questions": Imagine a philosophy seminar on knowledge during the 1950th, when contextualist approaches to knowledge ascriptions weren't on people's radars.[13] During this seminar a student asks: "What if the truth values of knowledge ascriptions are context dependent?" The student raises this question in order to draw attention to a worthy subject of inquiry. Her professor however dismisses this question, because she believes that students don't have the philosophical skills to determine which questions do and do not merit philosophical inquiry. Here, the student also suffers a kind of epistemic injustice. Due to her identity—just a student—she is not taken to be the kind of person who can contribute to a philosophical inquiry by raising questions that might shape said inquiry.

One might object to this assessment on the basis that the student, unlike Marge, didn't assert something. Hence, it is not the case that the student is wronged in her capacity as an informant or source of knowledge.

assessed against the credibility of others. Greenleaf, to pick up on the above example, assesses Marge's credibility against Tom's. And the credibility deficit that Marge suffers is, at least to some extent, due to the credibility excess Greenleaf assigns to Tom. (Thanks to an anonymous reviewer for pointing this out).

[12] I use this locution to distinguish the kinds of questions I am interested in in this paper from more mundane questions, such as "What time is it?", "What's your favourite ice cream?", "Are you eating that?", and so on.

[13] Although there are some precursors, contextualist approaches to knowledge ascriptions became prominent in the 1990th due to e.g. DeRose (1992), Lewis (1996), Cohen (1999).

Nevertheless, I believe that an epistemic injustice is committed against the student. After all, an inquiry is an epistemic activity in so far as it is aimed at generating knowledge or fostering understanding.[14] And the student is, due to a negative identity stereotype against her, unfairly prevented from contributing to this activity.

As should be apparent from the discussion of the above examples, the reason why I shall focus on assertions and inquiry questions is that they are especially important from an epistemic perspective. Assertions serve the function of sharing information and, if all goes well, spreading knowledge.[15] And inquiry questions serve the function of shaping activities aimed at gathering knowledge and promoting understanding. That being said, I don't deny that other speech acts (e.g. conjectures) are of epistemic importance as well.

What both testimonial injustice and inquiring injustice, as presented here, have in common is that the contents of certain contributions are dismissed when they shouldn't be, and that this happens due to negative identity stereotypes against the speakers who try to make these contributions. Andrew Peet (2017), however, points out that epistemic injustice can happen even earlier in the process of communication. One can also become a victim of epistemic injustice in virtue of the way one's utterance is interpreted; before the utterance is then further assessed for credibility and/or merit. Due to one's identity, one's utterance might be interpreted in ways one never intended.

As an example, following Fricker's (2007) reading of Harper Lee's *To Kill a Mockingbird*, Peet imagines a black man saying about a white woman "she seemed vulnerable to me". What the latter tries to convey is that he wanted to help her, because she didn't have anyone and he felt sorry for her. However, his interlocutors interpret his statement as saying that she would make an easy victim. In interpreting his statement like this, they are

[14] For more on why inquiring should be understood as a distinctly epistemic activity and what the characteristics of this activity are, cf. Freedman (2019, forthcoming).

[15] Cf. Kelp (2018) for related discussion.

guided by a negative identity stereotype that takes black men to be aggressive and prone to criminality.[16]

Although, intuitively, an injustice is committed in Peet's example, it might, compared to the two previous examples, be less obvious why we are dealing with a case of epistemic injustice here. After all, it's not that the black man's interlocutors don't believe him when he says that the woman seems vulnerable to him. Rather, the problem here is that they unfairly misinterpret the attitude expressed in his statement. The statement is taken to express malice rather than sympathy and pity. Nevertheless, there clearly is an epistemic dimension to the injustice committed here. The unfair misinterpretation happens because the black man's interlocutors have false background beliefs about black people and are not open to counterevidence that might challenge these beliefs. In fact, they fail to see his statement as potential counterevidence against their preconceived notions and instead unfairly interpret it in a way that fits these preconceived notions.[17]

[16] Cf. Peet (2017, 3432). The phenomenon Peet draws our attention to bears resemblance to what is often discussed under the headline "silencing" (cf. e.g. Langton 1993; Hornsby 1995a, 1995b; Langton & Hornsby 1998; Maitra 2009, 2017; McGowan 2009; Dotson 2011; McKinnon 2016; Caponetto 2020). A paradigm case to illustrate this phenomenon is the following: "A woman says 'No' to a man, intending to refuse sex. The man understands the conventional meaning of her utterance, and recognizes the content it expresses. Nevertheless, the utterance does not do what she wants it to do: it does not deter him from forcing sex on her" (Maitra 2009, 313). Although different accounts of silencing differ in their diagnosis of what goes wrong in this scenario, they are united by the thesis that the woman is silenced in so far as she gets disabled in her ability to communicate her refusal of sex. One possible explanation for this, that is given in the literature, is that her refusal gets misinterpreted. The man interprets her "No" not as a genuine refusal, but rather as an attempt to not appear too sexually forward. According to this interpretation, like in Peet's example, the message that is received is drastically different from the one that was intended to be conveyed.

[17] One might wonder whether the cases that were presented to illustrate testimonial injustice and inquiring injustice might instead be regarded as cases of interpretative injustice. For instance, Mr. Greenleaf might be regarded as misinterpreting Marge's utterance by taking it to be an expression of her fears, rather than an assertion issued with the required 'illocutionary' authority. Moreover, one might take the professor to misinterpret the student's utterance as a case of showing-off, rather than as a genuine question aimed at shaping the course of inquiry. While it is certainly

Let us briefly take stock: I have given an overview over three varieties of epistemic injustice—testimonial injustice, inquiring injustice and interpretative injustice. In each case, a speaker is treated unfairly due to a negative identity stereotype against her on the part of her respective interlocutor. Yet, apart from their common origin, these three varieties of epistemic injustice look quite different. Testimonial injustice is concerned with assertions. Inquiring injustice is concerned with questions of a certain kind. And interpretative injustice is concerned with how a speaker's message is understood by the hearers. In spite of these apparent differences, I want to suggest that these varieties of epistemic injustice are united by their effect. To do so, I will be utilizing the common ground framework laid out in the last section.

However, before exploring this idea, let me briefly address what might appear to be a tension between the common ground framework and the debate on epistemic injustice, especially regarding the first two examples. The common ground framework seems to be linked to the assumption that speakers engaged in a conversation are typically co-operative in a Gricean (e.g. 1975) sense. For example, a natural explanation for why the common ground between interlocutors can often be updated quite effortlessly is that the interlocutors implicitly assume each other to be co-operative. But such an assumption of cooperation seems not to be (at least not fully) present in conversations in which epistemic injustice takes place—e.g. those committing epistemic injustice tend to assume that their interlocutors' contributions are not relevant or of low quality.[18]

Yet, this tension disappears upon closer inspection. It's not that instances of epistemic injustice constitute counterexamples against Grice's

possible that things like this might happen, we can simply stipulate that this is not what happens in the examples given in the main text. That is, we can stipulate that Mr. Greenleaf (correctly) understands Marge's utterance as an assertion. It's just that he dismisses said assertion due to a negative identity stereotype. Similarly, we can stipulate that the professor understands the student's contribution as a genuine inquiring question. It's just that, due to a negative identity stereotype, the professor doesn't engage with this question—she simply thinks that students are in no position to evaluate what are worthy subjects of philosophical inquiry. (Thanks to Maciej Witek for raising this issue).

[18] Thanks to an anonymous reviewer for raising this worry.

cooperative principle. It's rather that we can use this principle to diagnose what, among other things, goes wrong when epistemic injustice takes place. In cases of epistemic injustice, an interlocutor falsely and unfairly assumes that the speaker violates the cooperative principle—e.g. an interlocutor falsely and unfairly assumes that the speaker's contribution falls short of certain standards of relevance or quality.

After having addressed this potential worry, let's take another look at the examples considered so far to explore the thesis put forward at the end of the last section. The thesis that in cases of epistemic injustice the speaker is, due to a negative identity stereotype against her, unfairly prevented from adding to the common ground in the way intended by her: Marge wants to update the common ground by adding the proposition that her fiancé, Dickie Greenleaf, has been murdered by Tom Ripley. However, her attempt to do so is resisted by Herbert Greenleaf. He deems her justification for her assertion—Tom is wearing Dickie's ring and Dickie promised her to never take it off—to be unsatisfactory. He brushes her accusation against Ripley off by saying "Marge, there's female intuition, and then there are facts". Consequently, the content of Marge's assertion does not become part of the common ground (it only becomes part of the common ground that Marge beliefs that Tom Ripley is guilty).

The student wants to update the common ground by proposing a question she takes to be a worthy starting point for philosophical inquiry. However, her professor believes that students are not in a position to determine what is and what isn't worthy of such inquiry. Therefore, she dismisses her question. Hence, although the question is registered, it doesn't become part of the common ground in the sense that it is accepted that perusing it might lead to interesting results. It won't be taken as worthy of further investigation in the context of the seminar.

And the black man in Peet's example intends to make it part of the common ground that there was a vulnerable woman that might have benefited from getting some help. However, this is not how his statement is interpreted by his interlocutors. They rather take him to classify her as a potential victim of his. What they take away from the conversation is very different from what he had in mind when he made his utterance.

It is worth stressing here that, in all of these examples, it might very well be the case that some proposition becomes common ground between the speaker's interlocutors as a result of her utterance. Crucially, however, in each instance, this won't be the proposition intended by the speaker. And what is more, said proposition will likely be anathema to the speaker's intention and unfairly proof harmful to her[19]—"Marge is a hysterical woman" instead of "Tom Ripley has murdered Dickie Greenleaf"; "This

[19] The qualifications "harmful", "unfairly", and "as a result of her utterance" are crucial for characterizing instances of epistemic injustice. After all, it is frequently the case that a speaker's utterance will lead to propositions becoming part of the common ground between all or some of those involved, although the speaker didn't intend them to become part of the common ground. Imagine someone says "I just picked up a new suit from my tailor". As a result of this utterance, it might become part of the common ground between his interlocutors that the speaker is well off and that he is particular about his clothes, although this wasn't intended by the speaker. But, intuitively, we wouldn't say that these interlocutors commit an epistemic injustice against the speaker. Presumably, part of the explanation for this assessment is that we wouldn't say that these propositions are anathema to the speaker's intentions, or that they will proof unfairly harmful to him.

Or imagine that a male driver says "Sorry for being late, there were so many slow women drivers on the road today" (this example is adapted from Faulkner (2011)). Here, the speaker intends to give his interlocutors a reason for him being late. But it might very well become part of the common ground between his interlocutors that he is a sexist, although he did not intend this to happen. While this addition to the common ground might lead to the others viewing the speaker in an unfavourable light, we would hardly say that this is unfair towards the speaker.

Finally, imagine the following scenario (suggested by an anonymous reviewer): Suppose that Marge accuses Tom of being Dickie's murderer while bursting into tears. Moreover, suppose Greenleaf believes her, but also takes her tears to be a symptom of her overly emotional female nature. This thought then becomes common ground between Greenleaf and a few other male figures who listen to Marge's accusation. Here, the common ground would be updated in a way that is both unfair to Marge and harmful to women. Nevertheless, intuitively, we wouldn't be dealing with a case of epistemic injustice. The reason is that the common ground doesn't get updated in this way as a result of Marge's utterance per se. Rather, the common ground gets updated in this way because of the emotion Marge displays while making her utterance. (Thanks to an anonymous reviewer for pressing me to be clearer on these issues.).

student hasn't understood part of the lecture" instead of "Context-sensitivity is a worthy subject of inquiry"; "He wanted to rob her" instead of "She needs our help".

Moreover, it is worth stressing that the proposition might not become part of the common ground of the conversation that the speaker herself is part of. After all, the speakers in question will likely not accept the respective proposition. Still, the proposition might well become part of the common ground of conversations that the speaker's interlocutors subsequently (or even simultaneously) have among themselves. Marge, for example, will likely not accept the proposition that she is a hysterical woman. Hence, this proposition will not become part of the common ground of the conversation she herself is part of. Still, the proposition that she is a hysterical woman might well become part of the common ground of conversations that her male interlocutors subsequently (or even simultaneously) have among themselves.[20]

The considerations presented in this section confirm the thesis that was put forward: *Different varieties of epistemic injustice are united by their effect*: *A speaker is, due to a negative identity stereotype against her, unfairly prevented from adding to the common ground in the way intended by her.* Moreover, it was indicated in this section that this is bound to have negative effects on the speaker. I am going to end by discussing some of the negative consequences that might arise as a result; for the speaker and for the conversational project she wants to contribute to. To do so, I am going to make use of Mitchell Green's examination of different conversation-types.

3. The conversational harms of epistemic injustice

Green (2017a, 1593 ff.) provides a taxonomy of conversation-types. In doing so, he, following Stalnaker (1970), differentiates between *inquiries*

[20] That's why it says in the preceding paragraph "it might very well be the case that some proposition becomes common ground between *the speaker's interlocutors* as a result of her utterance" (emphasis added), rather than "it might very well be the case that some proposition becomes common ground between *the speaker and her interlocutors* as a result of her utterance".

and *deliberations*. Inquiries aim at answering theoretical questions while deliberations aim at answering practical questions.[21] In both types of conversations, speakers can play different roles—symmetrical and asymmetrical ones. When speakers occupy symmetrical conversational roles, they pool their information and negotiate their desires to achieve the common goal of the conversation. If the roles are asymmetrical, a speaker answers theoretical or practical questions for the others (asymmetrical didactic). Or a speaker leads the others to answering theoretical or practical questions for themselves (asymmetrical socratic).

This analysis of conversational roles can help us to better understand the negative effects of epistemic injustice. As was shown in the last section, different varieties of epistemic injustice are united by the common feature of unfairly preventing speakers from adding to the common ground in the ways intended by them. This in turn severely limits the roles these speakers can play in a given conversation. As they are typically not believed, not taken seriously, or misinterpreted when they assert something or raise an inquiry question, they are curtailed in their ability to make contributions that might help to solve theoretical or practical questions.

Consider the examples from the previous section: As Marge is not believed, she is prevented from providing information that might help to answer the question what happened to Dickie Greenleaf. Because the student is not taken seriously, she is not in a position to ask questions that might shape the course of an inquiry concerning knowledge ascriptions. And because the black man in Peet's example is viewed as a predator, rather than as a good Samaritan, he is not seen as someone who might be a force for good in his community. Put in more general terms: *Victims of epistemic injustice are impeded in their ability to play a symmetrical role in a given conversation.*

To avoid a potential misunderstanding, a qualification is in order here. By "impeded" I don't mean that victims of epistemic injustice won't be able to play symmetrical roles in a given conversation at all. It's just that, in virtue of the negative identity stereotypes against them, the symmetrical

21 Similarly, Stalnaker, roughly speaking, takes inquiries to be concerned with finding out what is the case, and deliberations to be concerned with finding out what to do (cf. Stalnaker 1970, 280).

contributions they can make will be unfairly curtailed. That is, these contributions will be limited to topics where the speaker is not viewed in light of a negative identity stereotype, or where the negative identity stereotype does not deflate the assessment of credibility or competence, or where there might even be a positive identity stereotype associated with the speaker; e.g. a woman might well play a symmetrical role in a conversation on how to raise a child, a student when it comes to the latest university gossip, or a black man when it comes to the athletic performance of a sports team.[22]

That being said, victims of epistemic injustice will often be forced into an asymmetrical role, if they are allowed to participate in the conversation at all. They are told things. Or things are explained to them. Think of "mansplaining" as a prototypical example for such an asymmetrical conversation dynamic. For our purposes we can understand mansplaining as the phenomenon where a man condescendingly explains something to a woman, although he possesses less knowledge on the issue in question than she does.[23] Here, the man automatically (and wrongly) assumes that a woman is not on equal footing with him regarding the topic at issue. She might learn from him but not vice versa. Thus, he will be prone to dismiss her attempts to contribute to the conversation in ways meant to educate him on the issue in question. Rather, her best shot at getting his attention and getting him to engage with her contributions will be to ask him for information, asks him to explain things to her, asks him how to do stuff, or makes statements that play into his preconceived notions, like "I don't understand", "Ah, I see now", "Thank you for explaining". In short, mansplaining constitutes a case of epistemic injustice in so far as the man doesn't take his female interlocutor seriously as a source of knowledge from whom he might learn, but one-sidedly sees her as someone who could benefit from his knowledge.

Somewhat generalizing these observations, we can say that epistemic injustice fosters a power imbalance between conversationalists. The victims of epistemic injustice will likely be cast into a passive role. They will often not be seen as equals who can make a valuable contribution to inquiries or

[22] Thanks to an anonymous reviewer for pressing me to be clearer on this issue.

[23] Cf. https://www.merriam-webster.com/words-at-play/mansplaining-definition-history.

deliberations. Rather, they will typically be cast into the role of an inferior. That is, they will be seen as being dependent on others, and they will be prevented from conversing with others as equals. In short: *Their conversational agency gets impaired.*

Some practical negative effects of this impairment are quite obvious. When people are prevented from adding to the common ground in the way outlined above, they can't get credit for their work or their ideas. After all, substantial contributions they are trying to make are prone to be ignored, dismissed or misinterpreted. Again, consider the previous examples: Marge won't get credit for solving the disappearance of Dickie Greenleaf. The student in the philosophy seminar won't be mentioned by future epistemologists when they discuss contextualism. And the black man who just wanted to help won't be seen as someone who might make his neighborhood a better place. They are all, to borrow a phrase from Ishani Maitra (2009, 331), "unfairly deprived of the benefits of speech".[24]

Moreover, the positive effects that the speakers intended to bring about with their utterances might not take place at all, or at least might take place later than they could have otherwise. There might never be justice for Dickie Greenleaf, as Marge was the only one to see through Tom Ripley. An important strand in contemporary epistemology might have gotten off to a much earlier start. The quality of life in the neighborhood might not improve, or improve later than it could have.

However, I believe that the negative effects of this impairment cut even deeper. To see why, it will help to look at what Fricker (2007) identifies as a crucial harm inflicted by epistemic injustice, especially testimonial injustice. According to Fricker, in cases of epistemic injustice "a person is wronged in her capacity as a knower" and "[t]o be wronged in one's capacity as a knower is to be wronged in a capacity essential to human value"

[24] I say "borrow" here because the point I make is somewhat different from Maitra's. My point is that victims of epistemic injustice are deprived of the benefits of speech in so far as these victims won't get credit for the ideas they express. In contrast to this, Maitra is concerned with how speech gives people a voice in democratic societies (e.g. to protest or to criticise) and how silencing (cf. fn. 17) deprives people of this voice. (Thanks to an anonymous reviewer for pressing me to be explicit on this difference.).

(Fricker 2007, 44).[25] The key idea here is that our capacities to know things and to spread knowledge are crucial parts of our identity qua human beings. Thus, not to be taken seriously in these capacities, as happens in cases of epistemic injustice, is essentially "dehumanizing" (cf. Fricker 2007, 44).

If what I have been saying so far is on the right track, then the dehumanizing effect of epistemic injustice goes even further. While I agree with Fricker that our capacities to know things and to spread knowledge are crucial to our identity qua human beings, surely other capacities are essential here as well. I take it that another essential and related capacity is our capacity to communicate with each other more broadly—a capacity that importantly includes our ability to partake in conversations. But if victims of epistemic injustice are, as I have been arguing, impaired in their conversational agency (in virtue of being impeded in their ability to play a symmetrical role in a given conversation), then this means that their ability to partake in conversations gets curtailed.

The latter is crucial for assessing the dehumanizing effect of epistemic injustice: if the ability to partake in conversations is an important part of our capacity to communicate, and if this capacity in turn is an important part of our identity qua human beings, then this means that victims of epistemic injustice are also dehumanized in virtue of being curtailed in their ability to partake in conversations. Put differently, it's not just that victims of epistemic injustice are dehumanized in so far as they are wronged in their capacity as knowers. It's also that they are dehumanized in so far as they are wronged in their capacity as communicators more broadly.

4. Conclusion

In this paper, I have looked at epistemic injustice through the prism of Mitchell Green's work on the common ground and on conversation-types. This has allowed me to unearth a feature that different varieties of epistemic injustice have in common. When one is a victim of epistemic injustice, one is, due to this, unfairly prevented from adding to the common ground in

[25] Here, Fricker in turn draws on Craig's (1990) genealogical account of our concept of knowledge.

the way one intends. This in turn has an impact on the conversational roles one can play. Victims of epistemic injustice are likely to be forced into asymmetrical conversational roles. That is, they are bound not to be taken as equals who can make valuable contributions to inquiries or deliberations. Rather, they are bound to be perceived as people who need things to be explained to them, or to be shown to them. On a practical level, this is likely to have negative consequences for the victims of epistemic injustice as well as for the conversational projects they are prevented from engaging in properly. Moreover, this is even bound to dehumanize victims of epistemic injustice in so far as they are wronged in a capacity that is essential to human identity, the capacity to communicate.

Acknowledgements

This paper grew out of talk I gave at the workshop "The Origins of Meaning and the Nature of Speech Acts. A Workshop with Mitchell Green" (University of Szczecin). I want to thank Maciej Witek for inviting me to participate. Moreover, I want to thank Mitchell Green, as well as the other participants, for their very insightful comments. Finally, I want to thank the anonymous reviewers, as well as Maciej Witek, for their detailed feedback. You all helped to significantly improve this paper!

References

Caponetto, Laura. 2020. "A Comprehensive Definition of Illocutionalry Silencing." *Topoi*, Online First: 1–12. https://doi.org/10.1007/s11245-020-09705-2

Cohen, Steward. 1999. "Contextualism, Skepticism, and the Structure of Reasons." *Philosophical Perspectives* 13: 57–89.

Craig, Edward. 1990. *Knowledge and the State of Nature*. Oxford: Clarendon Press. https://doi.org/10.1093/0198238797.001.0001

DeRose, Keith. 1992. "Contextualism and Knowledge Attributions." *Philosophy and Phenomenological Research* 52 (4): 913–29. https://doi.org/10.2307/2107917

Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Patterns of Silencing." *Hypatia* 26 (2): 236–57. https://doi.org/10.1111/j.1527-2001.2011.01177.x

Faulkner, Paul. 2011. *Knowledge on Trust*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199589784.001.0001

Freedman, Jane. 2019. “Inquiry and Belief.” *Noûs* 53 (2): 296–315. https://doi.org/10.1111/nous.12222

Freedman, J. (forthcoming). “The Epistemic and the Zetetic”. *The Philosophical Review*.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: University Press. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001

Goldberg, Sanford. 2011. “Putting the Norm of Assertion to Work: The Case of Testimony.” In *Assertion: New Philosophical Essays*, edited by Jessica Brown, and Herman Capellen, 175–95. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199573004.001.0001

Green, Mitchell. 2017a. “Conversation and Common Ground.” *Philosophical Studies* 174 (6): 1587–604. https://doi.org/10.1007/s11098-016-0779-z

Green, Mitchell. 2017b. “Assertion”. *Oxford Handbooks Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-8 .

Grice, Paul. 1975. “Logic and Conversation.” In *Studies in Syntax and Semantics III: Speech Acts*, edited by *Peter Cole*, and Jerry L. *Morgan,* 183-98. New York: Academic Press: https://doi.org/10.1163/9789004368811_003

Hinchman, Edward. 2005. “Telling as Inviting to Trust.” *Philosophy and Phenomenological Research* 70 (3): 562–87. https://doi.org/10.1111/j.1933-1592.2005.tb00415.x

Hookway, Christopher. 2010. “Some Varieties on Epistemic Injustice: Reflections on Fricker.” *Episteme* 7 (2): 151–63. https://doi.org/10.3366/E1742360010000882

Hornsby, Jennifer. 1995a. “Illocution and its Significance”. In *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, edited by Savas L. Tsohatzidis, 187–207. New York: Routledge. https://doi.org/10.4324/9780203206478

Hornsby, Jennifer. 1995b. “Disempowered Speech.” *Philosophical Topics* 23: 127–47. https://doi.org/10.5840/philtopics199523211

Hornsby, Jennifer and, Rae Langton. 1998. “Free Speech and Illocution.” *Legal Theory* 4 (1): 21–37. https://doi.org/10.1017/S1352325200000902

Kelp, Christoph. 2018. “Assertion: A Function First Account.” *Noûs* 52 (2). 411–42. https://doi.org/10.1111/nous.12153

Langton, Rae. 1993. “Speech Acts and Unspeakable Acts.” *Philosophy & Public Affairs* 22 (4): 293–330.

Lewis, David. 1979. “Scorekeeping in a Language Game.” *Journal of Philosophical Logic* 8 (1): 339–59. https://doi.org/10.1007/BF00258436

Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74 (4). 549–67. https://doi.org/10.1080/00048409612347521

McGowan, Mary K. 2009. "On Silencing and Sexual Refusal." *The Journal of Political Philosophy* 17 (4): 487–94. https://doi.org/10.1111/j.1467-9760.2009.00346.x

Maitra, Ishani. 2009. "Silencing Speech." *Canadian Journal of Philosophy* 39 (2): 309–38. https://doi.org/10.1353/cjp.0.0050

Maitra, Ishani. 2017. "Speech and Silencing." In *The Routledge Companion to Feminist Philosophy*, edited by Ann Garry et al., 279–91. New York: Routledge. https://doi.org/10.4324/9781315758152

Mansplaining. https://www.merriam-webster.com/words-at-play/mansplaining-definition-history .

McKinnon, Rachel. 2016. "Epistemic Injustice." *Philosophy Compass* 11(8): 437–46. https://doi.org/10.1111/phc3.12336

Medina, José. 2011. "The Relevance of Credibility Excess in a Proportional View of Epistemic Injustice: Differential Epistemic Authority and the Social Imaginary." *Social Epistemology* 25 (1): 15–35. https://doi.org/10.1080/02691728.2010.534568

Medina, Jos. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and the Social Imagination*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199929023.001.0001

Peet, Andrew. 2017. "Epistemic Injustice in Utterance Interpretation." *Synthese* 194 (9): 3421–443. https://doi.org/10.1007/s11229-015-0942-7

Pohlhaus, Gaile. 2012. "Relational Knowing and Epistemic Injustice: Towards a Theory of Willful Hermeneutical Ignorance." *Hypatia* 27 (4): 715–35. https://doi.org/10.1111/j.1527-2001.2011.01222.x

Stalnaker, Robert. 1970. "Pragmatics." *Synthese* 22 (1-2): 272–89. https://doi.org/10.1007/BF00413603

Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* 25: 701–21. https://doi.org/10.1023/A:1020867916902

Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199645169.001.0001

Organon F 28 (2) 2021: 420–457
https://doi.org/10.31577/orgf.2021.28207

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Conclusions of Practical Argument: A Speech Act Analysis

Marcin Lewiński*

Received: 19 June 2019 / Accepted: 21 April 2020

Abstract: Conclusions of theoretical reasoning are assertions—or at least speech acts belonging to the class of assertives, such as hypotheses, predictions or estimates. What, however, are the conclusions of practical reasoning? Employing the concepts of speech act theory, in this paper I investigate which speech acts we perform when we're done with an instance of a practical argument and present its result in a linguistic form. To this end, I first offer a detailed scheme of practical argument suitable for an external pragmatic account (rather than an internal cognitive account). Resorting to actual examples, I then identify a class of action-inducing speech acts as characteristic conclusions of practical argument. I argue that these speech acts—promises, orders, pieces of advice, proposals, and others—differ chiefly depending on the agent of the action induced (me, us, you, them) and their illocutionary strength.

Keywords: Illocution; practical argument; practical reasoning; speech acts.

1. Introduction

Practical reasoning and theoretical reasoning are typically defined as, respectively, reasoning about what to do and about what to believe. Yet,

* Nova Institute of Philosophy, Nova University of Lisbon
https://orcid.org/0000-0002-7218-3948
Campus de Campolide – Colégio Almada Negreiros 1099-032 Lisboa, Portugal
m.lewinski@fcsh.unl.pt

the seemingly innocent "about" might be quite misleading here, and in a dual sense. The distinction does not in fact pertain to 1) the *content* of 2) the *premises* which we are reasoning "about," such as when we reason (whether practically or theoretically) "about" Brexit or Donald Trump. Rather, it refers to 1) the *function* of 2) the *conclusion* of reasoning. We thus reason "about" what we conclude we can do about Brexit or "about" what we conclude we should believe about Trump. Given this, the analysis of the function of the conclusion of reasoning is, by definition, crucial. This is my task for this paper.

In this task, I will avail myself of the basic idea of speech act theory, namely, that various functions of our language use can be comprehensively elucidated via the concept of speech acts.[1] This requires attention to the conclusions of our reasoning as linguistically constituted via practical argumentation—or at least linguistically represented; I give attention to this issue directly below, in Sections 2 and 3. Looking from the perspective of practical argumentation, and not just reasoning, the basic problem of this investigation can be represented as follows:

Conclusions of theoretical reasoning are assertions—or at least speech acts belonging to the class of assertives, such as hypotheses, predictions or estimates. This follows directly from standard definitions of assertives and of theoretical reasoning: both are about how things are and, as such, can be true or false. On a standard view (but see Section 2.4 for a challenge), theoretical reasoning is thus assertoric through and through: we insert various types of assertives as premises and conclude with another assertive. *What, however, are the conclusions of practical reasoning?* What do we do with words when we arrive at a conclusion of a practical argument? Which speech acts do we perform when we're done with an instance of a practical argument? (Other than: "I'm done!")

To answer these questions, in Section 3 I will offer a detailed scheme of practical argument suitable for an external pragmatic account (rather than an internal cognitive account). Resorting to actual examples (Section 4), I then identify a class of *action-inducing speech acts* as characteristic conclusions of practical argument (Section 5). I argue that these speech acts—

[1] For recent accounts see (Green 2020) and (Fogal, et. al eds. 2018); for classic works see, of course, (Austin 1975) and (Searle 1969).

promises, orders, pieces of advice, proposals, and others—differ chiefly depending on the agent of the action induced (me, us, you, them) and their illocutionary strength.

2. What concludes practical reason?

2.1. Attitudes or acts of a reasoning agent

Philosophical accounts of practical reasoning (henceforth: PR) are still dominated by the first-person perspective of a single reasoning agent (Audi 2006; Broome 2013; Davidson 1963; Gauthier 1963; Pollock 1995; Searle 2001)—even if work on collective intentions and actions is ever-more prominent (Bratman 2014; Gilbert 1990; Searle 2010; Tuomela 2013). Accordingly, the mainstream philosophical discussion over how to conclude PR revolves around the issue of the nature of the propositional attitude, or intentional state, which properly concludes PR. The conclusion is a result of reasoning from other states (premises) such as desires/intentions and beliefs (*Figure 1*).

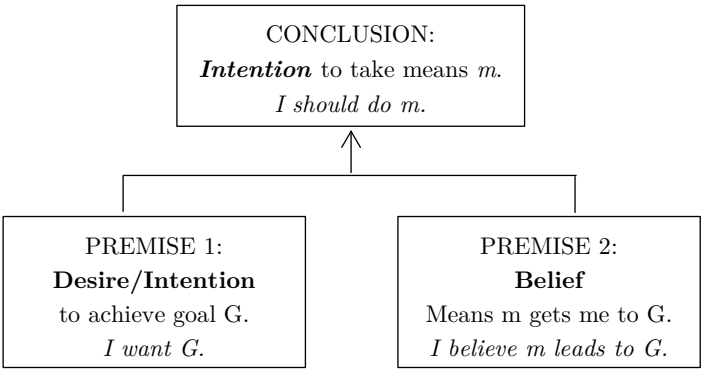

Figure 1: The basic scheme of practical reasoning

According to Searle, there is "a bewildering variety" of accounts of the elements of PR: they can be "desires, intentions, fiats, imperatives, norms, *noemata*, actions" (Searle 2001, 242)—and many of these can feature as

PR's conclusions.[2] From the weakest to the strongest conclusion, one can recognise the following continuum:

disposition to act (pro-attitude, secondary desire, practical judgment)
decision to act (prior intention)
intention to act (intention-in-action)
action itself

The crucial difference of opinion exists between those who think action itself is the proper conclusion of PR (Aristotle, *Nicomachean Ethics*; Searle 2001, 136) and those who think this is plain wrong, for reasoning is limited to propositional entities and cannot extend beyond them—therefore an intention to act is as far as we can get (Gauthier 1963; Broome 2013).

Clearly, we are dealing here with a single reasoning agent pondering over the right course of action for her to take. This would be perfectly fine if the extrapolation from the simplest unit of individual reasoning to various forms of collective reasoning was warranted. But it seems it is not. According to Hitchcock (2002), an individualistic approach is at risk of producing a "solipsistic, egoistic and unsocial" understanding of PR. Referring to Pollock's (1995) account of PR where the basic scheme of *Beliefs*, *Desires* and *Intentions* is supplemented by a reasoning agent's *Likings*, Hitchcock describes it as solipsistic, since "there is no provision for verbal input from, or verbal output to, other autonomous rational agents, still less for back-and-forth discussion, whether argumentative or non-argumentative" (Hitchcock 2002, 254). Further, "it is egoistic, in that the function of the entire system is to make the world more to the liking of that system itself" (2002, 254). As a result, "nothing […] permits rational criticism" (2002, 255) of an agent's hierarchy of desires and likings. Finally, the "model is unsocial, in that his [Pollock's] rational agent does not (and cannot) belong to any groups of autonomous rational agents with governance structures for making decisions about the actions of the group" (2002, 255). As Hitchcock concludes,

[2] Searle is clearly echoing Davidson's classic account where the pro-attitudes constitutive of PR include "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind." (Davidson 1963, 686).

"[a] comprehensive theory of good practical reasoning would have to remedy all three of these lacks" (2002, 255).

2.2. The speech act of advice

Of special importance in grasping the nature of the conclusion of PR is Gauthier's early work on various forms of representing the conclusion of PR, namely: *practical judgement*, in practical discourse. Focussing on practical *discourse*, and not merely practical *inference*, is crucial: "For it is in discourse that people actually make practical judgements, and support or criticize them" (Gauthier 1963, 50). As such, "[a]n examination of practical discourse may be expected to make clear the actual criteria used in appraising practical arguments, and hence in establishing practical judgements" (Gauthier 1963, 50). Practical discourse (practical argumentation, deliberative practices) is thus not only a display mechanism for inner practical reasoning but also an important entry point into the elements and standards of practical reasoning (Lewiński 2017).[3]

Gauthier's central idea is fairly straightforward:

> The basic practical conversation may be formulated simply: 'What shall I do?' 'Do x!' The response is most naturally put in the imperative mood, although it need not be; one might say, 'I advise you to do x', or 'Why not do x?', or 'You should do x'. […]
>
> Grammatically, the imperative mood is restricted in person to the second, and the first plural, and in tense to the present. These restrictions are of considerable importance in determining the relationship between imperatives and practical judgements, which are expressed in the indicative mood, and hence in any person and any tense. (Gauthier 1963, 50-51).

[3] Following Witek's (2021) terminology, practical reasoning/inference would have the "thinking-to-speaking direction of influence," while practical argumentation/discourse the "speaking-to-thinking direction of influence." However, Mercier and Sperber's (2011) idea that reasoning is for argumentation puts into question the very idea of purely discourse-independent reasons with a uniquely one-way, thinking-to-speaking direction of influence.

Based on this, Gauthier proposes to study five broad classes of imperatives: i) commands, orders, directions; ii) instructions and directions (sic!); iii) advice, recommendations; iv) urges, exhortations; v) requests.[4] While all these classes can be grammatically represented via the basic imperative phrase ("Do *x*!"), they are obviously distinct. According to Gauthier, the paradigmatic form is *advice*—contrary to the other imperatives of practical discourse, the very *function* of advice among men is "to assist their fellows with their practical problems" (Gauthier 1963, 77): "Advice is characteristically sought as a result of practical *concern*. Confronted with a practical problem, a man may not only make a personal judgement of what he should do, but may also ask for the judgements of others" (Gauthier 1963, 53).[5] The bulk of Gauthier's work is thus dedicated to analysing the conditions for successful advice and the intricate *differentia specifica* of advice in the broader *genus* of imperatives, esp. instructions, recommendations, hortations, commands and "moral counsel" (Gauthier 1963, esp. Chs. 4, 5, 10). In particular:

> In giving advice, and in determining what advice to give, the adviser is expected to reason from the advisee's practical basis. In advising, we treat someone else's problem from his point of view. In recommending, we consider whether our experience is relevant to recommendee's problem, regarded from his standpoint. Thus, in arguing from the situation in which advisee finds himself to a conclusion about what he should do, the adviser must take, as premisses with practical force, those which he believes to be held by advisee. (Gauthier 1963, 54-55)[6]

[4] See (Condoravdi and Lauer 2012) and (Portner 2018) for recent categorisations of imperatives which, while in many aspects similar, do not refer to Gauthier's work.

[5] This idea has clearly Aristotelian provenance: "We call in others to aid us in deliberation on important questions, distrusting ourselves as not being equal to deciding" (*Nicomachean Ethics*, 1112b11).

[6] Furthermore: "Directly hortatory terms, such as 'urge' and 'exhort', do not entail the commitment to concern with the problems of the person addressed that terms of advice imply. To urge someone to do something is to seek to move him to do it by open verbal means" (Gauthier 1963, 59); "Imperatives of advice concern practical problems of advisee. Imperatives of command do not. If you are authorized to command me to carry out certain actions, then it is inappropriate for me to consider

Two important comments are immediately necessary here. First, Gauthier does not use any of the conceptual vocabulary of speech acts—illocutionary forces, felicity conditions, etc. This is natural, given the timing of his work. Similarities are, of course, striking: Gauthier clearly analyses various *directive* speech acts (see Searle, 1975). He also specifically thanks his original supervisor at Oxford—"the late Professor J. L. Austin"—for illuminating discussions on the concept of advice. However, despite most precious insights into the nature of speech acts used in practical discourse, the lack of this specific conceptual framework is a serious limitation.[7]

Second, and perhaps an intellectually graver limitation, is Gauthier's understanding of practical reasoning as restricted to two basic forms: prudential and moral reasoning. Prudential problems, deriving from the practical base of an agent's individual wants and desires, "are essentially private, personal, not affecting others" (Gauthier 1963, 149). By contrast, "moral problems are not and cannot be purely personal, since they involve situations in which the interests of others are affected" (Gauthier 1963, 149). Using Hitchcock's terms adduced above, *prudential* discourse is thus not solipsistic (it is discourse, after all), but it is still egoistic and unsocial. Indeed, the speech acts mentioned above pertain to prudential discourse, and they are essentially private transactions between one reasoning agent (esp., advisee) and another reasoning agent (advisor), meant to solve the private practical problem of the former. In this context, advice is understandably the most central speech act. In the context of *moral* discourse, however, "advice" morphs into "moral counsel" by virtue of extending the content of the premisses of PR, namely, by including wants and interests of others beyond the advisee: "The schema which served for prudential practical reasoning is then adapted to the general case by substituting the extended basis, and the class of premisses with practical force derivable from it" (Gauthier 1963, 86). In this way, moral discourse would additionally not be egoistic, since "*all* wants of *all* persons are to be included in the

whether I shall benefit from the actions, or even whether to perform the actions. If you are entitled to command, I am obliged to obey. The responsibility for the performance of the actions is assumed by the commander." (Gauthier 1963, 61).

[7] For a recent account of advising, building on Searle's (1969, 1975) formulation of the felicity conditions of the speech act of advising, see (van Poppel 2019).

basis of [moral] practical reasoning of any agent" (Gauthier 1963, 86, emphasis in original).

However, both forms of PR studied by Gauthier remain largely unsocial. There is no room for joint, collective action and collective reasoning that is not immediately moral. Taking a walk with a friend, jointly fixing a car, or carrying a table together, are all practical problems that typically require practical reasoning and do affect others, yet do not call for moral counsel. Gauthier has very little to say about how we coordinate and resolve such issues in practical discourse. As a result, despite his undeniably relevant insights, some element is still missing.

2.3. The speech act of proposal

A good starting point towards investigating this missing element is to argue PR is a social—or, in principle, socializable—activity. This argument has quite some tradition in philosophy, as it characteristically connects PR to an argumentative activity of deliberation, a link stressed since Aristotle (EN).[8] One main consequence of it is a shift of focus away from the internal propositional attitude of intention to some externalised and collective speech act, notably, that of proposal.[9] Accordingly, the analyses of proposals have attracted some attention—especially in argumentation theory

[8] See (Lewiński 2017; 2019), (Corredor 2020), (Dascal 2005), (Green 2017), and (Walton 2006; 2007). While some authors claim that this connection is a sign that "Aristotle has confused the *psychological process* by which a person comes to resolve a practical problem with the *logical* argument in which the steps leading to the resolution are formally set out" (Gauthier 1963, 26; cf. Chang, 2016), others argue this might have been a deliberate choice in Aristotle's conception, where the process of deliberation is constitutive of practical reason, and thus in-principle social and open to the back-and-forth of argumentation (Lewiński 2017; 2019; Dascal 2005).

[9] Notice, though, that Broome, somewhat inconsistently, also speaks of speech acts which the reasoner performs to herself: "the speech-act you perform is the act of expressing an attitude of yours" (Broome 2013, 253). While expressing (propositional) attitudes, such as beliefs, desires and intentions, is one key job speech acts do (Green 2009; Witek 2021), they are communicative, rather than purely mental, acts which therefore always involve at least two parties, the speaker and the hearer, as well as the complex social commitments between them and larger groups (Lewiński 2021).

(Aakhus 2006; Corredor 2020; Ihnen Jory 2015; Kauffeld 1998; Walton 2006). Crucially, proposals shift the agent of the conclusion of PR from an individual "I" or "you" to plural "we": "so I should do m" or "so you should do m" is reformulated to "so let us do m".[10] While Gauthier was clearly aware of the fact that "the imperative mood is restricted in person to the second, and the first plural" (Gauthier 1963, 51), he focussed exclusively on the former, while having nothing to say of the latter, performed characteristically via the "let us…" construction.

Aakhus (2006) analyses proposals in deliberation as speech acts located between Searle's (1969; 1975) *commissives* (such as promises) and *directives* (such as requests) (see *Table 1*). Commissives are about future acts of the speaker who, in performing the speech act, commits her/himself to this act ("I will clean the room tomorrow"). Directives are about future acts of the hearer, whom the speaker wants to get to do something ("Clean the room tomorrow, will you?!"). Proposals concern future acts of *both* the speaker and the hearer, and their illocutionary point is "to enlist H[earer] in mutually bringing about [act] A" (Aakhus 2006, 406). They would thus be typically expressed by constructions such as "Let's (clean the room tomorrow)!" or "How about we (clean the room tomorrow)?"

According to Aakhus, "[w]hen proposing, a speaker puts forward a future act that requires a joint performance by the speaker and hearer" (Aakhus 2006, 405) and, additionally, "the speaker frames the proposed actions as mutually beneficial" (Aakhus 2006, 404). In this way, proposing is a speech act through which the conclusion of PR is put forward for consideration in the argumentative activity of deliberation: "A proposer (P) puts forward the proposal in part to get agreement but also to test for doubts and objections […] that may in turn help P design a more acceptable proposal" (Aakhus 2006, 406). Therefore, proposing belongs to the kind of illocutionary acts in which "speakers necessarily or typically incur probative burdens," that is, "a speaker cannot, other things being equal, responsibly dismiss an addressee's demands for proof" (Kauffeld 1998, 247).[11] What follows is that felicitous

[10] In Walton's formulation, the conclusion of PR in "multi-agent deliberation" is a "practical ought-statement" (Walton 2006, 204): *We* ought to do it.

[11] This condition is often, e.g. in (Pagin 2016), (Watson 2004), and (Williamson 1996), seen as a distinguishing characteristic of assertions—but it does not seem

proposals concern actions which are: 1) communicated and open for discussion, thus surely not solipsistic; 2) mutually beneficial rather than purely egoistic; 3) jointly performed, and therefore social. In this way, the analysis of proposals addresses all three concerns regarding individualistic approaches to PR identified by Hitchcock (see Sec. 2.1). However, as I will argue below—even if a paradigmatic case—proposal is only one of the possible speech acts which can convey the conclusions of PR.

Act	**Request (Searle, 1969)**	**Propose**	**Promise (Searle, 1969)**
Propositional Content	Future act A of H.	Future act A of H + S.	Future act A of S.
Preparatory Condition	H is able to do A. S believes H is able to do A. It is not obvious to both S and H that H will do A in the normal course of events of his own accord.	H and S are able to contribute to the accomplishment of A. It is not obvious to both S and H that either S or H can do A of their own accord in the normal course of events. That A will leave neither S nor H worse off than not doing A.	S is able to do A. S believes S is able to do A. It is not obvious to both S and H that S will do A in the normal course of events of his own accord.
Sincerity Condition	S wants H to do A.	S believes A will mutually benefit H and S or that if it benefits S it will leave H no worse off.	S intends that in uttering to do A he is under the obligation to do A.
Essential Condition	Counts as an attempt to get H to do A.	Counts as an attempt to enlist H in mutually bringing about A.	Counts as an attempt to commit S to do A.

Table 1: Felicity conditions for requesting, promising and proposing (Aakhus 2006, 406)

right. Quite the opposite, each speech act can be challenged and its felicity conditions tested, whereby some kind of "proof" by the original speaker must be provided. This proof would typically relate to its sincerity and preparatory conditions. This applies even to expressives: "Ouch!"—"No, this cannot really hurt, don't exaggerate." See also below, Section 2.5.

2.4. Proviso: Any speech act can conclude PR

All speech acts, including assertive speech acts, are, well, acts. They are intentionally performed human acts, based on some kind of linguistic, cultural and societal conventions.[12] As such, while they may be performed without profound deliberation—think of common expressives such as "Ouch!" or "Sorry!"—speech acts typically result from some prior judgment. That is, we need to practically reason, inside of us, to perform this and no other speech act in this very situation. In any communicative activity, we thus constantly conclude our internal deliberations with a conclusion "I should say *X* now" or "I should perform speech act of the kind *Y* (apologise, deny, object to, approve)." In this respect, there is similar PR behind commissive speech acts such as "So I shall catch the 2:30 train to London" and assertions such as "So the cat is on the mat." In Searle's words:

> There is thus a sense in which all reasoning is practical, because it all issues in doing something. In the case of theoretical reason, the doing is typically a matter of accepting a conclusion or hypothesis on the basis of argument or evidence. Theoretical reason is, thus, a special case of practical reason. (Searle 2001, 90-91)

Yet this sense is indeed special and perhaps trivial: speech acts are our intentional contributions to communicative activities and are all, in this sense, direct executions of some inner practical inference regarding our communicative involvement (roughly, our communicative plan).

Importantly, this practical communicative reasoning can be either a matter of descriptive psychology or of normative reconstruction. In descriptive studies, the behind-the-scenes working of practical inference in specifically argumentative communicative activities has been well documented by Hample (2005), Mercier and Sperber (2011), Paglieri (2013), and others. In general, forms of instrumental or strategic PR, characterised by a cost-benefit analysis of what and how efficiently one can achieve with a given argumentative contribution, have been identified. Arguers decide to perform and edit their arguments based on considerations such as chance of

[12] See (Austin 1975), (Strawson 1964), and (Searle 1969); for a recent overview, see (Harris et al. 2018).

success, identity and relation management, negative and positive politeness, situational appropriateness, as well as truth and relevance of their arguments.[13] While these results have not been cast in the language of PR about what to do, they clearly can be.

But the PR behind the performance of a given speech can also be understood in terms of normative reconstruction that can be paraphrased as: "a speaker who performs the act of *φ*-ing is to be ready to rationalize her speech act and reconstruct a practical reasoning whose conclusion would be *So I should have φ-ed.*" This reconstruction would thus reveal the structure of reasons *motivating* a speaker's performance of this, and no other, speech act. As such, it would always be an explanatory reconstruction having the structure of PR. This should not be confused with the possible reconstruction of the reasons *justifying* my performance of a given speech act. Theoretical reasons need justification "on the basis of argument or evidence" (Searle 2001, 91) and are concluded via any of the speech acts belonging to "the 'assertive' family": from mere guesses, to justified presumptions and conjectures, to strong assertions grounded in knowledge (Green 2009). Practical reasons, by contrast, need justification grounded in values, desires, preferences, and a comparative assessment of means, and are concluded via any of the action-inducing speech acts described below in Section 5.

Take a simple example: I have knowledge of the contextually required standard of a certain finding F. I am thus justified in issuing a straightforward assertion F. However, given this finding directly challenges results of a revered professor and my close colleague X, I decide to put forth my finding with the illocutionary force of a conjecture. I thus downgrade my assertive speech act for reasons of mutual respect and amicable cooperation. My conjecture is thus theoretically justified, and excessively so, by the evidence I have in hand, and practically motivated by concern for values my community (allegedly) holds dear. All this is different from performing specific speech acts *as* conclusions directly justified by PR: an issue this study focusses on.

To conclude, there is some practical inference, even if largely unconscious, behind performance of any speech act, including assertives and all argumentative speech acts. The reconstructible practical inference leading

[13] For a discussion, see Chapter 4 of (Hample 2005).

to the use of speech on a given occasion is, however, quite different from performing speech acts presented *as* conclusions of one's deliberate practical reasoning. Concluding that "Aristotle could well have written *Rhetoric* all by himself" requires very different supporting arguments than the conclusion "let's employ Aristotle."[14]

2.5. Counter-Proviso: Only assertives conclude our reasoning, whether theoretical or practical

Let's consider the following example, due to van Eemeren and Grootendorst (1984, 97-98).

Let's take an umbrella, or do you want to get wet?

It doesn't take much imagination, nor work, to reconstruct this utterance as a commonly experienced instance of PR. The speaker starts with a conclusion—"Let's take an umbrella"—that expresses an intention to act in a specific way. This conclusion is grounded in the desire—here explicitly expressed via a rhetorical question—that "we don't want to get wet." Unexpressed is the (obvious, hence the enthymematic form) belief that in a 20th-century Western society one *good* (perhaps *the best*, or at least *satisfactory*) way of not getting wet while going out is using an umbrella. Further, the conclusion here is clearly a speech act of proposal, just as defined above, involving a joint action and matching desires of both speaker and hearer (walking together in the rain, staying dry under the umbrella).[15]

Van Eemeren and Grootendorst (1984, 98), however, see this sort of interpretation as "erroneous," at least within their framework of speech acts in specifically argumentative discussions:

> The utterance 'Let's take an umbrella' should not be seen as the expressed opinion at the centre of the dispute that the speaker is trying to resolve with the aid of the utterance 'or do you want to get wet?' Rather it is a statement indicating that the speaker recognizes the possibility of a dispute arising about his proposal.

[14] I am indebted to Maciej Witek for discussion in this section.

[15] See (Gilbert 1990) for a classic account of this minimal kind of collective activity of walking together.

> This dispute might then be centred on the question of whether the proposal was a good one. The expressed opinion on which this particular dispute is centred is not formulated explicitly, but that is always possible. For example, the speaker might say: 'It is advisable to take an umbrella.' This statement is an elementary illocutionary act of the assertive type.
>
> In fully externalized discussions the expressed opinions and the argumentation must always in our view consist of elementary illocutions belonging to the class of the assertives. Expressed opinions and argumentations consisting superficially of illocutions of some other type must first be analysed in such a way that it is clear exactly what assertives are involved. If these expressed opinions and argumentations could not be construed as assertives a resolution of the dispute would be impossible, since it is only possible to resolve disputes thanks to the specific committedness associated with the performcance of assertives. (van Eemeren and Grootendorst 1984, 98, italics in original)

To understand their misgivings here, we need to grasp "the specific committedness" of assertives. According to van Eemeren and Grootendorst, "*the illocutionary point of the members of the class of assertives [is] to commit the speaker (to a greater or lesser degree) to the acceptability or unacceptability of the expressed proposition*" (van Eemeren and Grootendorst 1984, 97, italics in original). Further, this "specific committedness" results in the "obligation to defend" the acceptability of a speech act, which they consider "to be a general feature of assertives, distinguishing them as a class from other illocutionary acts" (van Eemeren and Grootendorst 1984, 96).[16]

As already mentioned, these claims are made in their task of developing a speech act-based theory of idealised argumentative discussions. What is at stake in such discussions is defensibility of a standpoint (+, -, ø; this is a conversational equivalent of a conclusion) in respect of an expressed opinion (O). Symbolically,

$+/(\mathrm{O}(p))$

[16] See also (Green 2009), (Houtlosser 1998), (Pagin 2016), (Watson 2004), and (Williamson 1996).

reads: my standpoint in respect of O is that O is the case (see van Eemeren and Grootendorst (1984, 114; '-' would mean "isn't the case", and 'ø' "I have no standpoint on the issue"). More precisely, we should add: my standpoint in respect of O is that O is an acceptable assertive with respect to its propositional content *p*. As we could see, this extends to other classes of speech acts, so for example,

+/(O(P(*p*)))

would read: my standpoint in respect of the expressed opinion O that the proposal P to take an umbrella (*p*) is good, is that O is an acceptable assertive with respect to the Proposal's being good.

Three arguments can be lodged against this, otherwise interesting, reconstruction of the conclusions of PR. First, each and every speech act would, at least for the purposes of argumentation analysis, be, eventually, an assertive.[17] A reduction of all speech acts to assertives is arguably not the claim van Eemeren and Grootendorst propound. Instead, they seem to defend a weaker claim that when we conclude an argument by making an utterance that in a non-argumentative context would be naturally taken to constitute the performance of a non-assertoric act only, the actual function of the utterance is that of making an assertion about the non-assertoric illocution in question. But even this weaker claim undermines the "central insight" of speech act theory, namely, that "language is a medium for many kinds of action, but its superficial uniformity tends to mask this fact" (Harris et al. 2018, 1). Indeed, van Eemeren and Grootendorst's definition of assertives seems to point to a very "general committedness" of speech acts, rather than "specific committedness" of assertives: all speech acts have satisfaction conditions (truth con-

[17] This should not be confused with Bach's (1975) argument that "performatives are statements too" (note the "too"), on the ground that performative utterances self-referentially *state* their illocutionary force and the vehicle for that force, the very performative utterance itself. This can be made explicit by adding an assertoric formula such as: '[In uttering this sentence,] I order you to leave' (Bach 1975, 234). As such, speech acts other than assertives "comprise two simultaneous illocutionary acts" (Bach 1975, 229); for further discussion, see also (Searle 1989), (Bach and Harnish 1992), and (Reimer 1995).

ditions, fulfilment conditions, "answerhood" conditions) and as such all obligate the speaker to defend these conditions when challenged, thus justifying the speech act's acceptability. Second, their approach is prone to infinite regress: I Propose *p*, and my expressed Opinion is that P is a good proposal, and my Standpoint is that my O is acceptable and… why not: Claim that my Standpoint is correct, and then Endorse the Claim, and Assert the Endorsement as true, etc.? Third, any "assertorification" of directive and commissive speech acts—proposals, recommendations, (moral) imperatives, pieces of advice—following the 'It is advisable to take an umbrella' gloss would need to seriously tackle the issue of moral descriptivism / realism. "It is true that 'it is not good to kill people'" as a gloss of "Thou shalt not kill" carries a heavy burden of proof that a simple statement that "ethical, aesthetic or other normative statements" (van Eemeren and Grootendorst 1984, 96) can be treated much in the same way does not quite carry.

Shortly, the idea that PR (just as any other form of linguistically expressed reasoning) always concludes in some kind of assertive is not exactly defensible. It is also rather cumbersome, and not very speech-acty (see also Jacobs 1989). There is, however, an important intuition in van Eemeren and Grootendorst's interpretation of *proposals*: they should be defensible as "good" proposals. This, I argue in the next section, is an inherent feature of PR.

3. Detailed scheme of practical argumentation

Before examining the various forms of conclusion of PR—which, when publicly performed, can better be called practical argumentation (henceforth: PA)—it seems necessary to understand what PA in general consists of.[18] The scheme of PA presented in *Figure 2* stems from a rich literature on practical argument in philosophy, Artificial Intelligence, and argumentation theory (Audi 2006; Broome 2013; Davidson 1963; Gauthier 1963; Pollock 1995; Searle 2001; Walton 2006; 2007; see Lewiński 2015; 2017 for a more detailed

[18] The *publicity* element is more than a mere mode of presentation. Publicity of practical arguments invokes socially and institutionally recognizable commitments and, as such, can generate "desire-independent reasons for action" and forms of collective, rather than only individual, intentionality (Searle 2001; 2010).

discussion). In particular, it is derived from a useful representation of PA by Fairclough and Fairclough (2012). While referring to their work for an analysis of all the premises constituting the scheme (*Circumstances*, *Goal*, *Values*), I will briefly mention five basic advantages of the scheme, focussing further on the last one: the speech acts which can conclude PA.

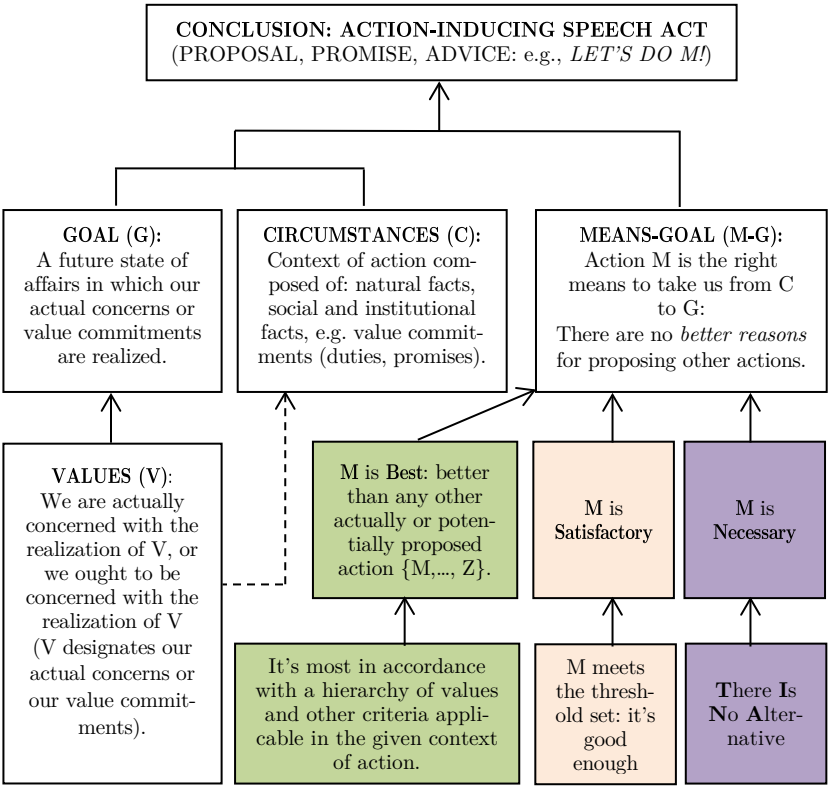

Figure 2: The scheme of Practical Argumentation
Derived from: Fairclough and Fairclough (2012)

Possible, contextually-determined decision criteria:
direct or indirect costs and benefits (negative and beneficial consequences / side effects) •
opportunity costs •

practical feasibility •
ethical, moral, or legal implications •
likelihood of realization or of success •
conformance with other goals or strategies, their timing, duration, or location •
Derived from (McBurney et al. 2007, 99)

First, the scheme shapes the framework of relevance for (multi-party) deliberation. Typically, different parties argue for the contextual betterness of their proposals for action {M, N, O… Z} (see the "M is Best" box). Their deliberation develops then as an *argumentative polylogue* (Lewiński and Aakhus 2014; Lewiński 2017; 2021) along the lines of possible disagreements over the various elements of the structure (basic premises, inference rules and contextual criteria).

Second, the scheme distinguishes between context-independent and context-dependent elements of PA. Its basic general structure (as per Fairclough and Fairclough: all the white boxes in *Figure 2*) remains constant, while contextual criteria for choosing "the right means" (below the diagram) fluctuate.

Third, the scheme clarifies the notion of the means-goal premise. This premise is grounded in one of the three basic inference licences warranting the choice of "the right means" taking us from the current (unwelcome) *Circumstances* to the (desired) *Goals*. We can thus warrant our conclusions by issuing a comparative claim (Chang 2016) that either the means are necessary, or that they are satisfactory (good enough), or that they are the best among all the possible alternatives.[19]

Fourth, it provides a new account of how to criticise and evaluate PA. It clearly demarcates the attacks on the main premises (Circumstances, Goals, Values, Means-Goals), from those concerning context-dependent decision criteria, and from criticisms of the inference licences related to the type of inferential step made (necessary, satisfactory, or the best means) (see Lewiński 2017 for further discussion).

Fifth, the scheme defines its conclusion in terms of a class of action-inducing speech acts, thus pointing to a "unity in diversity" of what we can argue for in PA.

[19] For a detailed analysis, see (Lewiński 2015) and (Lewiński 2017).

I will now apply this scheme to the analysis of an actual case of complex practical argumentation and then discuss in detail the last point.

4. Case study: Keep it in the ground!

On the 16th of March 2015, the British newspaper *The Guardian* launched a media campaign to divest (dis-invest) from fossil fuels such as coal, oil and gas. The campaign was entitled "Keep it in the ground" and included, at its start, a very short petition, a "note from Alan Rusbridger, [*The Guardian's* then] editor-in-chief", a "full story" in which Rusbridger presents an elaborate argument for the petition, as well as other multimedia materials (videos, frequently asked questions, interviews, reportage) related to the dangers of our massive continuing reliance on fossil fuels.[20] These texts, while originating in the same source, contain various examples of illocutionary acts concluding practical argument.

Let me start with a little note on the very slogan of the campaign: "Keep it in the ground." As an utterance in an imperative mood, it seems to be addressed to others as a directive speech act explicitly identified as (a part of) "petition" (a formal request? Appeal? Plea?). However, given the nature of such campaigns, and the immediate contextual and co-textual information (see below), it can also be understood as elided "(*We* should) keep it in the ground" or, even better, "(Let *us*) keep it in the ground"—by analogy with "(Let's) give peace a chance" and other such slogans. Here, the addressee—the agent of PA—is the first plural, inclusive we.

In what follows, I focus on Rusbridger's "full story" entitled: "The argument for divesting from fossil fuels is becoming overwhelming."[21] This "story" or "argument"—in fact, a well-structured complex argument—is introduced with the following lead:

[20] See http://www.theguardian.com/environment/ng-interactive/2015/mar/16/keep-it-in-the-ground-guardian-climate-change-campaign, last consulted on the 6th of January 2021.

[21] See https://www.theguardian.com/environment/2015/mar/16/argument-divesting-fossil-fuels-overwhelming-climate-change, last consulted on the 6th of January 2021. All following quotations are from this source.

> As progressive institutions, the Gates Foundation and Wellcome Trust should commit to taking their money out of the companies that are driving global warming, says the Guardian's editor-in-chief as he launches our climate campaign.

As is appropriate for a journalistic article, it starts with a conclusion. Here, it is clearly a conclusion of a PA—that some agents (the Gates Foundation and Wellcome Trust) "should" undertake certain actions ("should commit to taking their money out of the companies that are driving global warming").

According to Rusbridger, "[t]here are two arguments in favour of moving money out of the biggest and most aggressive fossil fuel companies—one moral, the other financial." In saying this, he explicitly refers to two basic types of PR recognised in philosophy: on the one-hand, *moral* or *value-based* reasoning, and on the other hand, *instrumental*, *prudential*, *strategic* or *means-end* reasoning.[22] The moral argument is analogous to "the push to pull money out of tobacco, arms, apartheid South Africa—or even slavery." Investing big money in the fossil fuel business, even if profitable, is *per se* a bad thing to do—just like making money from arms or slave trade is. The chief value in the fossil fuel argument is intergenerational justice, that is, "concern for future generations": through our current recklessness, we are burdening the future generations with all the negative consequences of climate change (see Campos 2018). *The Guardian's* financial (or "pragmatic") argument is, interestingly, much more profoundly argued for, likely with the view towards the target agents of change, financial managers. This argument is best summarised in a quote from the Bank of England's deputy head of supervision for banks and insurance companies, Paul Fisher: "As the world increasingly limits carbon emissions, and moves to alternative energy sources, investments in fossil fuels—a growing financial market in recent decades—may take a huge hit." That is to say, those who do not care about the moral implications of climate change per se, or even do not believe in it at all, might be driven out of further investment by prudential risk-assessment.

[22] See (Gauthier 1963), (Fairclough and Fairclough 2012), and (Walton 2007).

The "Keep it in the ground" campaign—as is appropriate for any other campaign—has a precise target: managers of endowments, investment and pension funds and, very specifically, the Gates Foundation and Wellcome Trust. It is clear in the text, though, that the problems of climate change policies and energy models based on fossil fuels have a number of other relevant stakeholders (see Rodrigues, Lewiński, and Üzelgün 2019). All of them can potentially be the target of the campaign, the addressees of its arguments and, eventually, the agents of the conclusion of the argument. Rusbridger mentions that at first *The Guardian* thought of addressing "governments" and "politicians" in general ("MPs, presidents, prime ministers and members of congress")—but political action on climate change has proven far from satisfying: "the people who represent us around the UN negotiating tables have moved inches, not miles." That's why they decided to address the above-mentioned big institutional investors instead. Another stakeholder in the "story" are the scientists, here endowed with authority and treated with reverence: "If only science were enough […] finance will eventually have to surrender to physics […] the physics is unarguable." Of course, the general public is another crucial stakeholder—here, it is divided into present and future generations, with the responsibilities laid on those who can decide and act now. Finally, there are the fossil-fuel companies' directors, who—by the logic of global capitalistic economy—are compelled to "behaviour that is overwhelmingly driven by short-term returns."

We can pretty straightforwardly reconstruct the complex structure of PA from Rusbridger's argument which, in this case, is remarkably explicit.[23] The current *Circumstance* is that of a climate change crisis caused by the overreliance on fossil fuels. This premise is briefly stated in the very first sentence of the piece: "The world has much more coal, oil and gas in the ground than it can safely burn." To address, or at least attenuate, this crisis we need to strive for a concrete *Goal*: "80% of the known coal reserves will have to stay underground, along with half the gas and a third of the oil

[23] As correctly noted by one of the anonymous reviewers, many public arguments have implicit premises to reconstruct—a major challenge for argumentation analysis—while "some instances of practical reasoning in the public sphere have explicit premises and implicit conclusions (the encouragement for an action is conversationally or conventionally implied)." This complication does not directly affect this very case.

reserves." Attaining this goal is necessary "if we and our children are to have a reasonable chance of living stable and secure lives 30 or so years from now." In general, the main *Values* and duties underlying the entire argument are "concern for future generations," "the protection of the public," "human health and science." But which means can be conducive to reaching this *Goal* and embodying the *Values*? As already mentioned, in *The Guardian's* view the best action to take ("the best" *Means-Goal* premise) is to directly call to divest from major fossil fuel companies—rather than, for instance, "campaigning for a paragraph to be inserted into the negotiating text at the UN climate talks in Paris this December" (see Lewiński and Mohammed 2019). Appropriately for deliberations, different options are thus considered and one is chosen as *the* option.

Finally, let us look at the *Conclusions* of the argument. These are explicitly indicated with the "so" connective:

> So we ask that the Gates Foundation and Wellcome Trust commit now to divesting from the top 200 fossil fuel companies within five years. And that they immediately freeze any new investment in the same companies.

This, however, is not the only conclusion of the text: "We will, of course, suggest that the Guardian Media Group does the same, and keeps you informed about its own deliberations and decisions." Finally, the readers are requested to act on the argument too: "Please sign, retweet and generally spread news about the petition."

Shortly, we have here a rather motley assortment of conclusions, at least when compared to a typical first-person-singular conclusion of most PR, as examined in philosophy ("so I shall take the 2:30 train to London"). What can be done about it?

5. Conclusions as speech acts

I have already argued that a speech act approach can bring about the required level of nuance to our understanding of the conclusions of PA. Indeed, Austin's (1975) initial distinction between fact-relevant *constatives*

and action-relevant *performatives* can be understood against the background of the distinction between theoretical and practical uses of reason: we constate on the basis of our theoretical reasons and we perform of the basis of our practical reasons.[24] Importantly, while the constative-performative distinction does not mirror differences between assertive and non-assertive speech acts as later elaborated by Austin himself and many others, e.g., Urmson (1977), there is a sense in which the dichotomy between theoretical and practical reason permeates the two. "Assertion and its cousins," such as conjectures, presumptions or guesses are all truth-relevant speech acts, whose norms require a psychological state of belief or at least acceptance, justified by evidence of an adequate sort.[25] As briefly put by Stalnaker, "[t]o accept a proposition is to treat it as true for some reason" (2002, 716), and this reason is theoretical reason. Some of the non-assertive speech acts, by contrast, are justified by practical reason and performed as its conclusion.

Against this background, the central question of this paper can be formulated as follows: Which speech acts "convey" or "express" the conclusion of practical reasoning (PR) externalised in a text, that is, of practical argumentation (PA)?

As a preliminary, consider Searle's claim that "to function in deliberation a reason must be for a type of action, it must be for the agent, and it must be known to the agent" (Searle 2001, 99). PA, being public, takes care of the last condition all by itself: the reasons are openly presented so—as long as the "normal input and output conditions" for communication obtain (Searle 1969, 57)—they are known to the agent. What remains to be

[24] As is well known, Austin in fact did not endorse this distinction: his work instead shows that he either abandoned it in favour of the analysis of the locutionary/illocutionary/perlocutionary acts or, as Sbisà (2007) argues, from the very beginning presented it rather as a premise of a *reductio ad absurdum* argument: assuming that the constative-performative dichotomy holds leads to absurd consequences (e.g., statements too function as performatives leaving the set of pure constatives empty), so it cannot be right.

[25] Strong for assertions, weak for guesses; for a discussion, see (Green 2009) and (Witek 2021).

analysed are the type of action and the type of agent that practical argumentation is "for."

For Fairclough and Fairclough (2012), the conclusion of PA has the form: "Claim for action: Agent (presumably) ought to do A." In their discussion of this point (Fairclough and Fairclough 2012, 45-46), the Faircloughs notice that the first-person-singular analyses of PR opportunistically conceal the fact that the arguer and the agent of action are not always the same. We can, and often do, argue practically with someone else's action in the conclusion ("All things considered, the best thing to do now is for you to go and apologise"). Shortly, the philosophers' *agent-self* (Searle 2001, 99), can be a self different than the arguer-self.[26]

The scope of the claim for action is further complicated by formulations such as "should" or "ought to." These are notoriously ambiguous, even underdetermined—if not semantically, then surely pragmatically.[27] In everyday spoken language we manage this ambiguity very well trough skilful, even if unconscious, use of prosody and contextual information. Compare:

> You should *just* let him go – a friendly suggested advice.
>
> You should let him go *please* – an appeal or entreaty.
>
> You should let him go *right now* – a command.
>
> (Note that in spoken discourse, the illocutionary force equivalent to "just," "please," and "right now" can be conveyed solely by prosodic elements, such as rising/falling intonation or accent.)

Even in such simple examples the pragmatic analysis of expressions such as "should" leaves the notion of a "claim for action" largely underspecified. The analysis in the previous section likewise identified that, quite tellingly, *The Guardian* editors conclude their arguments with varied "claims for actions" by agents other than themselves. What exactly can this "claim" be?

As analysed above, in the case of the conclusion of PA, the speech act of *proposal* has been considered a paradigm case in (Aakhus 2006), (Cor-

[26] This limitation to the agent-self has clear Aristotelian roots: "Deliberation is about the things to be done by the agent himself" (*EN*, 1112b32).

[27] See esp. (Portner 2007) and (Portner 2018).

redor 2020), (Ihnen Jory 2016), (Kauffeld 1998), and (Walton 2006). Proposals are speech acts located between *commissives* (such as promises) and *directives* (such as requests).[28] Due to this peculiarity they belong—next to bets, offers and bids—to a class of speech acts inescapably requiring an action of *both* the speaker and the hearer. Taken together, commissives, directives, and their hybrids (such as proposals or offers) form a class of what I will call *action-inducing speech acts* (see *Table 2*)—in opposition to assertives (representatives), expressives and declarations. Action-inducing speech acts are characterised by their world-to-words direction of fit (Searle 1975),[29] as their point is to get an agent (whether "I", "you", or "we") to

[28] Searle and Vanderveken treat proposals, somewhat in passing, as instances of directives: "Thus, for example, the illocutionary force of accepting (a proposal, an invitation, etc.) has one more preparatory condition than the illocutionary force of commitment to a future action, namely that *the speaker has been given a directive that allows for the possibility of refusal*" (Searle and Vanderveken 1985, 66, italics added). For reasons described in Section 2.3 this doesn't seem right. That proposals do more than direct others to do something is clear in that they are, inherently, also conditional promises (just like bets and offers): Let's go for a walk = You should go for a walk + If you do, I will come join you. Requests and invitations, discussed by Lance and Kukla (2013, 461), do not have this *conditional commissive* component, but at most some weak *conditional expressive* component: Do something for me + If you do, I'll be happy (the terms are mine). It is precisely this addition of a conditional commissive component that turns acts such as requests or invitations into offers and proposals. Compare a) with b):

Parent to a child:

a) Clean your room, please!

b) Clean your room – and I'll let you play games the whole night!

If the speech act is perlocutionarily successful (i.e., the child cleans the room), the child's natural response to b) would be: "I've cleaned the room, can I turn on my PlayStation now?" If the parent hesitates ("Well…"), the child's reaction would almost certainly be "You promised!"

[29] This, of course, has not escaped Searle's attention: "Since the direction of fit is the same for commissives and directives, it would give us a more elegant taxonomy if we could show that they are really members of the same category. I am unable to do this […] and am left with the inelegant solution of two separate categories with the same direction of fit" (Searle 1975, 356); see also (Searle 2001). Lance and Kukla (2013) characterise promises as "reverse imperatives", since both these types "oblige

perform an action that will bring the world into a state captured in the intentional content of the speech act. (By contrast, the goal of assertive speech acts with the words-to-world direction of fit is to capture in their content some existing state of the world.)[30] Naturally, so to speak, it is exactly speech acts with an "upward" (world-to-words, world-to-mind) direction of fit that PR is "about." Given their different direction of fit, the similarity between practical and theoretical reason breaks:

> The difference between theoretical and practical reason is in the direction of fit of the conclusion: mind-to-world, in the case of drawing a conclusion from evidence or premises, and world-to-mind, in the case of forming a decision and hence an intention on the basis of considerations. (Searle 2001, 91)

Now, the elements of the class of action-inducing speech acts can be distinguished, as discussed above, along two dimensions: (1) their primary agent: the speaker, the hearer[31], or both; and (2) their illocutionary strength, ranging from the cancellable and nearly off-record to fully endorsed and on-record.[32] Together, these two axes create an ordered matrix of "claims for action"—or speech acts which can conclude PA (*Table 2*).

---

someone to do something for someone" (Lance and Kukla 2013, 458-459, fn. 3). In his proposal challenging the very bases of Searle's taxonomy of speech acts, Hanks (2018) treats all the speech acts with a world-to-word direction of fit as a "unified category" with an "imperative propositional content," regardless of whether they involve commitments of a speaker (as in promises) or a hearer (as in orders). See (Green 2018) for a challenge to Hanks's challenge.

[30] Note that for Searle (1975) expressives have no direction of fit at all, while declarations have dual direction: both words-to-world and world-to-words.

[31] For the current purposes, it is inconsequential to carefully distinguish between the directly addressed Hearer ("you") and some potential, indirect Hearers or bystanders, such as third parties ("s/he", "they"). These differences do not affect the analysis proposed here—they do, however, play a significant role in understanding how speech acts function in a multi-party (n>2) context; see (Clark and Carlson 1982), (Levinson 1988), and (Lewiński 2021).

[32] Jacobs and Jackson (1983) study some weak directive speech acts (such as "indirect requests," "hints and prompts" and even seemingly "innocent remarks") and place them "on a continuum according to the degree to which the act-type is dissociated from the illocutionary force of a [direct] request" (Jacobs and Jackson 1983, 285).

Strength / Agent	Weak (suggesting)	Neutral (committed)	Strong (solemn)
I / We (exclusive) *commissive*	Declaration (of intention)	Promise	(Joint) Vow, Oath, Pledge, Guarantee
You / They *Directive*	Recommendation, Advice, Invitation	Request, Appeal, Warning, Instruction	Order, Demand, Plea, Call, Entreaty
We (inclusive) *commissive/directive*	Suggestion	Proposal, Offer	Bid, Contract, Call

Table 2: Action-inducing speech acts as conclusions of practical argumentation

What is common to these various action-inducing speech acts is, of course, that they can all be complemented with the phrase "to do A": "I promise to do A," "I advise you to do A," "we guarantee to do A." Their strength—here divided for illustrative purposes into three levels—can quite straightforwardly be grasped by our ordinary intuitions: "I've thought about it a lot and: maybe we can go and apologise? / let's go and apologise / we must go and apologise."

Given these distinctions, we can now more precisely characterise the explicit conclusions of *The Guardian's* campaign: First, its very title—"Keep it in the ground"—in the interpretation suggested above, takes as an agent the *inclusive we*. It is, in terms of speech acts, a joint *proposal* or *call*: neutral-to-strong combination of a directive (you should) and a commissive (we should too).[33] Second, the main conclusion is addressed to "them"—the

[33] While paradigmatically calls would be "second-person calls" (Lance and Kukla 2013), they can also be "inclusive we" calls, such as in "I call on all of us gathered

Gates Foundation and Wellcome Trust—and explicitly identified as a (second-person) "call," a strong directive speech act. Third, "[we] the Guardian Media Group"—the *exclusive we* (us)—concludes the argument with a declaration of intention or "suggestion," a weak commissive speech act. Finally, "you," "the readers" are requested to spread the news. This is a conclusion expressed through a directive speech act, of a rather neutral force (it is neither only suggested nor strongly demanded or called for).

6. Extensions and challenges

In this paper I mostly did two things. I asked which illocutionary acts "convey" or "express" the conclusion of practical reasoning and I responded these are various action-inducing speech acts. This broad category of acts shares the world-to-word direction of fit and can be preliminarily organised by the agent of the action induced (not to be confused with the concluder of the practical argument) and the illocutionary strength of the act. But between these two things I also reviewed various other takes on what the nature or function of the conclusion of practical argument could be (to intend, to advise, to propose, to state), proposed one possible scheme of practical argument, and briefly analysed some naturally occurring instances of practical argument with their varied conclusions. In this final section, I start dealing with some possible challenges to the account offered, but with an eye towards useful extensions.

The most common formulation of the conclusion of practical argument contains the somewhat enigmatic expression "I should" / "I ought to" perform a certain action. In this way, PR ends with something like a moral obligation or imperative of the individual agent-self. Accordingly, it is the "I / exclusive we" form that constitutes the traditional area of philosophical investigation into practical and moral reasoning. Both the actions and obligations of single and, more recently, collective subjects are treated in detail far exceeding this work. However, for speech act theorists such as Searle, reasons for action and their conclusions are best understood not through

here to exercise maximum restraint." In this latter case, they are pragmatically equivalent to "We [inclusive] must exercise maximum restraint."

some external ethical systems containing moral obligations "I" or "we" should fulfil, but rather through "the human ability to create, through the use of language, a *public* set of commitments" (Searle 2001, 183, emphasis in original). This set consists of the consequences of various speech acts we perform. It is therefore the analysis of speech acts that can illuminate the structure and conclusions of our *public* practical argumentation.

Throughout this paper I have largely followed Searle's (1975) well-known taxonomy of speech acts into five broad categories (assertives/representatives, commissives, directives, expressives and declarations) rather than Austin's (1975, Ch. XII) original taxonomy (verdictives, exercitives, commissives, behabitives and expositives). Indeed, one might argue that Austin's exercitives are particularly relevant here[34]:

> An exercitive is the giving of a decision in favour of or against a certain course of action, or advocacy of it. It is a decision that something is to be so, as distinct from a judgement that it is so: it is advocacy that it should be so, as opposed to an estimate that it is so; (Austin 1975, 154)

Many of the illocutionary acts discussed here—orders, pieces of advice, recommendations, entreaties, pleas—are listed by Austin in this category.[35] However, in line with his incipient institutional ontology, Austin delineated a somewhat more specific category of largely "legislative or executive acts": vetoes, appointments, excommunications, proclamations, authorizations, etc. (Austin 1975, Ch. XII). As such, however, the class of exercitives can

[34] Austin's commissives have been unreservedly appropriated by Searle (1975, 356), so this is not a point of concern. Further, "I conclude…" is obviously a discursive expositive on Austin's account, a type of a speech act instrumental in "the clarifying of reasons, arguments, and communications" (Austin 1975, 162). As is clear from my arguments, I did not aim at this sort of metalinguistic characterization here.

[35] In a similar vein, Corredor (2020) classifies the act of collectively "accepting a proposal" in a deliberation dialogue as an Austinian exercitive, while maintaining that (an earlier) speech act of "making a proposal" belongs to the class of verdictives. Since her work was published after this paper was originally submitted, I have no space to fully engage with her account. However, given Austin's definitions, it seems the reverse might be correct: making a proposal seems to be an exercitive, while (officially) accepting it is a verdictive.

be justifiably criticized for lacking a straightforward principle of classification and, hence, for being too heterogeneous, with many examples defying the very definition of the class: "nominating, appointing, and excommunicating are not the 'giving of a decision in favor of or against a certain course of action,' much less are they 'advocating' it. Rather they are, as Austin himself might have said, *performances* of these actions, not *advocacies* of anything" (Searle 1975, 353). Given this, Searle counts them among declarations, with all their peculiarities regarding double direction of fit, institutional preparatory conditions, void sincerity conditions, etc. Altogether, Searle's organizing effort on Austin's original ideas works well precisely in the context where some clear distinctions are needed as a background for further work.

In this context, recent work on imperative speech acts, while challenging some of the simplifications of traditional speech act analysis, sheds additional light on the "you / they" category of agents. In particular, Lance and Kukla (2013) have examined the class of speech acts they simply call "second-person calls," whose prime examples include imperatives (orders, commands), requests and entreaties. Their basic argument is that second-person calls can hardly be regarded as belonging to a unified class of directives or imperatives, characterised by the same illocutionary point—namely, to get or even oblige someone to do something—and differing mostly in strength, from meek invitations to forceful orders. Rather, Lance and Kukla maintain, "there are indefinitely many kinds of calls, with distinctive structures, whose subtleties help to constitute a rich moral and social space" (Lance and Kukla 2013, 458). According to them, especially between imperatives and requests there exist "deep differences":

> It is in the nature of imperatives that they impute obligations, and in doing so, they do not present their target with a choice of whether to obey them. […] [W]hile all calls give their targets reasons to act, different kinds of calls create different kinds of reasons, and these kinds often cannot be understood except in relation to the types of second-personal transactions that institute them. We may give the name "petitionary reasons" to the distinctive kinds of reasons created by requests—that is, those that give the one requested the right kind of reason to act that opens

> up the right kind of space of freedom, pressure, and so forth. […] An imperative is structurally incapable of giving its target a petitionary reason to act. Petitionary reasons are not just weak obligations, nor are they obligations backed up by weak desires on the part of the requester; they are a different variety of reason altogether. (Lance and Kukla 2013, 462)[36]

Further, requests and invitations can also be distinguished by their pragmatic subtleties: requests amount to a favour done by the requestee to the requester, while invitations appeal to matching desires of the invitee and inviter. As a consequence, successful invitations call for gratitude from both parties ("Thanks for inviting me!"—"Thanks for coming!") whereas granted requests only from the requester—the recipient of a favor (Lance and Kukla 2013, 461).

These arguments make it obvious that a unidimensional continuum ordering action-inducing speech acts based on their illocutionary strength is merely a heuristic appealed to for the sake of the economy of exposition. As described in section 2.2, this was also the basic insight of Gauthier (1963), who without using speech act concepts such as illocutionary force or illocutionary strength, analysed the subtle differences among various directive speech acts much in the same way Lance and Kukla, and others, recently do.[37] Other concerns beyond mere strength of the illocutionary point, such

[36] In their account of imperatives, Condoravdi and Lauer (2012) even more emphatically criticize the view according to which "the addressee of an imperative automatically becomes committed to making the content of the imperative true." As they observe: "While this may be right for order uses, which intuitively create hearer obligations, most other uses of imperatives, even other directive uses, do not (directly) induce hearer commitments. A crucial feature of requests, pleas, warnings, etc. is that they do not create addressee commitments (though they may be uttered in the hope that the addressee takes on a commitment)." (Condoravdi and Lauer 2012, 55). Still, as observed by Lance and Kukla, and in line with the traditional speech act theory, unless the speech act is infelicitous, the hearer is expected to respond to an imperative, and the preferred response is for the hearer to get committed (grant a request, accept an invitation) or refuse commitment with a good justification ("Sorry, I can't make it, because…").

[37] It is only a pity, then, that this work is completely ignored, including Gauthier's ideas connecting directive speech acts to practical argumentation.

as the mode of achievement (command vs. request) or degree of strength of the sincerity conditions (beg vs. request), can and should be factored in for a more nuanced taxonomy.[38] For now, however, the crucial point is to indicate that our practical argumentation often concludes in an appeal to or a call for an action of a second person. In this way, despite all their intriguing differences, second-person calls form a category homogenous enough for the current purposes.

The current proposal also clearly allows for the practical reasoning from the second / third person's desires and goals. This might resolve some of the apparently Moorean paradoxes of directive speech acts, such as those discussed by Condoravdi and Lauer (2012):

> #Call him at home! I don't want you to but he is fine with that.
>
> #OK, go to Paris then since you want it so much! But, don't forget, I don't want you to.

These examples seem to demonstrate that "it is not felicitous to follow an imperative with an assertion that the realization of the content goes against the speaker's desires" (Condoravdi and Lauer 2012, 41-43).

However, such cases can be made felicitous, as long as the structure of PR is relative to hearer's desires.[39] Good advice is often idiomaticised in English to "*If I were you*, I would never talk to him again." There seem to be no pragmatic inconsistency whatsoever in adding "although, as you know, I myself actually quite hate that kind of thing." But even in the case

[38] "Mode of achievement" and "degree of strength of the sincerity conditions" are among the seven basic factors distinguishing various types of speech acts in Searle and Vanderveken's work. The institutional mode of authority is needed for commands, and a speaker who "*begs, beseeches,* or *implores,* […] expresses a stronger desire than if he merely requests" (Searle and Vanderveken 1985, 19). Lance and Kukla are therefore partly wrong in attributing ignorance of these additional factors to Searle's typology of speech acts.

[39] Or, even stronger, as Gauthier argues, this is the very basic condition of an advice: "It would clearly be impossible for the adviser to substitute his own practical basis in giving advice. For, from premisses about the situation, plus premisses about *my* aims, nothing follows at least prudentially, about what *you* should do" (1963, 54); cf. (van Poppel 2019).

of straightforward grammatical imperatives, reasoning from the second person's desires conflicting the speaker's desires *is* pragmatically possible:

> Parent to an early-adult child:
>
> OK, go to Paris then, if you want it so much! But don't forget I've been against it all along![40]

Overall, however, the current proposal is compatible with much of the recent work on imperative speech acts. This is the case with Condoravdi and Lauer's approach who treat imperatives—including their various illocutionary variants, from orders, pleas and requests to invitations, permissions and advice—as expressions of effective preferences of the speaker. These *effective preferences* are sets of possible choices structured in a consistent way from the best to the worst—a condition necessary if "the [rational] agent is to decide on a course of action" (Condoravdi and Lauer 2012, 45). These elements are included in the model of PA presented above in Section 3. Indeed, while for Condoravdi and Lauer the utterers of imperatives are committed to their effective preferences and beliefs, in the model suggested here they are committed to all the premises of the PA, including their choice of the best, or at least satisfactory, action-inducing speech act.

According to Portner (2007; 2018) imperatives are linguistically performed via *priority modals* of three varieties: deontic, bouletic and teleological. As he puts it, "The idea behind the term 'priority' is that some choice is given priority over another" (Portner 2007, 355). Importantly, various subtypes of imperatives can be distinguished based on the ordering source: orders are deontic as they involve obligations of the hearer, invitations are bouletic as they appeal to the desires of the hearer, and suggestions are

[40] Arguably, in this case a "permissive" rather than "directive" form of grammatical imperative is used, as argued in (Condoravdi and Lauer 2012), whereby the parent is reluctantly permitting the adult child to do something that s/he as a parent can no longer prohibit. Still, the same form of Moore's paradox applies here, based on the generalization that "It seems that there *is* a bouletic component conventionally associated with imperatives. For if it were not conventional, we would expect this constraint to be absent in scenarios in which the speaker can be assumed to not share the goals of the addressee, as in disinterested advice uses." (Condoravdi and Lauer 2012, 42).

teleological in that they refer to the goals of the hearer. Again, the structuring of acts based on their priority in relation to desires, goals and obligations is a central element of the approach proposed here.

This is not to deny that what I offered here is hardly more than a simple matrix ordered by the agent of the action and the illocutionary strength of a speech act. Even this simple matrix, however, can help us better understand what practical argumentation is "about," that is, "why" or what "for" we argue practically. We argue to issue various speech acts: from innocuous private announcements and suggestions to strong commands and solemn joint pledges. Further careful analysis of all such speech acts, in their natural context, with their respective felicity conditions and further consequences, can sharpen our understanding of practical argument.

Therefore, we should do it.

References

Aakhus, Mark. 2006. "The Act and Activity of Proposing in Deliberation." In *Engaging Argument: Selected papers from the 2005 NCA/AFA Summer Conference on Argumentation*, edited by P. Riley, 402-408. Washington, DC: National Communication Association.

Aristotle. 1984. *Nicomachean ethics*. In *The Complete Works of Aristotle, Vol.* II, edited by Jonathan Barnes, 1729-1867. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400835850

Audi, Robert. 2006. *Practical Reasoning and Ethical Decision*. New York: Routledge. https://doi.org/10.4324/9780203015681

Austin, John L. 1975. *How To Do things With Words*. 2nd ed. Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198245537.001.0001

Bach, Kent. 1975. "Performatives are Statements Too." *Philosophical Studies* 28: 229–236. https://doi.org/10.1007/BF00353970

Bach, Kent, and Robert M. Harnish. 1992. "How Performatives Really Work: A Reply to Searle." *Linguistics & Philosophy* 15: 93–110. https://doi.org/10.1007/BF00635834

Bratman, Michael E. (2014). *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199897933.001.0001

Broome, John. 2013. *Rationality Through Reasoning*. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118609088

Campos, Andre S. 2018. "Intergenerational Justice Today." *Philosophy Compass* 13 (3): 1–12. https://doi.org/10.1111/phc3.12477

Chang, Ruth. 2016. "Comparativism: The Grounds of Rational Choice." In *Weighing reasons*, edited by Errol Lord and Barry McGuire, 213–240. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199315192.003.0011

Clark, Herbert H., and Thomas B. Carlson. 1982. "Hearers and Speech Acts." *Language* 58 (2): 332–73. https://doi.org/10.2307/414102

Condoravdi, Cleo, and Sven Lauer. 2012. "Imperatives: Meaning and Illocutionary Force." In *Empirical Issues in Syntax and Semantics 9: Papers from the Colloque de Syntaxe et Sémantique à Paris 2011*, edited by Christopher Piñón, 37–58. Paris: CNRS.

Corredor, Cristina. 2020. "Deliberative Speech Acts: An Interactional Approach." *Language & Communication* 71: 136–48. https://doi.org/10.1016/j.langcom.2020.01.005

Dascal, Marcelo. 2005. "Debating with Myself and Debating with Others." In *Controversies and Subjectivity*, edited by Pierluigi Barrotta and Marcelo Dascal, 33–73) Amsterdam: John Benjamins. https://doi.org/10.1075/cvs.1.04das

Davidson, Donald. 1963). "Actions, Reasons, and Causes." *The Journal of Philosophy* 60 (23): 685–700. https://doi.org/10.2307/2023177

Eemeren, Frans H. van, and Rob Grootendorst. 1984. *Speech Acts in Argumentative Discussions*. Dordrecht: De Gruyter Mouton. https://doi.org/10.1515/9783110846089

Fairclough, Isabela, and Norman Fairclough. 2012. *Political Discourse Analysis.* London: Routledge. https://doi.org/10.4324/9780203137888

Fogal, Daniel, Daniel W. Harris, and Matt Moss, eds. 2018. *New Work on Speech Acts*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.001.0001

Gauthier, David P. 1963. *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and Their Exemplification in Discourse.* Oxford: Clarendon Press.

Gilbert, Margaret. 1990. "Walking Together: A Paradigmatic Social Phenomenon." *Midwest Studies in Philosophy* 15: 1–14. https://doi.org/10.1111/j.1475-4975.1990.tb00202.x

Green, Michell S. 2009. "Speech Acts, the Handicap Principle and the Expression of Psychological States." *Mind & Language* 24 (2): 139–63. https://doi.org/10.1111/j.1468-0017.2008.01357.x

Green, Mitchell S. 2017. "Conversation and Common Ground." *Philosophical Studies* 174: 1587–1604. https://doi.org/10.1007/s11098-016-0779-z

Green, Mitchell S. 2018. "A Refinement and Defense of the Force/Content Distinction." In *New work on speech acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 99–122. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0004

Green, Mitchell S. 2020. "Speech Acts." *Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), edited by Edward N. Zalta. Last updated September 24, 2020. https://plato.stanford.edu/archives/win2020/entries/speech-acts/

Hample, Dale. 2005. *Arguing: Exchanging Reasons Face to Face.* Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781410613486

Hanks, Peter. 2018. "Types of Speech Acts." In *New work on speech acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 123–43. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0005

Harris, Daniel W., Daniel Fogal, and Matt Moss. 2018. "Speech Acts: The Contemporary Theoretical Landscape." In *New work on speech acts*, edited by Daniel Fogal, Daniel W. Harris, and Matt Moss, 1–39. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198738831.003.0001

Hitchcock, David. 2002. "Pollock on Practical Reasoning." *Informal Logic* 22 (3): 247–56. https://doi.org/10.22329/il.v22i3.2591

Houtlosser, Peter. 1998. "Points of View." *Argumentation* 12 (3): 387–405. https://doi.org/10.1023/A:1007770813424

Ihnen Jory, Constanza. 2016. "Negotiation and Deliberation: Grasping the Difference." *Argumentation* 30 (2): 145–65. https://doi.org/10.1007/s10503-014-9343-1

Jacobs, Scott. 1989. "Speech Acts and Arguments." *Argumentation* 3 (4): 345–65. https://doi.org/10.1007/BF00182603

Jacobs, Scott, and Sally Jackson. 1983. "Strategy and Structure in Conversational Influence Attempts." *Communication Monographs* 50 (4): 285–304. https://doi.org/10.1080/03637758309390171

Kauffeld, Fred J. 1998. "Presumptions and the Distribution of Argumentative Burdens in Acts of Proposing and Accusing." *Argumentation* 12 (2): 245–66. https://doi.org/10.1023/A:1007704116379

Lance, Mark, and Rebecca Kukla. 2013. "'Leave the Gun; Take the Cannoli': The Pragmatic Topography of Second-person Calls." *Ethics* 123 (3): 456–78. https://doi.org/10.1086/669565

Levinson, Stephen C. 1988. "Putting Linguistics on a Proper Footing: Explorations in Goffman's concepts of participation." In *Erving Goffman: Exploring the Interaction Order*, edited by Paul Drew, and Anthony Wootton, 161–227. Cambridge, Mass.: Polity Press.

Lewiński, Marcin. 2015. "Practical Reasoning and Multi-party Deliberation: The Best, the Good Enough and the Necessary." In *The Eighth Conference of the International Society for the Study of Argumentation (ISSA)*, edited by Bart

Garssen, David Godden, Gordon Mitchell and A. Francisca Snoeck Henkemans, 851–62. Amsterdam: SicSat.

Lewiński, Marcin. 2017. "Practical Argumentation as Reasoned Advocacy." *Informal Logic* 37 (2): 85–113. https://doi.org/10.22329/il.v37i2.4775

Lewiński, Marcin. 2019. "Argumentative Discussion: The Rationality of What?" *TOPOI: An International Review of Philosophy* 38 (4): 645–58. https://doi.org/10.1007/s11245-015-9361-0

Lewiński, Marcin. 2021. "Illocutionary Pluralism." *Synthese*, https://doi.org/10.1007/s11229-021-03087-7

Lewiński, Marcin, and Dima Mohammed. 2019. "The 2015 Paris Climate Conference: Arguing for the fragile consensus in global multilateral diplomacy." *Journal of Argumentation in Context* 8 (1): 65–90. https://doi.org/10.1075/jaic.18017.lew

McBurney, Peter, David Hitchcock, and Simon Parsons. 2007. "The Eightfold Way of Deliberation Dialogue." *International Journal of Intelligent Systems* 22 (1): 95–132. https://doi.org/10.1002/int.20191

Mercier, Hugo, and Dan Sperber. 2011. "Why Do Humans Reason? Arguments for an Argumentative Theory." *Behavioral and Brain Sciences* 34 (2): 57–74. https://doi.org/10.1017/S0140525X10000968

Pagin, Peter. 2016. "Assertion." In *Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. Last updated December 31, 2016. https://plato.stanford.edu/archives/win2016/entries/assertion/

Paglieri, Fabio. 2013. "Choosing to Argue: Towards a Theory of Argumentative Decisions." *Journal of Pragmatics* 59 (Part B): 153–63. https://doi.org/10.1016/j.pragma.2013.07.010

Pollock, John L. 1995. *Cognitive Carpentry: A Blueprint for how to Build a Person.* Cambridge, Mass.: MIT Press. https://doi.org/10.7551/mitpress/1887.001.0001

Poppel, Lotte van. 2019. "Anticipating Resistance to Health Advice: A Speech Act Perspective." *Discourse Processes* 56 (3): 256–69. https://doi.org/10.1080/0163853X.2018.1470440

Portner, Paul. 2007. "Imperatives and Modals". *Natural Language Semantics* 15: 351–83. https://doi.org/10.1007/s11050-007-9022-y

Portner, Paul. 2018. *Mood.* Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199547524.001.0001

Reimer, Marga. 1995. "Performative Utterances: A Reply to Bach and Harnish." *Linguistics and Philosophy* 18 (6): 655–75. https://doi.org/10.1007/BF00983301

Rodrigues, Soledade, Marcin Lewiński, and Mehmet A. Üzelgün. 2019. "Environmental Manifestoes: Argumentative Strategies in the *Ecomodernist Manifesto*."

Journal of Argumentation in Context 8 (1): 12–39. https://doi.org/10.1075/jaic.18036.rod

Sbisà, Marina. 2007. "How to Read Austin." *Pragmatics* 17 (3): 461–73. https://doi.org/10.1075/prag.17.3.06sbi

Searle, John R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Searle, John R. 1975. "A Taxonomy of Illocutionary Acts." In *Language, mind, and knowledge. Minnesota studies in the philosophy of science, Volume 7*, edited by Keith Günderson, 344–69. Minneapolis: University of Minnesota Press.

Searle, John R. 1989. "How Performatives Work." *Linguistics and Philosophy* 12: 535–58. https://doi.org/10.1007/BF00627773

Searle, John R. 2001. *Rationality in Action*. Cambridge, Mass.: MIT Press. https://doi.org/10.7551/mitpress/5759.001.0001

Searle, John R. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780195396171.001.0001

Searle, John R., and Daniel Vanderveken. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Stalnaker, Robert. 2002. "Common Ground." *Linguistics and Philosophy* 25 (5–6): 701–21. https://doi.org/10.1023/A:1020867916902

Strawson, Peter F. 1964. "Intention and Convention in Speech Acts." *The Philosophical Review* 73 (4): 439–60. https://doi.org/10.2307/2183301

Tuomela, Raimo. 2013. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199978267.001.0001

Urmson, J. O. 1977. "Performative Utterances." *Midwest Studies in Philosophy* 2: 120–27. https://doi.org/10.1111/j.1475-4975.1977.tb00033.x

Walton, Douglas. 2006. "How to Make and Defend a Proposal in a Deliberation Dialogue." *Artificial Intelligence and Law* 14 (3): 177–239. https://doi.org/10.1007/s10506-006-9025-x

Walton, Douglas. 2007. "Evaluating Practical Reasoning." *Synthese* 157 (2): 197–240. https://doi.org/10.1007/s11229-007-9157-x

Watson, Gary. 2004. "Asserting and Promising." *Philosophical Studies* 117: 57–77. https://doi.org/10.1023/B:PHIL.0000014525.93335.9e

Williamson, Timothy. 1996. "Knowing and Asserting." *Philosophical Review* 105: 489–523. https://doi.org/10.2307/2998423

Witek, Maciej. 2021. "Self-Expression in Speech Acts." *Organon F* 28 (2): 326-59. https://doi.org/10.31577/orgf.2021.28204

Organon F 28 (2) 2021: 458–483
https://doi.org/10.31577/orgf.2021.28208
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Do Kuhn's Cases of the Theory-Change from Newtonian to Einsteinian Physics Support His Incommensurability Thesis?

Kyoung-Eun Yang*

Received: 28 April 2020 / Revised: 22 September 2020 / Accepted: 8 October 2020

Abstract: In order to support his revolutionary view on scientific change, Kuhn suggests that there exist two separate aspects of the theory-change from Newtonian to Einsteinian physics that support his incommensurability thesis. For the evidence of his thesis Kuhn offers the conceptual change in the meaning of notion of "mass" in the theory-change. And he claims the absence of any neutral observational basis to evaluate the strengths of the two theories. This essay argues that these two cases fail to support his incommensurability thesis.

Keywords: Conceptual change of mass; Kuhn's incommensurability thesis; neutral observational basis; scientific revolution; space-time measurement; the theory-change from Newtonian to Einsteinian physics.

* Korea National University of Education
https://orcid.org/0000-0001-8967-6127
Graduate School, Korea National University of Education, 250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do, South Korea
newtleib@gmail.com

1. Introduction

The theory-change from Newtonian to Einsteinian physics has been regarded as an archetypal example of a revolution in science by both scientists and philosophers. For example, the physicist Max Born identified the development of the theories of relativity as "the Einsteinian revolution" which opened "the beginning of a new era" (Born 1965, 2). And the philosopher of science Karl Popper wrote that "Einstein revolutionized physics" (Withrow 1967, 25). After the prediction made by general relativity of the deflection of light due to the sun's gravitational field was confirmed by observations, the London *Times* on November 7th 1919 famously ran the headline "Revolution in Science" with a sub-heading "Newtonian Ideas Overthrown." In view of the fact that Einsteinian physics is founded on his novel understanding of the concepts of space-time and of the equivalence of mass and energy, which are quite alien to the Newtonian frameworks, it seems difficult to deny the intuition behind the idea that this episode of theory-change was indeed revolutionary.

Along these lines, Kuhn (1962) explicitly articulated and defended a revolutionary view involving "paradigm changes." According to Kuhn, two different paradigms are incommensurable in their assertions about the world, aims, criteria of appraisal, conceptual frameworks, and even observational basis. As supporting cases, Kuhn considers the conceptual discontinuities concerning notions such as 'mass' and 'space-time' as evidence of the occurrence of a revolution brought about by Einsteinian physics.

Einstein himself, however, rarely employed the term "revolution" in order to characterize his theories of relativity (Cohen 1985). He instead warned that the term "revolution" mischaracterizes the way that the special and the general theories were developed. Their development is considered as one which "slowly leads to a deeper conception of the laws of nature" based on results of "the best brains of successive generations" (Klein 1975, 113). Although Einstein referred to the theory-change from Newton to Maxwell as a revolution in that "action at a distance is replaced by the field" (Einstein 1949, 35), he did not maintain that the theories of relativity were new fundamental theories. The special theory of relativity is claimed as "simply a systematic development of the electromagnetics of Maxwell and

Lorentz" (Einstein 1934, 57). As for the general theory it was "the last step in the development of the program of field theory, … [and] it modified Newton's theory only slightly" (ibid., 37).

This paper follows Einstein's intuition—that the development of his two theories of relativity is not in fact revolutionary—in critically examining the aforementioned Kuhn's two cases in the theory-change from Newtonian to Einsteinian physics. This essay will argue that these two cases fail to support the incommensurability thesis.[1]

2. Kuhn's scientific revolution and conceptual changes in Einsteinian physics

Thomas Kuhn famously argued in his *Structure of Scientific Revolutions* that there have been major discontinuous changes in the history of science. He explicitly cites the theory-change from Newtonian to Einsteinian physics as a case strongly supporting his claim (Kuhn 1962).

The concept of paradigm in Kuhn's view plays a key role in characterizing the different stages of science. In the stable stage during which a specific scientific discipline matures, the discipline, which Kuhn calls a 'normal science,' "is predicated on the assumption that the scientific community knows what the world is like" (Kuhn 1962, 5). Scientists make great efforts to preserve this assumption, to the extent that "normal science often suppresses fundamental novelties because they are necessarily subversive of its basic commitments" (ibid.). Research within normal science is thus based on a 'paradigm,' which consists of the scientific community's metaphysics, conceptual frameworks, theories, methodology, and goals. A paradigm is essential to normal science, in that "no natural history can be interpreted in the absence of at least some implicit body of intertwined theoretical and methodological belief" (ibid., 16–7). Adopting a paradigm is "an attempt to force nature into the preformed and relatively inflexible [conceptual] box that the paradigm supplies" (ibid., 24). So, "[n]ormal-scientific research is

[1] Surely, I cannot dismiss Khun's general argument just because his specific example is dubious. This paper targets only Kuhn's argument that is based on the the theory-change from Newtonian to Einsteinian physics.

directed to the articulation of those phenomena and theories that the paradigm already supplies" (ibid.).

However, with the advent of a crisis within the paradigm (if only in some vague sense)—caused by the continuing failure to solve anomalies, the paradigm confronts challenges from competitors that question the fundamental assumptions underlying the earlier normal science: "[N]ature has somehow violated the paradigm-induced expectations that govern normal science. It then continues with a more or less extended exploration of the area of anomaly" (ibid. 52–3). And the "crisis may end with the emergence of a new candidate for paradigm and with the ensuing battle over its acceptance" (ibid. 84).

One of the competing paradigms, because of its success in solving the anomalies, attracts advocates who set the direction of their future research according to the new paradigm. "The transition from a paradigm in crisis to a new one from which a new tradition of normal science can emerge is far from a cumulative process … Rather it is a reconstruction of field from new fundamentals, a reconstruction that changes some of the field's most elementary theoretical generalizations" (ibid. 84–5). A scientific revolution occurs when the new paradigm replaces the old one. One result is that different paradigms are "incommensurable" in their aims, conceptual frameworks, and even observational bases: "The normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before" (ibid. 103). Kuhn's incommensurability thesis involves several radical claims—for example that even the empirical data for a given theory cannot be translated in a way that is neutral between competing paradigms:

> Scientists then often speak of the "scales falling from the eyes" or of the "lightning flash" that "inundates" a previously obscure puzzle, enabling its components to be seen in a new way that for the first time permits its solution. … No ordinary sense of the term 'interpretation' fits these flashes of intuition through which a new paradigm is born. Though such intuitions depend upon the experience, both anomalous and congruent, gained with the old paradigm, they are not logically or piecemeal linked to particular items of that experience as an interpretation would be. Instead,

> they gather up large portions of that experience and transform them to the rather different bundle of experience that will therefore be linked piecemeal to the new paradigm but not to the old. (ibid., 122–23)

In order to support his revolutionary view, Kuhn suggests that there were two separate aspects of the theory-change from Newtonian to Einsteinian physics that support the incommensurability thesis. These are the conceptual change in the meaning of notion of "mass" (ibid., 102), and the absence of any neutral observational basis to evaluate the strengths of the two theories due to the "theory ladenness" of the space-time measurements in the two theories (ibid., 149–50).

In the following sections, I will argue that these two claims fail to support the incommensurability thesis.

3. The case of the concept of mass

According to Kuhn, although the terms employed in Newtonian physics such as mass are also employed in Einsteinian physics, the referents of these terms are not the same:

> [T]he physical referents of these Einsteinian concepts [mass] are by no means identical with those of the Newtonian concepts that bear the same name. … Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same. (Kuhn 1962, 102)

In the same spirit, Paul Feyerabend pointed out:

> [I]n classical, prerelativistic physics the concept of mass (and, for that matter, the concept of length and the concept of time duration) was absolute in the sense that the mass of a system was not influenced (except, perhaps, causally) by its motion in the coordinate system chosen. Within relativity, however, mass has become a relational concept whose specification is incomplete without indication of the coordinate system to which the spatio-temporal descriptions are all to be referred. … what is measured in the classical case is an *intrinsic property* of the system under

> consideration; what is measured in the case of relativity is a *relation* between the system and certain characteristics of [the coordinate system] D'. (Feyerabend 1962, 80)

Although the term 'mass' appears in both theoretical frameworks, the Newtonian mass of a body is the same irrespective of its state of motion, whereas the Einsteinian mass varies depending on the motion of a body relative to the frame within which its mass is measured:

> [T]he total mass of a system is not a scalar quantity in relativity theory, so that its value depends on the reference frame with respect to which it is measured. For example a particle whose mass is m, as measured in its own rest frame, appears to have a larger mass when measured in a second frame with respect to which it is moving. (Penrose 2004, 435)

In other words, the Newtonian mass is a scalar quantity m, which is invariant under any coordinate transformation, while the Einsteinian mass (expressed as $M = \gamma m$, where γ is the Lorentz factor, i.e., $1/\sqrt{(1 - v^2/c^2)}$) is a variable quantity which increases with the velocity v of the body.

Feyerabend claims that the change in the concept of mass shows that there are enormous difficulties in relating the two successive scientific theories:

> It is also impossible to define the exact classical concepts in relativistic terms or to relate them with the help of an empirical generalization. … It is therefore again necessary to abandon completely the classical concepual scheme once the theory of relativity has been introduced … Our argument against meaning invariance is simple and clear. It proceeds from the fact that usually some of the principles involved in the determination of the meanings of older theories or points of view are inconsistent with the new, and better, theories. (Feyerabend 1962, 80–2)

The concepts of classical and relativistic mass essentially belong in "different and incommensurable frameworks" (ibid., 81). In his recent book on concepts of mass, Jammer describes the relationship between the concepts of invariant and relativistic masses as "ultimately the disparity between two competing views of the development of physical science" (Jammer 2000,

61). Kuhn also sees this conceptual change as a classic illustration of the incommensurability thesis:

> the normal-scientific tradition that emerges from a scientific revolution is not only incompatible but often actually incommensurable with that which has gone before. (Kuhn 1962, 103)

This case, however, fails to provide legitimate evidence supporting the occurrence of a radical conceptual change. Einsteinian relativistic mass cannot in fact be considered as a physically significant concept that is the counterpart of the Newtonian mass. In fact, the former is not a legitimate physical quantity that respects the principles of special relativity. Kuhn explicitly takes the 'relativistic mass' $M = \gamma m$, which increases with the velocity of the body. Yet this concept of mass does not properly fit within the framework of special relativity.

The special theory of relativity is essentially based on two fundamental postulates: (1) all physical laws take the same form in all inertial frames, and (2) the speed of light is always the constant c in all such frames. From these two hypotheses, Einstein derived the coordinate transformation which implements the principle of special relativity.[2] As a result, these two hypotheses yield predictions about the kinematical effects of time dilation, length contraction, and the addition of velocities. On the basis of this kinematics, a dynamical framework can be developed by positing the concepts of mass and momentum. In this context, as the above expression suggests, one can employ the *relativistic* concept of mass M_{rel} while maintaining the *classical* concept of velocity $v_{cla}(dx/dt)$ in order to relativitize the concept of momentum, i.e. $p_{rel} = M_{rel}v_{cla}$. In other words, the Newtonian momentum ($p_{cla} = m_{cla}v_{cla}$) is modified into the relativistic expression by adopting the relativistic mass (rather than by changing laboratory time dt to proper time $d\tau$), i.e., $\gamma mdx/dt$. At this point, however, one is forced to adopt a primitive concept of improper 4-velocity, which is not Lorentz covariant. (Oas 2006, 4) "The improper velocity being a direct result of the imposition of RM [the relativistic mass] means that RM is at odds with the accepted kinematics of special relativity" (ibid.).

[2] Accordingly, one of essences of the special theory of relativity is that the laws of physics must be covariant with respect to Lorentz transformations.

In a similar spirit, Wheeler and Taylor claim that:

> Any difference between [relativistic] formulae for momentum (for example, $mdx/d\tau$) and the corresponding Newtonian formula (mdx/dt) is therefore to be attributed to the difference between proper time and laboratory time, not to any difference in the value of m in the two descriptions of nature. (Wheeler and Taylor 1963, 108)

Given that the Lorentz factor $1/\sqrt{(1 - v^2/c^2)}$ measures the ratio between laboratory time and proper time, modifying the kinematical concept of velocity dt to $d\tau$ is more natural than modifying the concept of mass. So, in order to be consistent with the kinematics of special relativity, the Lorentz factor $1/\sqrt{(1 - v^2/c^2)}$ needs be associated with velocity, rather than with mass. This is expressed by Resnik as follows:

> Indeed, it should be noted that, whether we identify the factor $1/\sqrt{(1 - v^2/c^2)}$ with mass or with velocity, the origin of this factor in collision measurements is kinematical; that is, it is caused by the relativity of time measurement. (Resnik 1968, 199)

Because of this, Einstein himself considered the rest mass m as the only physically significant concept and substituted the energy-momentum 4-vector for the relativistic mass. Thus, he wrote:

> It is not good to introduce the concept of the mass $M = m/\sqrt{(1 - v^2/c^2)}$ of a body for which no clear definition can be given. It is better to introduce no other mass than 'the rest mass' m. Instead of introducing M, it is better to mention the expression for the momentum and energy of a body in motion. (Einstein 1948, a letter to Lincoln Barnett, quoted from Okun 1989, p. 42)

In the special theory, just as space and time are incorporated within the single entity space-time whose components are (t, x), so energy and momentum are united to form the energy-momentum 4-vector whose components are (E, – p), where we use $c = 1$ convention. This quantity satisfies a conservation law, which, given the equivalence of energy and mass, incorporates the law of mass conservation, the law of energy conservation, and the law of momentum conservation. The squared magnitude of this four-

vector represents the rest mass, $m^2 = E^2 - p^2$, which is invariant regardless of the choice of inertial frames. In this context, 'relativistic mass' is only the temporal component of the energy-momentum 4-vector of a given body. This single component is measured as being larger in motion than when the body concerned is at rest. However, all four components of the energy-momentum 4-vector are transformed in the proper way to maintain the vector's invariance under the change of the reference frame. Along these lines, Wheeler and Taylor's answer to the question "is the mass of a moving object greater than the mass of the same object at rest?" is "no":

> The concept of 'relativistic mass' is subject to misunderstanding. That's why we don't use it. First, it applies the name mass—belonging to the magnitude of a four-vector—to a very different concept, the time component of a four-vector. Second, it makes increase of energy of an object with velocity or momentum appear to be connected with some change in internal structure of the object. In reality, the increase of energy with velocity originates not in the object but in the [kinematical] properties of space-time itself. (Taylor and Wheeler 1992, 250–51)

It seems that Kuhn's reading of the concept of mass is mistaken.

Defenders of Kuhn's thesis that radical change is involved in this shift could respond that although Kuhn may have misunderstood the way in which the concept of mass has been radically transformed, adopting the concept of energy momentum still shows that the concept of "mass" undergoes a radical change. For example, the difference between the classical notion of mass and the energy momentum 4-vector is very apparent. Contrasted with classical mass, the components of the energy-momentum 4-vector are described differently with the change of the coordinate systems representing inertially moving observers. Moreover, unlike classical mass, its relativistic counterpart is implicated in the interactions among mass, energy, and momentum. This shows the ontological difference between the notions of classical mass and energy-momentum. Accordingly, it could be argued that my argument resurrects Kuhn's view by introducing the concept of the energy-momentum 4-vector.

To answer this criticism, I will analyze the reference and denotation of the theoretical terms, i.e., classical mass and energy-momentum 4-vector.

Bird (2000) and Field (1973) characterize the comparison of theoretical terms in theory-change as a matter of denotating and extending them. Consider Kuhn's claim "the *physical referents* of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name" (Kuhn 1962, 101). In response to this Kuhn's claim, Bird maintains that

> Kuhn assumes that laws fix intensional meaning and thereby also fix, in a strict manner, extension and reference. If it is further assumed that all laws play a part-thick intensionalism-then it follows that if the term in question refers, then the laws must be true; correspondingly if one or more of the laws fails to be true then there is no reference, in which case incommensurability of reference cannot arise. (Bird 2000, 176)

In response to this claim by Kuhn, Bird maintains that Newtonian mass does not refer to anything since Newtonian law is not true. Along a similar line, Field points out that "what Einstein showed is that there *is* no such quantity as "Newtonian mass"; and unless one holds that the world used to obey Newton's laws but started obeying Einstein's laws one day, it is clear that there was no "Newtonian mass" in Newton's time either" (Field 1973, 470).

However, physicists would disagree about this skeptical attitude toward Newtonian mechanics. Field also endorses the physicists' intuition since, to a certain extent, the reference of Newtonian mass is based on "a great many false-but-approximately-true beliefs" (ibid., 465). Accordingly, Field modifies Frege's descriptivist theory and embraces the idea that there can be the partial denotation of scientific terms such as Newtonian mass. Along the similar line, Bird claims that

> it need not be that the total content of a given law contributes to meaning, and indeed it is implausible that it should be so. But then the content of laws breaks up into an element that determines reference and an element that does not. (Bird 2000, 176)

In order to clarify the reference-determining elements of Newtonian mass, Field directs our attention to the theoretical terms of Einstein's special relativity, which converge on Newtonian mass. Given that "at low

velocities, the mass of a particle is *almost precisely equal to* twice its kinetic energy divided by the square of its velocity," this concept refers to either relativistic or proper mass (Field 1973, 464–5).

What is notable about Field's account is that the concept of the "false-but-approximately-true" Newtonian mass is *referentially indeterminate* (ibid., 467). The central idea of Newtonian mass can be captured by (1) momentum = mass · velocity, and (2) the quantity of mass is the same regardless of the choice of reference frame. However, (1) and (2) cannot both be true given that special relativity does not authorize the conjunction of (1) and (2). Therefore, it seems possible to specify the one that is not correct. However, Field claims that there are contexts in which physicists *legitimately* employ both the relativistic mass and the proper mass by endorsing only one of the two theoretical tenets. For example, they might suggest that momentum equals the relativistic mass times velocity, but that it by no means equals the proper mass times velocity. By contrast, the proper mass is the same regardless of the choice of reference frames, whereas the relativistic mass is not. According to Field,

> The conjunction of Newton's tenets (1) and (2) was objectively false, but there is no fact of the matter as to which of the conjuncts was true and which false, and hence no fact of the matter as to whether the word 'mass' as it occurred in them denoted relativistic mass or proper mass. (ibid., 468)

Newtonian mass can be approximated through the Newtonian limit of either the relativistic mass or the proper mass. However, "there is no fact of matter as to how the falsity of the theory as a whole is to be distributed among the individual sentences of the theory" (ibid., 474). Hence, there is no fact of the matter in theoretical terms as to what the approximate truth ascribes. Accordingly,

> there is simply no coherent way of using the term 'refers' in connection with Newton's word 'mass'. In spite of this, there are many of Newton's utterances containing the word 'mass' that we want to regard as true … It follows, then, that the truth and falsity of these utterances simply cannot be explained on the basis of what Newton was referring to when he used the word 'mass',

> for there is no coherent way of explaining what he was referring to. (ibid., 473)

Kuhn claims the incommensurability thesis since the term "Newtonian mass" refers to an invariant quantity, whereas the term "Einsteinian mass" does not. In opposition to Kuhn's claim that the referents of these two concepts are different, Field argues that these two concepts can both be the referents for the term "Newtonian mass." According to Field, the physical referent of both Newtonian mass and Einsteinian mass is either the relativistic mass or the proper mass. He maintains that the referentially indeterminate Newtonian mass and Einsteinian mass can refer to both, and he makes this claim while criticizing Frege's referential semantics.

I will modify Field's argument in line with my employment of the energy-momentum 4-vector. Field argues for referential indeterminacy since Newtonian mass can refer to both the relativistic and the proper mass. This is because physicists can legitimately employ both theoretical terms. However, as I have argued before, physicists prefer the *proper* energy momentum 4-vector to the *improper* relativistic mass. Nevertheless, I still endorse Field's notion of referential indeterminacy, even if I come at it from a different point of view. This is because the energy-momentum 4-vector *unifies* both the relativistic mass and the proper mass.

Einstein (1905) constructed the special theory of relativity in his attempt to incorporate the electric field and the magnetic fields, and correspondingly space and time: "The existence of the electric field [and the magnetic field] was therefore a relative one, dependent on the coordinate system used, and only the electric and magnetic field taken *together* could be ascribed some kind of objective reality" (Einstein 1919). As a result, mass, energy, and momentum are all incorporated into the energy-momentum 4-vector. Thus, the energy-momentum 4-vector contains both referential elements for the relativistic and proper mass. The temporal aspect of the energy-momentum 4-vector represents the relativistic mass, while its all four components together represent the proper mass. The former is true for one observer, while the latter is true for another observer. This would be a plural denotation on top of an indeterminate one.

Morrison considers this case of unity as a synthetic one that "involves the integration of two separate processes or phenomena" under one concept.

In this case, there is no ontological reduction that "offers little in the way of support for claims about a physical unity in nature" (Morrison 2000, 5). For Field, the physical referent of both Newtonian mass and Einsteinian mass is either the relativistic mass or the proper mass. By contrast, I argue that the physical referent of both kinds of masses is the energy-momentum 4-vector, which incorporates both the relativistic mass and the proper mass. By taking this slightly different route, I arrive at the same conclusion as Field's one. Along the similar line, Bird maintains that

> It could be that the reference-determining element is sufficiently general as to be applicable to E-mass as well as N-mass, in which case those will be the same property. Thus there will be no incommensurability of reference in this case… Consequently his argument fails to establish incommensurability due to shifts in reference. (Bird 2000, 176)

Accordingly, Kuhn's incommensurability thesis could be undermined if one considered the energy-momentum 4-vector as this reference-determining element.

The reference-determining elements of scientific theories that undergo theory-change are best captured by the so-called convergent realists. Hardin and Rosenberg (1982) claim that although we can admit that central theoretical entities in theories from the past do not refer to their intended objects, we can still talk about approximate truth due to the commonality between the past and present-day theories about the same causal or explanatory role. Kitcher's distinction between "working posits" and "presuppositional posits" is based essentially on whether theoretical terms are referring or non-referring (Kitcher 1995).

At this point, it seems appropriate to add Duhem and Poincaré's idea into our context, given that they were concerned with the theory-change related to mechanics (Duhem 1914, Poincaré 1902). Their reference-determining elements are invariant (or quasi-invariant) mathematical equations or structures. In Newtonian mechanics, the equation of motion, $F_i = ma_i = dp_i/dt$, $a_i = dv_i/dt$, $v_i = dx_i/dt$ ($i = 1, 2, 3$), relates the concept of a body's mass with the force that is exerted on the body. In relativistic mechanics, the corresponding equations for motion, $f^\mu = ma^\mu = dP^\mu/d\tau$, $a^\mu = dU^\mu/d\tau$, $U^\mu = dx^\mu/d\tau$ (m is the rest mass, P^μ is the energy-momentum 4-vector, $\mu =$

0, 1, 2, 3), relates the concept of mass-energy with four-force. The idea of classical mass and the relativistic concept of mass-energy are all incorporated into Newtonian and relativistic mechanics. Just as classical mass is the ratio of force to acceleration in Newton's law of motion, so, too, is its relativistic counterpart the ratio of four-force to four-acceleration in the relativistic law of motion.

Just as classical mass is preserved before and after the collision of bodies, each component of mass-energy and the three-momentum vector is always preserved before and after the collision of bodies. That is, mass-energy is a quantity that is conserved and which inherits its classical counterpart. While Newtonian mechanics can explain the collision process, the emission of a photon from a body can only be explained by relativistic mechanics. In fact, the relativistic equation of motion, $f^{\mu} = ma^{\mu}$, is derived from these premises: (1) the satisfaction of the relativity principle, (2) the correspondence with the relativistic law of inertia ($dU^{\mu}/d\tau = 0$, as $f^{\mu} = 0$), and (3) the Newtonian limit ($f^{\mu} = ma^{\mu}$ approximates to $F_i = ma_i$, as $U^{\mu} \ll c$) (Hartle 2003). This shows that the mass at rest and the three-momentum, respectively, inherit the concept of mass and momentum from classical mechanics. Along these lines, structural realists, such as Duhem and Poincaré, claim that, although the contents of a given theory are discarded, its underlying mathematical structures are invariant (or structurally invariant) in the course of the theory change:

> There was a continuity or accumulation in the shift, but the continuity is one of form or structure, not of content. (Worrall 1987, 117)

This mathematical continuity becomes manifest from the perspective of the group structure of space-time viewpoint. The geometric properties of space-time consist of a topological Lie group, which involves the transformations of rotations, boosts and translation on $\mathbf{R}^4$. (see Appendix A) That is, the Lorentz transformation preserving the space-time metric can be represented as a combination of the three transformations. The concept of energy and momentum, within this framework, arise as the generators of transformations on the group of translations. Whilst this group in the relativistic case is the Lorentz group, the corresponding group in the non-relativistic case is the Galilean group. In both cases, the concept of energy,

which is used as total relativistic mass, appears as the generator of time translations. In this context, Saunders views the relationships between the concepts of the rest mass in both physics as involving the group structure:

> The non-relativistic mass, in contrast, has a quite different interpretation … , bound up with more detailed properties of the respective Lie algebras: in the case of the [inhomogeneous Galilean group], to the 'neutral elements' of the algebra (it therefore defines the momentum and the energy in conjunction with the velocity); in the relativistic case to the Casimir invariants (a function of elements of the Lie algebra, not a separate element). In both cases these quantities have vanishing Lie bracket with every element of the Lie algebra; they are therefore conserved. One has a quite reasonable understanding of their inter-relationships as provided by the theory of group contractions. (Saunders 1993, 304)

So, by introducing the concepts of the energy-momentum 4-vector and the rest mass, the relationships between the concepts of mass can be captured. This concept unifies mass, energy and momentum without attempting to unify its physical nature. This unification is incapsulated by the common mathematical equations and group structures. One can say that the common mathematical structures, which survive scientific revolution, are in fact what theoretical terms refers to. Accordingly, Kuhn's attempt to use the concept of mass to support the occurrence of a radical conceptual change does not succeed.

4. The case of incommensurable space-time measurements

While the above case for the revolutionary conceptual change in the notion of mass is not supported by space-time geometry, Kuhn's other case for the incommensurability thesis as applied to the Newton-Einstein shift is concerned with an alleged conceptual change in the notion of space-time itself. In the rest of this section, I will argue that this case also in fact fails to show that the development of Einsteinian physics was revolutionary.

According to Kuhn, given that the concept of space and time provides the foundation of both Newtonian and Einsteinian physics, the change of

these concepts generates a revolution in the conceptual network. The transition from Newtonian to Einsteinian physics can be characterized as a holistic one in the sense that it involves changes in a range of other interrelated concepts:

> To make the transition to Einstein's universe, the whole conceptual web whose strands are space, time, matter, force, and so on, had to be shifted and laid down again on nature whole. (Kuhn 1962, 149)

Within Newtonian physics, simultaneities and temporal intervals between any two events are all absolute in that they are the same for all inertial observers. In other words, two events that are simultaneous for an inertial observer are simultaneous for any other inertial observer regardless of her relative motion, and the same holds for temporal intervals. Spatial intervals between two simultaneous events are also absolute. Within Einsteinian physics, on the other hand, simultaneity, and both temporal and spatial intervals between any two events are all relative to the motions of inertial observers.

Kuhn maintains that these differences between the two theoretical frameworks show that the conceptual change involved was revolutionary. Advocates of the competing paradigms, according to Kuhn, experienced a "transition of vision," which meant that the two sets of scientists observed totally different worlds:

> One [set of scientists] is embedded in a flat, the other in a curved, matrix of space. Practicing in different worlds, the two groups of scientists see different things when they look from the same point in the same direction. (ibid., 150)[3]

First of all, we need to ask what the occurence of this "transition of vision" exactly involves in this context. In the aforementioned first quotation, Kuhn

[3] This clearly separates Kuhn's revolutionary view from any evolutionary view. The latter must maintains that the realtivistic modifications of kinematic and dynamic concepts conserve or quasi-conserve essential observational and theoretical components of Newtonian physics. An evolutionary view, for example, claims the continuity of relativistic kinematic and dynamic concepts in the classical limit.

suggests the occurrence of a change in the relationship between the fundamental conceptual elements, i.e., a structural change of the whole conceptual network. And the second quotation refers to a change of empirical substructures as a result of the conceptual change. So, given that measurements are essentially interwoven within different space-time concepts, they cannot provide a neutral basis from which to evaluate the relative empirical strengths of the two theories. Accordingly, the incommensurablity thesis involves the claim that no neutral observational basis exists due to theory ladenness.

Kuhn's conclusion is based on two premises, (1) the revolutionary conceptual change in notions of space and time, and (2) theory-ladenness of the measurements highlightened by those conceptual changes. From these two premises, Kuhn concludes that no neutral observational grounds exist for comparisons between the measurements in Newtonian and Einsteinian physics.

Two strategies might be employed to undermine Kuhn's argument. The first is to undermine one or both of the above premises. The second strategy is to argue that Kuhn's conclusion cannot be guaranteed even if we accept both premises. I will choose the second strategy. In this section, I will not attempt to undermine either the claim of the revolutionary development of the concept of space and time, or that of the theory-ladenness of the measurements based on the two different space-time theories. The strategy is instead to employ the very weapons of Kuhn's own argument. What will be argued is that *even if* we admit that radical conceptual changes occurred in the notions of space and time and we accept the theory-ladenness of the measurements at issue, there still exists a neutral observational basis from which to evaluate the relative evidential strengths of Newtonian and Einsteinian physics.

We need first to take a close look at the way Kuhn's premises can be understood within the context of this theory-change. As for theory-ladenness, the measurements of length are, as both sides agree, made with rigid rulers. It seems that all that the measurements with the rulers show is that if their two arms were aligned, their ends would coincide. How, then, can they be "theory-laden"?

Yet, it can in fact be argued that measurements with these instruments are dependent on a specific space-time theory, given that the length and

interval of a rigid ruler and a clock are interpreted differently with respect to the two possible embedding theories. This is because the embedding theories are concerned with the spatio-temporal intervals between events themselves. Given that the length of two different ends of the rigid rods represents the relationships between two different spatio-temporal events, it could be differently interpreted with respect to different space-time theories. The length of a rigid ruler and the interval measured by a clock, according to Newtonian physics, are interpreted as invariant irrespective of their state of motion. On the other hand, within Einsteinian physics, the same instruments are posited as possibly experiencing length contractions due to their state of motion.

If, on the contrary, the measurements of spatio-temporal intervals employ instruments such as a light pulse and a clock, then theory-ladenness of spatial measurements occurs in the opposite way. Geroch shows how the measurement of spatio-temporal intervals is possible with light pulses and clocks (Geroch 1981, 69–72). Light pulses can be employed to probe space-time because "light, once emitted, moves within the environment of space-time, independently of what the emitter was doing" (ibid., 72). And by employing mirrors, "one can arrange for the light to get back to us to tell us what space-time is like" (ibid.). Observers who carry a clock with them can evaluate the spatio-temporal intervals between the observer and a specific event by evaluating the time it takes for the light sent by them to be reflected back from the event. The clocks carried by moving observers are obviously neutral between Newtonian and Einsteinian theories. Within the latter framework, the temporal intervals of the clock are distorted only when an observer measures the clocks of the others which are in motion relative to the observer. In the case of measuring the observer's own clock, its measurement results will be identical regardless of whichever theory it is based on. Yet, the light pulses are laden with a specific space-time theory here. According to the special theory of relativity, the speed of the light pulses is constant regardless of the motion of an observer relative to the light source. Yet, in Newtonian physics, the speed of light pulse is posited to change depending on the observer's motion with respect to the light pulse. Given that the dimensions of the same instruments for the measurement of spatial intervals are differently interpreted in different theories, it

seems that the instruments by no means provide a neutral observational basis between the two theories.

Consider, for example, the Michelson-Morley experiment. As it was designed to measure the effect of the ether on the speed of light pulses, this experiment can in fact be interpreted as measuring the structure of space-time. To detect the effect of the ether, the apparatus is equipped with an interferometer, which is posited as being at rest with respect to the earth moving through the ether. The idea of the experiment is to compare the speed of light pulses moving through the ether frame with the speed of light pulses perpendicular to the frame. By using half-silvered mirrors, light pulses are reflected to travel back and forth along two different directions, once along the direction of motion of earth and once at right angle to that motion. Although this experiment was originally designed to measure the speed of light with respect to the frame of ether, we can also employ this experiment in order to examine whether or not Lorentz contraction occurs.

The occurrence of length differences for these two distinct round-trip journeys can be employed as an appraisal of Newtonian and relativity theories. In this experiment, the existence of a length contraction of the interferometer arms is a component that plays a crucial role in producing the length difference of the two round-trip journeys, which is the key to the appraisals of Newtonian and Einsteinian physics. When this experiment is prepared, an ordinary ruler is employed to determine the lengths of the interferometer arms. We also could think of the case that the experimental design employs light pulses and clock in order to determine the lengths of the interferometer arms. In both cases, when one assumes that the ruler (or light pulses and clock) can be employed to determine the lengths of the arms, one assumes the very theory to be appraised (Laymon 1988, 250). The former case employing the ruler assumes that there does not exist Lorentz contraction effect, while the latter using light pulses assumes the opposite. Then, do the difference of the concept of space-time and the theory-ladenness of space-time measurement guarantee the incommensurability Kuhn suggested?

Our answer is 'no.' Although one admits that a component of the experiment is theory-laden as mentioned, this experiment still produces a neutral observational basis for the appraisal of Newtonian and Einsteinian

physics (Laymon 1988, Cho 1996). Consider the lengths of round-trip journeys that are experienced by (1) the light pulses moving through the ether frame and (2) the light pulses moving perpendicular to the frame. In the original experimental setting, where c and v are the velocity of light and the earth in its orbit respectively and D is the distance of interferometer arm length, (1) is calculated as $2D/(1 + v^2/c^2)$, while (2) is $2D(1 + v^2/c^2)^{-1/2}$, which is $2D(1 + (v^2/2c^2))$ if terms higher than $(v/c)^2$ are neglected. (See Appendix B) As Michelson and Morley rotated the whole apparatus through 90°, the predicted displacement of the interference fringe becomes $2D(v^2/c^2)$ (Michelson and Morley 1887, 336).

In order to consider the possibility of the Lorentz contraction of the interferometer arms, we can consider a modified analysis of the result of the experiment, which assumes that the interferometer arm lengths can vary (Silverstein 1914). In other words, the interferometer arm that is parallel to the motion through the ether frame experiences Lorentz contraction. According to Laymon, whichever theory we employ, the length of the two paths of the light travels in the two different directions, is calculated as the same, if terms higher than $(v/c)^2$ are neglected (Laymon 1988). The existence or non-existence of contraction of the moving ruler by no means influences the anticipated effect of the difference of the lengths of the two paths. This result stems from the fact that the final outcome is a function of the sum of the lengths of the two interferometer arms, i.e., the lengths of arms initially parallel and the lengths of arms orthogonal to the direction of motion. Let the length of the former be D_h and the length of the latter be D_v. Given a 90o rotation of the interferometer, the anticipated path length change is $[D_h + D_v](v^2/c^2)$, which is a function of the sum of the interferometer arm lengths. When we assume the hypothesis of Lorentz contraction to decide the length of the rulers, the corrected D_h, when an ordinary ruler yields a value of 11 meters for both D_h and D_v, D_h is calculated as $11(1 - v^2/c^2)^{1/2}$. Ignoring higher order terms, D_h is obtained as $11(1 - (v^2/2c^2))$. Inserting this corrected value as an input value, the fringe shift becomes $2[11 + 11(1 - (v^2/2c^2))](v^2/c^2)$. This then yields $2[11 + 11](v^2/c^2)$ by expanding and ignoring terms of higher than second order. (See Appendix C)

This is the same as the anticipated measurement in the case that no contraction of the ruler is assumed. So whether or not the length contraction

is posited, the derived result of the fringe shift is the same. From this calculation, Laymon concludes:

> All of this means that while it is true that the actual length of the interferometer arms is a varying function of the very theories to be tested, when measuring that length the *computational* effect of assuming different theories from among the set to be tested is inconsequential. Hence, the phenomena (of fringe shift in rotating equal-arm interferometers) can be determined with an accuracy sufficient for testing the relevant theories *regardless* of which of the competing theories is chosen to specify the measurement procedures to be used to determine the experimental initial condition of length. (Laymon 1988, 252–53)

We can see that, although a component of the measurement depends on a specific space-time theory, the experiments, whichever theories we employ, still provide a neutral observational basis for the appraisal of Newtonian and Einsteinian physics. So, theory-ladenness of the measurements highlightened by those conceptual changes within space-time theories does not guarantee Kuhn's conclusion of the non-existence of neutral observational grounds in the comparisons between the measurements in Newtonian and Einsteinian physics.

5. Conclusion

It has been argued that Kuhn's incommensurability thesis is by no means supported by his own cases employing the concept of mass and space-time measurement within Newtonian and Einsteinian physics. Kuhn claims that the conceptual change in the notion of mass supports his incommensurability thesis, and also that this thesis is supported by the non-existence of a neutral observational basis for space-time measurements, stemming from the conceptual change in the notion of space-time. Yet, as regards the first, it has been argued that the concept of "relativistic mass," which Kuhn claims to be incommensurable with its classical counterpart (classical mass), is not in fact a physically meaningful concept. Furthermore, by introducing the energy-momentum 4-vector, we can clearly see the inter-relationship

between classical mass and its relativistic counterpart. And as regards the second case, it has been argued that although such space-time measurements are in a sense theory-laden, a neutral observational basis between Newtonian and Einsteinian physics can nonetheless really be secured. Accordingly, these two cases fail to support Kuhn's incommensurability thesis.

Appendix A

The concept of continuous symmetry, such as rotational symmetry in 3D Euclidean space and Lorentz symmetry in Minkowski space-time, can be captured by Lie groups. Lie groups are manifolds that form groups. A manifold is a space which resembles Euclidean space in a small neighborhood of each point. For the manifold to be a Lie group, a group multiplication between every pairs of points of the manifold can be defined.

This employment of a Lie group inherits the intuition of Felix Klein's Erlangen program, which emphasizes the formal properties of symmetry groups in characterizing differential equations and geometrical objects.

According to this program, symmetry groups determine the essence of equations and geometries, since their genuine characteristics are the invariant quantities under a group of specific transformations. For example, in Newtonian space-time, the transformations between coordinate systems representing inertial frames form Lie groups, and the symmetries of these groups distinguish the invariant geometric objects, independently of the choice of any specific reference frame. Along these lines, in Minkowski space-time, the Lorentz transformation, which incorporates the transformations of rotations, boosts and translation on $\mathbf{R}^4$, forms a Lie group. That is, the Lorentz transformation preserving the space-time metric can be represented as a combination of the three transformations.

A set of group generators is a set of group elements such that repeated application of the generators on themselves and each other can produce all the group elements. At this point, one can see the connection between the existence of conserved quantities and symmetry properties of a given physical system. The concept of energy and momentum, within Newtonian framework, arise as the generators of the group of temporal and spatial translations.

Whilst this group in the relativistic case is the Lorentz group, the corresponding group in the non-relativistic case is the Galilean group. In both cases, the concept of energy, which is used as total relativistic mass, appears as the generator of time translations.

Appendix B

As Michelson and Morley (1887) define, c is the velocity of light, v is the velocity of the earth in its orbit respectively and D is the distance of interferometer arm length. The calculations of the lengths of round-trip journeys that are experienced by (1) and (2) can be given as follows.

Let T = time that light occupies to outgoing pass, and T_1 = time that light occupies to returning pass. Then, T is calculated as $D/(c + v)$, and T_1 is calculated as $D/(c + v)$. And the whole time of the round-trip is $T + T1 = D/(c - v) + D/(c + v) = 2Dc/(c^2 - v^2) = 2Dc/[c^2(1 - v^2/c^2)] \sim [2D(1 + v^2/c^2)]/c$, if terms higher than $(v/c)^2$ is neglected. So, the length of round-trip journeys, i.e., (1), is $c(T + T_1) = 2D(1 + v^2/c^2)$.

The calculation of the time of the round-trip of other path is similar as the derivation of the time dilation employing a light pulse. The travel time of the round-trip of a light pulse is $2D/(c^2 - v^2)^{1/2} = 2D/[c^2(1 - v^2/c^2)]^{1/2} = [2D(1 - v^2/c^2)^{-1/2}]/c$, which is $[2D(1 - (v^2/2c^2))]/c$ if terms higher than $(v/c)^2$ is neglected. Accordingly, its length, i.e., (2), is $2D(1 + (v^2/2c^2))$.

Appendix C

Lorentz constructed his ether theory which explains the undetectability of the relative motion of ether with respect to earth. According to this theory, the frame of ether can be distinguished as Newton's absolute space. And the length contraction of a body in line with its motion explains the null result of the Michelson-Morley experiment.

A central idea of Lorentz ether theory is summarized as the "theorem of corresponding states". The theorem says that the laws of electromagnetics are invariant irrespective of the choice of the reference frame, i.e., whether it is measured with respect to the stationary ether frame or the moving one.

The length contraction of a moving body in its moving direction enables physicists to identify the "preferred" reference frame based on Lorentz's

ether. The formula of the length contraction is given as $l = l_o(1 - v^2/c^2)^{1/2} \sim l_o(1 - (v^2/2c^2))$, where lo is the length with respect to the stationary ether.

In the Michelson-Morley experiment, the distance of interferometer arm length is set to D whether it is measured in its moving direction or its perpendicular one. Then, the predicted displacement of the interference fringe is calculated as $2D(v^2/c^2)$. (see Appendix 1) In contrast, under the assumption that the interferometer arm lengths can vary in accordance with Lorentz ether theory, the predicted displacement of the interference fringe is calculated as $2(D_h + D_v)(v^2/c^2)$.

Since Dh is what Lorentz contraction is applied to, D_h is calculated as $11(1 - v^2/c^2)^{1/2} \sim 11(1 - (v^2/2c^2))$. ($D_h$ with respect to the stationary ether is measured as 11 meters, as Michelson and Morley prepared.) Inserting this corrected value as an input value, the predicted displacement of the interference fringe becomes $2[11 + 11(1 - (v^2/2c^2))](v^2/c^2) = 2[11 + 11 - 11(v^2/2c^2)](v^2/c^2)$. This then yields $2[11 + 11](v^2/c^2)$ by ignoring terms of $(v^2/c^2)^2$.

References

Bird, Alexander. 2000. *Thomas Kuhn*. Princeton: Princeton University Press.

Born, Max. 1965. *My Life and My Views*. New York: Charles Scribner's Sons.

Cho, Inrae. 1996. "The Methodological Challenge of Kuhn's Incommensurability Thesis." *Philosophy, Korea Philosophical Association* 46: 155–88.

Cohen, I. Bernard. 1985. *Revolution in Science*. Cambridge MA: Harvard University Press.

Duhem, Pierre. 1914. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Einstein, Albert. 1905. "On the Electrodynamics of Moving Bodies." In *Principle of Relativity*, edited by Einstein, Albert., Hendrik A. Lorentz, Hermann Minkowski, and Hermann Weyl. 1952. Translated by W. Perrett and G.B. Jeffery, 35–65. New York: Dover Books.

Einstein, Albert. 1919. "What is the Theory of Relativity?" In *Ideas and Opinions*. 1982. 227–32. New York: Bonanza Books.

Einstein, Albert. 1934. *Essay in Science*. New York: Philosophical Library

Einstein, Albert. 1949. "Autobiographical notes." In *Albert Einstein, Philosopher-Scientist*, edited by Paul Arther Schilpp. Chicago: Open Court.

Einstein, Albert. 1965. *My Life and My Views*. New York: Charles Scribner's Sons.

Feyerabend, Paul. K. 1962. "Explanation, Reduction, and Empiricism." In *Minnesota Studies in the Philosophy of Science*, vol III, edited by H. Feigl and G. Maxwell, 28–97. Dordrecht: D. Reidel Publishing Company.

Field, Harty. 1973. "Theory Change and The Indeterminacy of Reference." *The Journal of Philosophy* 70(14): 462–81. https://doi.org/10.2307/2025110

Geroch, Robert. 1978. *Relativity from A to B*. Chicago: University of Chicago Press.

Hardin, Clyde L. and Alexander Rosenberg. 1982. "In Defence of Convergent Realism." *Philosophy of Science* 49(4): 604–15. https://doi.org/10.1086/289080

Hartle, James B. 2003. *Gravity: An Introduction to Einstein's General Relativity*. New York: Pearson

Jammer, Max. 2000. *Concepts of Mass in Contemporary Physics and Philosophy*. Princeton: Princeton University Press.

Klein, Martin J. 1975. "Einstein on Scientific Revolutions," *Vistas in Astronomy* 17: 113–33. https://doi.org/10.1016/0083-6656(75)90053-7

Kitcher, Phillip. 1993. *The Advancement of Science*, New York: Oxford University Press.

Kuhn, Thomas I. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, Imre and Alan Musgrave. Eds. 1970. *Criticism and the Growth of Knowledge*, London: Cambridge University Press.

Laymon, Ronald. 1988. "The Michelson-Morley Experiment and the Appraisal of Theories." In *Scrutinizing Science*, edited by Donovan, Arthur, Larry Laudan, and Rachel Lauda., 245–66. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-2855-8_12

Michelson, Albert A. and Morley. Edward W. 1887. "On the Relative Motion of the Earth and the Luminiferous Ether," *American Journal of Science* 34: 333–45. https://doi.org/10.2475/ajs.s3-34.203.333

Morrison, Margaret, 2000. *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge: Cambridge University Press.

Oas, Gary. 2006. "On the Abuse and Use of Relativistic Mass." http://arxiv.org/abs/physics/0504110

Okun, Lev. B. 1989. "The Concept of Mass," *Physics Today* 42(6): 31–36. https://doi.org/10.1063/1.881171

Penrose, Roger. 2004. *The Road to Reality: A Complete Guide to the Laws of the Universe*. Knopf.

Poincaré, Henri. 1902, *Science and hypothesis*, New York: Dover.

Popper, Karl. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Resnik, Robert. 1968. *Introduction to Special Relativity*. New York: Wiley.

Saunders, Simon. 1993. "To What Physics Corresponds." In *Correspondence, Invariance, and Heuristics, Essays in Honour of Heinz Post*, 295–25, Dordrecht, Reidel. https://doi.org/10.1007/978-94-017-1185-2_15

Silberstein, Ludwik. 1914. *The Theory of Relativity*. London: London, Macmillan.

Taylor, Edward. F. and John. Archibald Wheeler. 1966. *Spacetime Physics*. San Francisco: W. H. Freeman.

Taylor, Edward. F. and John. Archibald Wheeler. 1992. *Spacetime Physics 2nd edition*. San Francisco: W. H. Freeman.

Withrow, Gerald James. ed. 1967. *Einstein: The Man and His Achievement*. A series of broadcast talk. London: BBC.

Worrall, John. 1989. "Structural Realism: The Best of Both Worlds?" *Dialetica* 43: 99–124. https://doi.org/10.1111/j.1746-8361.1989.tb00933.x

Organon F 28 (2) 2021: 484–491
https://doi.org/10.31577/orgf.2021.28209
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Some Further Remarks on Hybrid View of Fictional Characters

Miloš Kosterec*

Received: 29 November 2019 / Accepted: 10 March 2020

Abstract: In this short paper, I focus on several properties of so-called Hybrid View of Fictional Characters. First, I present the theory to detail needed for the discussion I am to put forward. Then I present several remarks on the consequences of the theory, mainly the problem of identifying of fictional characters and the problem of modal properties of sentences containing fictional names.

Keywords: fictional character, fictional name, requisite, de dicto, de re

1. Introduction

Hybrid View of Fictional Characters (HV) was now presented and defended by Glavaničová on several occasions (see mainly Glavaničová (2017), (2018)).[1] The theory was questioned by Kosterec in (2018). Glavaničová

[1] The 'hybridity' of the view stems from its sharing of particular features of theories on both sides of the discussion between realists and anti-realists about the fictional characters. More on this later.

* Slovak Academy of Sciences
https://orcid.org/0000-0001-8398-2297
Institute of Philosophy SAS, Klemensova 19, 811 09 Bratislava, Slovakia
milos.kosterec@gmail.com

presented a response in (2020). This piece can be considered as another step within that discussion. First, let's present the theory in short.

The theory tries to pose itself within the field of discussion about the metaphysical (and semantical) properties of so called purely fictional characters. These are supposed entities that seem to stand for the proper-name-like terms of non-actual entities within the works of (literary) fiction.[2] The most prominent example in the discussions (still) seems to be *Sherlock Holmes*. HV is not (at least in its present state) yet applied to discussions about the properties of actual characters which take part in the works of fiction (e.g *Napoleon* in *War and Peace*).[3] The theory take stand between so called *realists*, who pose fictional characters as real (existent) entities and *anti-realists*, who are against such position. There are no fictional characters according to the position. HV presents its position form the outset of suppositions in which we use fictional names. HV is based on the assumption that the differences between proper names in normal non-fictional contexts and proper-name-like terms within (meta)fictional contexts are enough to enable us to consider their semantics and metaphysics as different (or at least not necessary identical).[4]

The fictional names (proper-name-like terms of characters within context of fiction), according to HV, are used in several different suppositions. HV prominently discusses two of them: supposition *de dicto* and *de re*. A

[2] We have to be careful here. Fictional names (or rather proper-name-like terms) *are not proper names* according to Glavaničová's theory (following Currie, Predelli and Zouhar, among others). Further, fictional characters (according to the Glavaničová's theory) are actual (albeit abstract) entities. Thanks to an anonymous reviewer for urging me to clarify on these points.

[3] Glavaničová did not discuss these cases much in her (2017) and (2018). The focus of those works is on purely fictional names only. The discussion about the applicability of the theory on these cases is yet to come. The theory seems to be at least prima facie applicable within that field. The application of the theory, however, depends on the position one takes (argues for) in the case of proper-names-like terms for actual individuals used in fiction. If these are not taken to be proper names, then HV is applicable. If these are to be taken as genuine proper names, HV does not seem to be applicable (according to the theory, genuine proper names are not fictional names). Thanks to an anonymous reviewer for pressing me on this point.

[4] The theory makes no claims about proper names and their semantics, however.

fictional name is used in supposition de dicto when we speak about the sense of fictional name. The theory states we speak about some hyperintension, when we use the fictional name with supposition de dicto. This enables us to consider the sentences like *Sherlock Holmes is a fictional character* as true without the need to postulate any real individual. We use fictional names with supposition de re, when we (try to) speak about the individual it stands for, as it seems to be the case in *Sherlock Holmes is a detective* when considered within the context of fiction. According to HV, there is necessarily no such individual and the sentence is either denotationless or false, yet it is meaningful. In short, according to HV the fictional names have a hyperintension as its content. Presumably different fictional names stand for different hyperintensions. Each such hyperintension denotes necessarily empty intension and there is no extension for such names. The theory presupposes a semantic theory detailed enough to differ among hyperintensions, intensions and extensions of fictional names. The theory has realist facets in that it supposes there is something connected to fictional names, at least when we use fictional names de dicto, i.e. some sense, hyperintension. On the other hand, the theory is anti-realist when speaking about the use of fictional names de re. From this the 'hybridity', at least when posed within the discussion between realists and antirealists. From semantical point of view, there is no 'hybridity' – fictional name has hyperintension, intension and lacks extension.[5] The supposed ambiguity of fictional names when used in different context (fictional, meta-fictional, …) stems from different supposition in which we use the same fictional name. It seems that we use fictional names in supposition de dicto in meta-fictional contexts and with supposition de re in fictional contexts.

2. The problem of identification

Kosterec (2018) criticized theory in state presented by Glavaničová (2017) mainly due to its supposed presupposition of use of Transparent Intensional Logic (TIL). Glavaničová (2018) and (2020) made it clear that

[5] Therefore, according to the theory, fictional names always lack referent.

the theory does not presuppose TIL.[6] The main problem seemed to be the individuation of fictional characters. Why is *Sherlock* different from *Watson*? According to the actual state of HV, it is because there are different hyperintensions for 'Sherlock' and 'Watson'. Fair enough. But wait a moment, why are these hyperintensions different? For example, in TIL, the difference between hyperintensions stems from the inductive definition based on non-hyperintensional level. Also, HV does not *just presuppose* that there are different hyperintensions for 'Sherlock' and 'Watson'. HV contains reasons, why these hyperintensions seem to differ. This in turn will provide some answer to the question of identifying and discerning among fictional characters. So, what is the root of the difference according to HV? In short, sets of requisites. The hyperintensions differ because there are different sets of requisites connected to them. For example, *Sherlock* is a detective, *Holmes* is a doctor. And because their sets of requisites are supposed to differ, they are different hyperintensions and consequently different fictional characters. HV, however, does not provide any mechanism or defined criteria for the sets of requisites, although there seems to be some work in progress (see Jespersen – Duží - Glavaničová (2020)). HV in its actual state just presupposes that fictional characters differ because they have different sets of requisites.

This is my main problem with the theory in its actual state. As soon as we have fictions with more than just one fictional character (e.g. *Holmes* stories), the theory is in danger of circularity when it comes to explaining the difference among sets of requisites of fictional characters. According to Glavaničová, there is some sense, in which it is essential for *Sherlock* to be a detective. It is essential *for the story.*[7] It seems to be essential for the story, that *Sherlock* is a detective. It also seems to be essential for the story, that *Watson* is a doctor. What else, however, seems to be essential for the story, is that *Sherlock* is a friend of *Watson* and that *Watson* is a friend of

[6] Although she assumes HV can be stated *also* within TIL. I have yet to see such proper formulation, which either respects seeming simplicity of content of ficitonal names (And I have big doubts that it is feasible within TIL) or presents some complex hyperintensions as meanings of fictional names. From what I have seen during several conferences, the second option seems to be the assumed possibility.

[7] See Glavaničová (2018, footnote 18). More on that later.

Holmes. Now, if we are to consider these relational properties to be requisites (and nothing blocks us in the actual state of HV), then we have a problem. In order to differentiate between fictional characters, we are to be able to discern between their correspondent sets of requisites. Ok, in order to do that, we have to be able to discern between those sets. The only identity criteria of sets (after all influences of interpretation, contexts and whatever are resolved) is given by its elements (if any). Now the elements of sets of requisites are requisites. Therefore, in order to distinguish between *Sherlock* and *Watson*, we have to be able to distinguish between the requisites within their requisite sets. But, then we shall be able to distinguish between the requisites *is a friend of Watson* and *is a friend of Sherlock*. But in order to do so, we have to be able to differ, once again, between *Sherlock* and *Watson*. That is a circle.[8] A reply could be posed, that these relational properties are not essential for this story. That would probably miss the point. First, *Sherlock* and *Holmes* are *famous duo*. It would be strange to consider their friendliness (or other convenient relation) as *accidental for the story*. Second, the point can be stated more generally, the HV conditions for differentiating among fictional characters are circular for fictions for which it is essential that a fictional character is in some relation to some (other) fictional character. And, because these seem to be very much abound, the use of HV is problematic for analysis of great deal of fictions. The point could be even stronger, if there is some essential relation a fictional character must be in with itself. Then the theory would not be usable at all. Be that as it may, other worries by Kosterec, connected to the semantical model of fictional names in HV, still stand: e.g.:

> Glavaničová seems to suggest the possibility of new criteria of identity of constructions. It can be done, but at the cost of leaving TIL (at least in its present forms). One way or the other, her proposal will still have problems in both intensional and extensional contexts, since the contexts respect the substitution of equivalent constructions salva veritate. (Kosterec 2018, 123)

[8] The same point, in fact, was posed by Fine in (1982) for theories not dissimilar to HV.

3. The modal profile

The question, what is essential for the story (again after all the mumbo-jumbo about the influences of interpretations, context and whatever is resolved) leads me to my second remark. Kosterec (2018) states:

> As far as "Sherlock Holmes need not be a detective" is true *within the fiction,* the proposal has bad results. According to it, properties are ascribed to individual offices as their requisites. That means, however, that a fictional character (denoted by the use of a fictional name) has all its properties necessarily. If Sherlock Holmes is a detective it is necessary that Sherlock Holmes is a detective. In short, as far as there is a nontrivial modality within fiction (i.e. characters have at least some properties merely possibly), Glavaničová's proposal fails. (Kosterec 2018, 123)

Now, the last clause of the quotation is wrong. Glavaničová made it clear in (2018) that the sentences containing fictional names, when considered in the context *within the fiction* are not with supposition *de dicto*:

> Still, it may be useful to apply the pretense theory in *de re* context. We could get into the context of the story and speak about Sherlock Holmes *as if* he was a real person (despite there being really no such person). (Glavaničová 2018, 71)

It is correct, that in supposition *de dicto* the sentence *Sherlock Holmes is a detective* is necessarily true as soon as *being a detective* is considered *Holmes'* requisite. But that seems to be mainly the feature of use of that sentence in *meta-fictional contexts*. It is not accurate to claim (or it is at least not the only possible option), however, as Kosterec presupposed in 2018), that HV assumes that the fictional name *within the fictional contexts* is used with the supposition de dicto. On the other hand, though, HV seems to be then dependent on some other theory (e.g. pretense) when dealing with the supposedly true readings of sentences *Sherlock Holmes is a detective* within the fictional contexts (and therefore presumably in supposition de re).[9] HV

[9] To be more precise, there are no such supposedly true readings according to the theory (for the fiction contexts) – so there is no such dependence, *once we agree with the assumption*, that these sentences are either false or truth-valueless. If we do

is to explain the truth conditions of sentences simpliciter. It does not contain or explains truth-in-fiction condition of investigated sentences, … yet:

> More should be said about the generation of requisites, and my suggestion is to employ some plausible account for the computing of truth in fiction and generate requisites in terms of it. But that is a very nontrivial task for another time. (Glavaničová 2018, 67)

In general, we should discern when speaking about *truth simpliciter* and *truth-in-fiction*. We should differ between *modal properties of stories* and *modal properties of sentences within the stories*. Glavaničová seems to consider stories analogous to axiomatic theories (see Glavaničová 2018, footnote 18). It seems possible, that Doyle could write different Holmes stories. What, on the other hand, seems to be a consequence of HV is that there are some essentials to the Holmes stories Doyle wrote in this possible world. That it is necessary for these stories, that *Sherlock is a detective*, however, shall not be care freely considered equivalent to the modality of that sentence *within the fiction*. These are in the end, *different* sentences according to HV.

Acknowledgements

I am indebted to anonymous reviewers for their helpful comments and suggestions. I would also like to thank Carolyn Benson for correcting my English.

Funding

This paper was supported by VEGA Grant No. 2/0117/19 Logic, Epistemology, and Metaphysics of Fiction.

References

Fine, Kit. 1982. "The Problem of Non-Existence I-Internalism." *Topoi* 1 (1-2): 97–140. https://doi.org/10.1007/BF00157548

Glavaničová, Daniela. 2017. "Tichý and Fictional Names." *Organon F* 24 (3): 384–404.

not, however, there is a need for something more. Again, I thank an anonymous reviewer for pressing me to clarify on this point.

Glavaničová, Daniela. 2018. "Fictional Names and Semantics: Towards a Hybrid View." In *Objects of Inquiry in Philosophy of Language and Literature. Studies in Philosophy of Language and Linguistics*, edited by Stalmaszczyk Piotr, 59–73. Berlin: Peter Lang. https://doi.org/10.3726/b14249

Glavaničová, Daniela. 2020. "Response to Kosterec." *Organon F* 27 (1): 129–34. https://doi.org/10.31577/orgf.2020.27106

Jespersen, Bjørn, Duží, Marie, Glavaničová, Daniela. 2020. "Impossible Individuals as Necessarily Empty Individual Concpets." In *Logic in High Definition: Developing Fine-Grained Semantics, Trends in Logic*, edited by Gioradni Alessandro and Malinowski Jacek. Dordrecht: Springer. https://doi.org/10.1007/978-3-030-53487-5

Kosterec, Miloš. 2018. "Some Problems of Glavaničová's Approach to Fictional Names." *Organon F* 25 (1): 120–25.

Organon F 28 (2) 2021: 492–495
https://doi.org/10.31577/orgf.2021.28210
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

BOOK REVIEW

Sanford Shieh: Necessity Lost: Modality and Logic in Early Analytic Philosophy, Vol. 1 Oxford University Press, Oxford, 2019, xxiv+441 pages.

Jaroslav Peregrin*

It seems to be clear that the truth value of a sentence such as *The most populous city in the world is Shanghai* can fluctuate: it is true now, but perhaps it was not true in the past and it will not be true in the future. And it is certainly not true in many imaginable states of affairs alternative to the current one. And it may, nowadays, seem obvious that when we talk about necessity and possibility we are talking about this kind of fluctuation: we say that something is *possible* if there are conceivable circumstances in which the corresponding sentence is true, and we say that it is *necessary* if all circumstances are of this kind. This intuition, it may further seem, was put on firm foundations by the Kripkean possible worlds semantics and its various elaborations.

However, it would be erroneous to suppose that such intuitions have been held by all eminent logicians. True, the logic of possibility and necessity was already considered as integral to the agenda of logic by Aristotle, and some medieval and post-medieval logicians even mused about possible worlds; but the fact is that some of the founding fathers of modern logic, notably Frege and Russell, considered necessity and possibility as something that does not belong to the core of the subject matter of logic, at least not so that we could have something as a modal logic.

Why this was so is explained in great detail in Sanford Shieh's book. This volume, the author tells us, will be followed by a second, to be devoted to Wittgenstein and C. I. Lewis, whose attitudes to modal logic were much more

* Institute of Philosophy of the Czech Academy of Sciences
https://orcid.org/0000-0002-0460-1933
Institute of Philosophy of the Czech Academy of Sciences, Jilska 1, 110 00 Prague 1, Czech Republic
peregrin@flu.cas.cz

positive than those of Frege and Russell. Given the ever expanding literature on the origins of modern logic and about Frege, Russell, Wittgenstein etc., one might ask whether new publications on this subject can add anything really new to the already existing body of findings. But Shieh, I think, has managed to hit on a novel viewpoint, from which some of the well-known events of the story appear in a relatively new light.

How could Frege, such a meticulous thinker, we may want to ask, overlook the fact that truth is relative to circumstances? The answer is simple: what is true or false, he was convinced, are primarily not sentences, but propositions (or *thoughts*, as he would call them), and their truth values are not relative to anything. Every proposition is either true, or false, and it is true or false once and for all.

So what about *The most populous city in the world is Shanghai*? Is it not so that it is true here and now, but it can be false in the future or in an alternative world? Frege would say it is true in force of the fact that it expresses a proposition that is true. And it can become false only in force of coming to express a *different* proposition. The proposition expressed by the sentence now is better expressed by *The most populous city in the world in 2021 is Shanghai*, while that expressed by *The most populous city in the world in 2121 is Shanghai* is a different proposition.

The part of the book devoted to Frege consists of five chapters. In the first, Shieh discusses Frege's early philosophy, as it appears especially in *Begriffsschrift*. This is where he discerns the main thesis of Frege's conception of judgment, namely that a judgement is the step from a representation (later called a *thought*, the sense of a sentence) to its truth value. This made him part ways with Kant, whose views form the baseline of Frege's philosophy. It is already here that Frege expresses his reluctance to consider necessity or possibility as pertaining to the content of judgment, or to accept that necessary truth and possible truth are subspecies of truth.

In Chapter 2 Shieh discusses what he calls Frege's *amodalism*, *viz.* the conviction that truth cannot be relativized in any of the ways we now know from Kripkean possible world semantics and its later variations. Why Frege holds this view is discussed in great detail in Chapter 3. Here the author pursues Frege's view, that a judgment is a step from a thought to its truth value, through to its consequences. Shieh argues that the theory of truth held by Frege was in the redundancy theory genre, according to which "recognizing the truth

of a thought supervenes on recognizing the obtaining of what that thought represents". Frege himself holds that truth is undefinable, and Shieh argues that it is this overall conception of truth that prevents Frege from accepting that it need not be absolute.

In Chapter 4 the author explores the details concerning Frege's explaining away the intuition that truth, especially the truth of sentences, is relative. As I mentioned above, Frege's basic explanation was that a sentence can change its truth value only when it changes the proposition (thought) it expresses; and propositions are absolute because they incorporate all the factors to which they could be relative: thus, a proposition is not relative to times, because it always incorporates a specific time to which it relates etc. In this chapter Shieh also discusses Frege's reluctance to assimilate necessity to other historically "tried and true" concepts, such as analyticity or apriority.

In Chapter 5 Shieh discusses Frege's general views on the nature of logic. According to him, the question central for Frege was "what is it for a thought to be self-justifying, and how do we know which thoughts are self-justifying?", and Frege's answer was "that a thought is self-justifying just in the case it is true in virtue of its logical structure."

So in the case of Frege, the situation is relatively transparent. In the third realm, where thoughts reside, there is no room for empirical circumstances, hence no room for non-absolute truth or falsity, and hence no room for necessity and possibility. The problem, of course, is that the truth values of the sentences we use to communicate do usually depend on the context of their utterance, on time, on the state of the world etc. The propositions expressed by such sentences would have to "absorb" all these dependencies. Not only would they have to contain the exact time to which they refer, but also an indication that we are in the world we are etc. This all makes Fregean thoughts extremely chimeric.

The situation is much more complex with Russell, whose intellectual journey was more tortuous. After flirting with Hegelian and Bradleyian idealism he (accompanied by Moore) developed his non-idealistic theory of propositions, after which he came to conclude that no such self-standing entities as propositions can exist. In Chapter 6 Shieh maps the twists and turns of Russell's philosophical journey from Bradley to his rejecting of idealism and his attempting at an account for the necessity of mathematics.

In Chapter 7 the author discusses the after effects of Russell's (and Moore's) parting ways with Bradley and developing his own theory of propositions. At the end of this period, Shieh claims, he reached a view of necessity not too

dissimilar to Frege's: he rejects necessity and possibility because he holds truth to be absolute. Shieh quotes Russell: "there seems to be no true proposition of which there is any sense in saying that it might have been false.... What is true, is true; what is false, is false; and concerning fundamentals, there is nothing more to be said. (*The principles of mathematics*, §430, 454)". In Chapter 8 Shieh anatomizes some further consequences of Russell's rejection of idealism.

In Chapter 9 Russell's general views of the nature of logic are discussed. Shieh claims that for Russell logic is primarily a theory of the relation of (material) implication standing among propositions and determining the logically valid inferences from propositions to propositions. Strangely, we come to know this relation by a process "akin to sense-perception". Chapter 10, the book's final chapter, then summarizes the reasons for Russell's rejection of necessity and possibility in his post-idealistic period. Russell, according to Shieh, maintains that the intuitions we have about necessity and possibility turn out, on close scrutiny, to be incoherent.

It might be considered an embarrassment not to present any criticism of a book under an extensive review. But Shieh's book is a fine piece of meticulous scholarship, with no glaring omissions. It clearly results from an immense amount of work. It is perhaps not deeply revelatory, for the details of Frege's and Russell's contributions to logic have already been thoroughly explored, but Shieh unleashes a novel slant, which allows him to bring to light some connections that were not discernible before. In this sense, the book ranks alongside the most important contributions to the exploration of the history of modern logic.