Croatian Journal of Philosophy

Articles

Book Reviews

Vol. XXI · No. 63 · 2021

Croatian Journal of Philosophy

1333-1108 (Print)
1847-6139 (Online)

Co-published by
"Kruzak d.o.o."
Naserov trg 6, 10020 Zagreb, Croatia
fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr
www.kruzak.hr

and

Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr
www.ifzg.hr

Available online at http://www.ceeol.com and www.pdcnet.org

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XXI · No. 63 · 2021

Articles

Book Reviews

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.1
Received: June 6, 2021
Accepted: November 15, 2021

Argumentation as a Speech Act: A (Provisional) Balance

PAOLO LABINAZ
Department of Humanities, University of Trieste, Trieste, Italy

This paper investigates whether, and if so, in what way, argumentation can be profitably described in speech-act theoretical terms. I suggest that the two theories of argumentation that are supposed to provide the most elaborate analysis of it in speech-act theoretical terms (namely van Eemeren and Rob Grootendorst's Pragma-Dialectics and Lilian Bermejo-Luque's linguistic normative model of argumentation) both suffer from the same two flaws: firstly, their "illocutionary act pluralism" assumption and secondly, a lack of interest in where arguing belongs in the classification of illocutionary acts. I argue that these flaws derive from the authors' reliance on an intention-based speech-theoretical framework. Finally, I adopt a deontic framework for speech acts in order to propose an alternative way of accounting for argumentation which seems to overcome the two limitations outlined above. According to this framework, argumentation may be conceived as a speech act sequence, characterized by the conventional effects brought about by the communicative moves (as illocutionary acts) of which it is composed.

Keywords: Speech act theory; argumentation; J. L. Austin; illocutionary force; verdictives.

1. *Introduction*

The goal of this paper is to investigate whether, and if so, in what way, argumentation can be profitably described in terms of speech-act theory. On the one hand, although most argumentation theorists conceive argumentation as a social and communicative activity, only a few of them have gone further by analyzing it in detail as a speech act. At the same time, while speech-act theorists have indeed referred to the act

of arguing as a kind of illocution, they have failed to provide any systematic description of its illocutionary force. It is thus still a relatively open question whether and to what extent construing argumentation as a speech act can help shed light on its social and communicative function.

The paper is organized as follows: the next section (Section 2) considers how argumentation has been characterized as a social and communicative activity in argumentation studies. I then turn to the two most elaborate accounts which have analyzed it in terms of speech-act theory, namely van Eemeren and Rob Grootendorst's Pragma-Dialectics and Lilian Bermejo-Luque's linguistic normative model of argumentation (Section 3). After that, I discuss some of the problems associated with their (intention-based) speech-act theoretical frameworks which I believe make their accounts descriptively inadequate (Section 4). Finally, I suggest an alternative way of accounting for argumentation from a speech-act theory perspective, considering it not as a specific kind of speech act complex, but rather as a speech act sequence, characterized by the conventional effects brought about by the communicative moves (as illocutionary acts) of which it is composed (Section 5).

2. *From argument to the act of arguing*

Over the last 30 years or so, there has been a gradual shift in interest in argumentation studies from arguments as semantic structures to arguments as communicative moves (van Eemeren et al. 2014: 27–39). This shift has led argumentation theorists to focus on the conditions under which argumentation is established and carried out within different communicative contexts. Their analyses are characterized by a pragmatic approach to argumentation: communication is conceived as a goal-oriented activity, in which argumentation is used for the attainment of goals. Despite this common approach, as van Eemeren and his colleagues have pointed out (2014: 29) "as yet, there is no unitary theory of argumentation that encompasses the logical, dialectical, and rhetorical dimensions of argumentation and is universally accepted". Let us now take a brief look at some of the most significant examples of how argumentation has been defined by argumentation theorists from their own different perspectives.

According to Frans van Eemeren and Rob Grootendorst (2004: 1), argumentation is to be conceived as "a verbal, social, and rational activity aimed at convincing a reasonable critic of the acceptability of a standpoint". By focusing on the argumentative interaction between proponents and (potential) opponents of a standpoint, they elaborate an ideal model of a critical discussion, which specifies stages and rules involved in a rational dialectical procedure (van Eemeren and Rob Grootendorst 2004: 57–62). Whereas Douglas Walton (1990: 411) focuses primarily on arguments, considering them as moves occurring in a certain con-

text of dialogue. According to him, there are different types of dialogue which are characterized by different goals (such as persuasion dialogue, negotiation, inquiry etc.) and each of these represents a different conversational context of argument use (see Walton 1998). Accordingly, argumentation is assumed to contribute to the attainment of the goal of the type of dialogue concerned. Ralph Johnson (2000: 12, 154) takes a difference stance towards argumentation, construing it as "the sociocultural activity of constructing, presenting, interpreting, criticizing, and revising arguments". An argument consists in a discourse (or text) through which an arguer aims to persuade her interlocutors of the truth of an already made claim by providing reasons in support of it. This is what Johnson considers the illative core of an argument, distinguishing it from the dialectical tier "in which the arguer discharges his dialectical obligations" (Johnson 2000: 168). Finally, David Hitchcock (2002: 326–327) has adopted a modified version of Johnson's definition of argumentation. According to him, in describing the goal of the practice of argumentation as that of persuading someone of the truth of a claim, Johnson wrongly focuses on the arguer's expected outcome, and not on the function of the practice itself. In Hitchcock's view, the purpose of argumentation is to be conceived as "reaching a shared rationally supported position on some issue" (Hitchcock 2002: 327) because it is only by conceiving its specific purpose in this way that we can avoid equating it to a rhetorically-driven communicative practice.

Whether argumentation is conceived as aimed at convincing a reasonable critic of the acceptability of a standpoint (van Eemeren and Grootendorst 2004), or at contributing to the goal of different types of dialogue (Walton 1998), or at reaching "a shared rationally supported position on some issue" (Hitchock 2002), or more simply at rationally persuading the other(s) of the truth of a thesis (Johnson 2000), all these goals are assumed to be pursued by engaging in the social and communicative activity of making arguments. In making an argument, we engage in a reasoning process: what we do is make connections between certain claims and an already made a claim in such a way as to support the latter with the former. When we present the argument to an addressee, we do so because we want to convince her that the justificatory link we have established between those claims holds. In so doing, we have performed an act of arguing. Regardless of its proper purpose, then, argumentation as a social and communicative activity necessarily involves (at least) one or more acts of arguing. This means that if we want to understand the communicative and social function of argumentation, we need to investigate the specific characteristics of this act. At this point, the first task to be dealt with is to establish what is constitutive of any communicative move that counts as an act of arguing. But one also has to identify its components: does it consist only in the claims provided in support of a certain claim, or is the claim being supported also part of it?

3. *Argumentation as a speech act*

It is widely assumed among argumentation theorists that the framework provided by speech act theory can be of help for an in-depth investigation of the act of arguing. According to the most representative speech-act theorists, arguing is to be understood as a type of illocution. John L. Austin, the founder of the speech act theory, includes the verb "to argue" within the class of the performative verbs and considers it as denoting an expositive illocutionary act (see Austin 1975: 160–162). According to John Searle and Daniel Vanderveken (1985: 184), on the other hand, the act of arguing belongs to the class of the assertives. They point out that arguing has much in common with another member of this class, namely assuring: indeed, both are associated with the perlocutionary intention of convincing an addressee of the truth of a claim since people rely on them when someone has doubts about the truth of an already made a claim. However, only in the case of arguing does the speaker try to fulfill the perlocutionary intention by providing reasons in support of that claim. Be that as it may, neither Austin nor Searle and Vanderveken have provided a systematic description of its illocutionary force. Argumentation theorists, too (with a few notable exceptions) have not devoted much energy into making an in-depth investigation of argumentation as an illocution. In the next two sub-sections, we consider the only two theories of argumentation which, in my opinion, analyze it thoroughly in speech-act theoretical terms. We begin with the first (historically) complete analysis of argumentation as a speech act, proposed by van Eemeren and Grootendorst (1983, 1992) within their pragma-dialectical theory. We then examine Lilian Bermejo-Luque's linguistic normative model of argumentation, which differs substantially from van Eemeren and Grootendorst's analysis (see Bermejo-Luque 2011).

3.1 *Argumentation as an illocutionary act complex*

In Van Eemeren and Grootendorst's normative model of a critical discussion, argumentation comes into play at its third stage. Indeed, argumentation is carried out after the parties involved in the discussion have established what dispute is at stake (confrontation stage) and after they have assumed the discussion roles of proponent and opponent (opening stage) (Van Eemeren and Grootendorst 1992: 35).

Based on a revised version of Searle's speech act theory (see Searle 1969, 1979), Van Eemeren and Grootendorst (1983: 39) describe argumentation as an illocutionary act complex (also speech act complex) consisting of a certain number of "elementary" illocutionary acts (more specifically, assertive speech acts) that stand "[…] in a justifying or refuting relation to an expressed opinion (which consists of statements acting as a claim or conclusion)".[1] A speaker usually performs a speech

[1] The expressed opinion can be any proposition because it "may refer to facts or ideas […] actions, attitudes, and so on" (van Eemeren and Grootendorst 1983: 5).

act of this kind in reaction to an addressee either expressing doubts about a claim previously made by her or putting forward a claim contradicting it. When this happens, in order to defend her claim, the speaker advances one or more assertive speech acts in its support. It is the complex of these assertive speech acts that, taken together, constitutes the performance of the complex illocutionary act of arguing. From a critical discussion perspective, this act should be aimed at contributing to resolving the difference of opinion as to the claim at issue between speaker and addressee (van Eemeren and Grootendorst 1992: 28–30).

Van Eemeren and Grootendorst (1992: 29) point out that the main difference between the illocutionary act complex of argumentation and the "elementary" illocutionary acts composing it is that while the latter's communicative function operates at the sentence level, that of the former works at some higher textual level. However, they hold that it has the same sort of felicity conditions as an "elementary" illocutionary act. In formulating its felicity conditions, van Eemeren and Grootendorst take inspiration from Searle's distinction between propositional, preparatory, sincerity and essential conditions, dividing them into identity and correctness conditions. As to the identity conditions, which comprise Searle's propositional and essential conditions (see Searle 1969: 57–61), these must be satisfied for a constellation of assertive speech acts to be recognized as the felicitous performance of the illocutionary act complex of argumentation. When speaking of correctness conditions, van Eemeren and Grootendorst refer to what is required for the performance of this act complex to be regarded as appropriate. These conditions include Searle's preparatory condition as well as what they call the "responsibility condition", which replaces Searle's sincerity condition in order to emphasize the commitment one incurs by virtue of being recognized as arguing (van Eemeren and Grootendorst 1992: 30–33).

Let us consider how these felicity conditions apply to the illocutionary act complex of argumentation.

As to the identity conditions, this illocutionary act is felicitously performed (and can thus be recognized as such) if the following two conditions are met:

1. Propositional content condition: utterances 1, 2, ..., n constitute the elementary speech acts 1, 2, ..., n, in which a commitment is undertaken to the propositions expressed.
2. Essential condition: the performance of the constellation of speech acts that consists of the elementary speech acts 1, 2, ..., n counts as an attempt by the speaker to justify p, that is to convince the listener of the acceptability of his standpoint with respect to p.

Furthermore, the performance of the illocutionary act complex of argumentation is regarded as appropriate if the following conditions are met:

3. Preparatory conditions:
   a. The speaker believes that the listener does not accept (or at least not automatically or wholly accept) his standpoint in respect to *p*.
   b. The speaker believes that the listener is prepared to accept the propositions expressed in the elementary speech acts 1, 2, ..., *n*.
   c. The speaker believes that the listener is prepared to accept the constellation of elementary speech acts 1, 2, ..., *n* as an acceptable justification of *p*.
4. Responsibility conditions:
   a. The speaker believes that his standpoint with respect to *p* is acceptable.
   b. The speaker believes that the propositions expressed in the elementary speech acts 1, 2, ..., *n* are acceptable.
   c. The speaker believes that the constellation of the elementary speech acts 1, 2, ..., *n* is an acceptable justification of *p*. (van Eemeren and Grootendorst 1992: 31)

When only the identity conditions are met, the illocutionary act complex of argumentation counts as felicitous, but is regarded as failing to convince the listener of the acceptability of the standpoint in support of which the act has been performed. If the identity conditions are not met either, then the constellation of elementary assertive illocutionary acts would not be recognized as counting as an attempt to argue in favor of a certain standpoint.

It is to be noted that van Eemeren and Grootendorst (1992: 26–27, 33) distinguish between the communicative and interactional aspects involved in the illocutionary act complex of argumentation. While the communicative aspect has to do with the speaker's communicative intention, that is, to make one's audience understand that in making a certain constellation of statements she intends to justify a previously expressed opinion, its interactional aspect is concerned with convincing the audience of the acceptability of that opinion. In van Eemeren and Grootendorst's view, the interactional aspect links the speech act of arguing with the perlocutionary act of convincing. A listener is said to be convinced when she is prepared to accept the expressed opinion to which the speaker's argument relates. According to them, this acceptance amounts to the perlocutionary effect conventionally associated with argumentation (van Eemeren and Grootendorst 1983: 65–69). This effect must be an intentional one on the part of the speaker: it requires a full understanding of the speech act and is "[…] partly dependent on rational considerations on the part of the listener" (van Eemeren and Grootendorst 1983: 28).

3.2 *Argumentation as a second-order illocutionary act complex*

Lilian Bermejo-Luque (2011: 62), too, considers argumentation as a composite illocutionary act, but conceives it as consisting of two distinct illocutionary acts, namely the act of adducing (a reason) and that of concluding (a target-claim).[2] In her view, these are second-order illocutionary acts, because "they can only be performed by means of first order speech-acts [...]" (Bermejo Luque 2011: 60). More specifically, two first-order speech acts count as an act of adducing (*R*) and an act of concluding (*C*), respectively, if an addressee attributes to the speaker performing them an implicit inference-claim (*I*) of the form "if *R* (the content of the reason adduced), then *C* (the content of the target-claim)". In doing so, the addressee interprets the speaker as performing the "second-order illocutionary act complex" of arguing. Consider the following example by Bermejo-Luque (2011: 60). When uttering something like "I promise I'll take care, so don't worry", at first sight the speaker appears to be performing the first-order illocutionary acts of promising and requesting, respectively. However, if the addressee takes it that these two illocutionary acts are connected with each other by an inference-claim of the type "If I promise I'll take care, then you should not be worry", then "I promise I'll take care" would count as the second-order illocutionary act of adducing (that the speaker commits herself to take care) and "Don't worry" as the second-order illocutionary act of concluding (that the addressee should not be worried), respectively. In uttering "I promise I'll take care, so don't worry", then a speaker would normally be regarded as arguing.

In order to account for the illocutionary force of the second-order illocutionary act complex of argumentation, Bermejo-Luque (2011: 61) relies on the Speech-Act Schema developed by Ken Bach and Robert Harnish (1979), considering it better suited to this job than van Eemeren and Grootendorst's speech-act theoretical framework. Bach and Harnish's Speech-Act Schema constitutes an inferential pattern that an addressee has to follow in order to determine the content and the force of an illocutionary act (Bach and Harnish 1979: 4–7). Most importantly, this inferential pattern is based upon three presumptions: the linguistic presumption, the communicative presumption and the presumption of literalness (Bach and Harnish 1979: 7, 12). Here, we are interested in the communicative presumption. According to this, whenever a speaker utters a certain sentence, she is doing so with some recognizable intention. Were this not so, it would be impossible, or nearly impossible, to identify the illocutionary force of her utterances. According to Bermejo-Luque (2011: 61), attributing to a speaker the implicit inference-claim (*I*) of the form "if *R*, then *C*" to which we

[2] According to Bermejo-Luque (2011: 59), van Eemeren and Grootendorst's analysis differs from hers because, by speaking of argumentation as consisting of the assertive speech acts put forward in support of a previously made claim, they erroneously equate it to the act of adducing.

referred previously is based on a similar presumption: it amounts to the mutual belief (of both addressee and speaker) that the speaker, together with performing the second-order speech acts of adducing (*R*) and of concluding (*C*), is also intending to implicitly assert "if *R*, then *C*" (argumentative presumption). Bermejo-Luque (2011: 62) points out that an addressee makes this presumption whenever the speaker uses epistemic qualifiers such as "probably," "necessarily," "evidently" and so on, or expressions like "so," "therefore," "since," and "consequently". It is precisely the fact that these expressions are attached to some first-order speech acts (constative ones, in particular) that usually prompt the addressee to interpret such speech acts as second-order speech acts constituting the illocutionary act complex of argumentation.

Given the conventional nature of illocutionary acts, Bermejo-Luque (2011: 68–69) too acknowledges that, as van Eemeren and Grootendorst have observed, there must be some necessary and jointly sufficient conditions the satisfaction of which makes a certain performance count as an act of arguing. Following Searle's distinction between preparatory, sincerity and essential conditions, she formulates the following conditions:

Preparatory conditions:

(i) *S* believes that a claim *R*, having such and such pragmatic force, may be taken to be correct by *L*

(ii) It makes sense to attribute to *S* a conditional claim, with a certain pragmatic force, whose antecedent is "*R is correct*," and whose consequent is "*C is correct*"

(iii) S takes the correctness of a claim C to be in question within the context of the speech-act

Propositional content conditions:

(v) The content of the reason is that a claim *R'* is correct

(vi) The content of the target-claim is that a claim *C'* is correct

Sincerity conditions:

(vii) *S* believes the propositional content of *R* in a certain way and to a certain extent, namely, the way and extent that correspond to the pragmatic force of the claim *R'*

(viii) *S* believes that *R* being correct is a means to show that a target-claim *C* is correct

(ix) *S* believes the propositional content of *C* in a certain way and to a certain extent, namely, the way and extent that correspond to the epistemic pragmatic force of the target-claim *C*

Essential conditions:

(x) Adducing *R* with such and such pragmatic force is a means to show that a target-claim *C* is correct

(xi) *S* aims to show that a target-claim *C* is correct. (Bermejo-Luque 2011: 70–71)

Although, according to Bach and Harnish, felicity conditions have no role to play in their Speech Act Schema, Bermejo-Luque thinks that it

can be a useful interpretative tool for considering whether conditions (i)–(xi) have been fulfilled in a certain situation.

If we look at the essential conditions, which specify what constitutes the performance of a given type of speech act, we notice that, according to Bermejo-Luque, what characterizes the act of arguing amounts to its "[…] being aimed at showing a target-claim to be correct—that is, aimed at justifying" (Bermejo-Luque 2011: 55). It is this goal that "[…] conventionally renders any piece of communication argumentation" (Bermejo-Luque 2011: 58). This means that an audience recognizes the act of arguing being made "by recognizing a speaker's communicative intention of showing a target-claim to be correct" (Bermejo-Luque 2011: 24), regardless of her ulterior (perlocutionary) intentions. In contrast with the standpoint of van Eemeren and Grootendorst, Bermejo-Luque does not consider the goal of convincing to be a necessary condition of the act of arguing since it is only one of the possible goals that we may pursue by means of it (Bermejo-Luque 2011: 59). However, argumentation as a persuasive device plays a fundamental role in communication: it aims at achieving "[…] rational persuasion rather than mere conviction" (Bermejo-Luque 2011: 58), since its persuasive power depends on the reasons provided in support of the target-claim.

4. *Illocutionary act pluralism and illocutionary classes*

The two accounts presented above have the great merit of drawing attention to the speech act status of argumentation. Both accounts focus on the illocutionary aspect of the act of arguing and conceive of it as comprising the speaker's having a certain communicative intention (basically, the intention of arguing), meant to be recognized as such by the audience. While van Eemeren and Grootendorst assume that to argue is to intend to make one's audience understand that in making one or more assertive speech acts one is attempting to justify a certain standpoint in order to convince them of its acceptability, Bermejo-Luque foregrounds the intention to show that the target-claim is correct or to justify it. Both accounts also pay some attention to the perlocutionary goal typical of arguing: in the former, the association of arguing with the perlocutionary goal of convincing the audience of the acceptability of the target-claim follows on directly from the way in which the communicative intention of arguing is specified, while in the latter, the act of arguing, the main aim of which is justificatory, is granted an additional persuasive function (which, however, does not amount to a condition for its performance). Furthermore, there is a fundamental difference in the way the two accounts characterize the elements composing the complex illocutionary act of arguing: van Eemeren and Grootendorst identify this act with the reason(s) presented in support of the target claim, whereas Bermejo-Luque refers to it as the core unit of reason(s) and target claim.

While I have some sympathy for van Eemeren and Grootendorst's and Bermejo-Luque's general approach to analyzing argumentation as a speech act, it seems to me that the speech-act theory frameworks on which they base their analyses suffer from two significant drawbacks common to both: the first regards the assumption of "illocutionary act pluralism" underlying both analyses, while the second has to do with where to place arguing in the classification of illocutionary acts. In my opinion, these problems must be addressed if we hope to gain an understanding of the actual explanatory power of an analysis of argumentation as a speech act.

Let me start with the issue related to the "illocutionary act pluralism" assumption, according to which an utterance token can carry out more than one illocutionary act. This assumption is at the core of both of the analyses presented above: van Eemeren and Grootendorst (1983: 32) state explicitly that "sentences uttered in argumentation in fact have *two illocutionary forces simultaneously*" and Bermejo-Luque (2011: 59, 60) agrees with them by defining the illocutionary act of arguing, as well as the illocutionary acts composing it (i.e., adducing and concluding), as "second-order" illocutionary acts. Indeed, according to their analyses, assertive (or, using Bermejo-Luque's terminology, constative) speech acts can be conceived in certain conditions as also having a different illocutionary force, and thereby another communicative function. While assertive speech acts are aimed at presenting their propositional content as true, when they are involved in argumentation, they are also designed to either express a certain standpoint or provide support for it.

In traditional speech act literature, the two most studied cases of illocutionary plurality are those in which (i) illocutionary force indicators either contained in a certain utterance or accompanying it (in the form of paralinguistic features) are taken to indicate one illocutionary force or another (e.g., the verb in the future tense in a sentence such as "I will call a lawyer" may be taken to indicate a promise, a warning or even a prediction) and (ii) utterances with linguistic features which indicate a certain illocutionary force (e.g., the interrogative form for a question such as "Can you pass the salt, please?") are interpreted as performing a different illocutionary act (e.g., an indirect request), the so-called "indirect speech acts" (Searle 1979: 30–57). In both of these cases, as well as in those involving argumentation, a crucial role is played by the addressee's uptake. While, on the one hand, the speaker may do everything possible to get the addressee to understand the illocutionary force of her utterance, on the other, recognition of the illocutionary act the former is purporting to perform must always be ratified by the latter. Indeed, the addressee's recognition of the speaker's communicative intention is partly constitutive of its fulfilment. If cases of illocutionary plurality such as (i) can be explained locally (that is, depending from time to time on the distinguishable features of the

context of utterance), cases of indirect speech acts and those involving argumentation require something more, and namely accounting for the general possibility of illocutionary plurality. In the standard account of indirect speech acts, illocutionary plurality is explained by identifying a primary illocutionary act (the indirect request) which is performed through the performance of a "secondary" (or literal) elementary illocutionary act (the question) (Searle 1979: 33–36). However, as Searle himself explains, in such cases the speaker is recognized as having the communicative intention to perform only the "primary" illocutionary act. What about cases involving argumentation? Can the standard account of indirect speech acts be applied to them? It seems to me that it cannot, for two main reasons. Firstly, indirect speech acts are elementary (or first-order) illocutionary acts (such as requests, rejections and so on), whereas in the case of argumentation, there is a leap from elementary (or first-order) illocutionary acts to a composite (or second-order) illocutionary act. Secondly, indirect speech acts involve the speaker having the communicative intention to perform only the "primary" illocutionary act, while according to the accounts presented above, when arguing, a speaker has the communicative intention(s) of performing both elementary (or first-order) and composite (or second-order) illocutionary acts. This means that when attempting to argue, a speaker should be expressing two distinct communicative intentions: the one associated with the performance of each of the assertive (or constative) speech acts which make up the illocutionary act complex of arguing, and the one associated with its performance. But how is it possible to express two distinct communicative intentions when issuing the same utterance? It may be conceded that the same sentence uttered at two different points in the same conversation can express different communicative intentions. If so, its utterances will be the vehicle of different illocutionary acts at different stages in the conversation. But that is not what van Eemeren and Grootendorst and Bermejo-Luque seem to suggest: indeed, as stated above, both sides are in agreement that "sentences uttered in argumentation in fact have *two illocutionary forces simultaneously*". What is needed, then, is an explanation of how an utterance token may be recognized as having two different communicative intentions and hence two different illocutionary forces at the same time. Moreover, it remains to be seen whether this condition only regards sentences uttered in the attempt to argue, or can be expanded to include sentences in discourses that are not argumentative. Indeed, if it only concerns sentences uttered in argumentation, one might think that what both van Eemeren and Grootendorst and Bermejo-Luque have provided is simply an *ad hoc* explanation of how some assertive (or constative) speech acts (taken together) can count as the performance of the illocutionary act complex of arguing.

The second problem concerns the place of arguing in the classification of illocutionary acts. It is widely recognized among speech act theo-

rists that illocutionary acts of different kinds can be divided into different classes (Searle 1969: 66–67, 1979: 1–29, Bach and Harnish 1981: 39–59, 108–119; for a general overview, see Kissine 2013). Surprisingly, neither van Eemeren and Grootendorst nor Bermejo-Luque suggest to which class of illocutionary acts arguing may belong. Since, according to them, it is composed of assertive (or constative) speech acts, one might think that it might be conceived as belonging to the assertive class. But this is not possible. Indeed, as said before, van Eemeren and Grootendorst hold that there are two illocutionary forces simultaneously associated with the sentences involved in argumentation: one is of the assertive kind, while the other is the force characteristic of the act of arguing. This means that the two illocutionary forces cannot coincide. Accordingly, arguing cannot belong (if not indirectly) to the class of the assertives. Another possibility is that the illocutionary act of arguing belongs to a class of illocutionary acts whose members are all illocutionary act complexes (or second-order illocutionary acts). However, an illocutionary act's being composed of more than one element is not a relevant criterion for identifying the members of a certain class. Think of another potential member of this class, namely explaining. It has something in common with arguing: both of them are composed of more than one element (explananda/explanandum, and reason(s)/target claim, respectively) and they also share some of the same linguistic indicators, such as the connectives "because" and "since" (which can be used as indicators for argumentation and for explanation), as well as "therefore" which, as well as indicating the conclusion of an argument, can also be used in explanations. However, these are superficial similarities. If we look at their communicative function, the differences between them become very evident. Indeed, when we have to give an explanation for some fact or event, and this occurs when we recognize that the addressee has some cognitive need or interest with respect to that fact or event, we do not take its occurrence (or the statement asserting that it occurs) as problematic. Instead, as we already know, in the case of argumentation, we aim to show that a certain claim (which has been questioned or may be questioned by someone) is well-grounded. It would therefore seem arduous to place argumentation and explanation in the same class of illocutionary acts, and the same might be said of other illocutionary act complexes (if they exist). If we rely, then, on the speech-act theoretical frameworks proposed by van Eemeren and Grootendorst and Bermejo-Luque, respectively, we must assume that the illocutionary act complex of arguing is situated outside any classification of illocutionary acts. But this is not in line with the most elaborate attempts to theorize about speech or illocutionary acts, all of which aim to establish very clear boundaries between different kinds of illocutionary acts.

Since the accounts of argumentation as a speech act proposed by van Eemeren and Grootendorst and Bermejo-Luque, respectively, do not seem to have the resources to deal with the problems just discussed,

I think it might be concluded that they are not descriptively adequate, at least from a speech-act theoretical perspective.

5. *Argumentation as a speech act sequence*

In this section, I would like to suggest an alternative way of accounting for argumentation from a speech-act theoretical perspective. To do so, in place of the intention-based framework for speech acts assumed by van Eemeren and Grootendorst and Bermejo-Luque, I shall adopt a deontic one.[3] According to the intention-based framework, the core illocutionary effect of a speech act consists of bringing about the addressee's recognition that the speaker intends to perform a certain kind of speech act, while the deontic one characterizes speech acts as having illocutionary effects that go beyond the "securing of uptake" (Sbisà 2007, 2009). These are conventional effects which come into being in virtue of intersubjective (and therefore social) agreement,[4] made possible by the securing of uptake. According to the deontic framework, an illocutionary act consists in the achievement of its characteristic conventional effect, which can be described in terms of deontic modal attributes (e.g. rights or authority, obligations or needs and so on) to be assigned or unassigned to the relevant participants in the communicative situation at hand (Sbisà 1984, 2006: 164–167). More specifically, this effect amounts to a change in the deontic dimension of their interpersonal relationship, since it concerns what they are entitled or obliged or committed to do with respect to each other.

As I will try to show, the deontic framework adopted here can help overcome the two problems discussed in the previous section by considering arguing not as a specific kind of speech act complex, but as a speech act sequence characterized by the conventional effects brought about by the illocutionary acts of which it is composed. While one the one hand, argumentation requires a complex sequence of speech acts to be recognized as such on the other, any speech act involved in such a sequence may be thought of as bringing about a change in the deontic statuses of the participants in a communicative situation, thereby establishing what can or must (or cannot or must not) be done in the next stage of the sequence. At the same time, that speech act is performed against a normative background established by the performance of the speech act preceding it (if there has already been one).

Let us now consider how argumentation as a speech act sequence may unfold using the deontic approach to speech acts adopted here. First of all, this requires identifying which type of illocutionary act may be the initiator of the sequence. Relying on Austin's illocutionary

[3] This deontic framework is grounded on J.L. Austin's conception of the illocutionary act and its effects as further elaborated by Marina Sbisà (see Austin 1975: 116–117; Sbisà 1984, 2006).

[4] It is to be noted that agreement can be presumed to occur by default, that is, in the absence of evidence to the contrary.

classes (1975: 150–163), we can assume that it is the verdictive type. Austin characterizes verdictives as consisting in "[…] the delivering of a finding, official or unofficial, upon evidence or reasons" regarding "something—fact, or value—which is for different reasons hard to be certain about" (Austin 1975: 151, 153).[5] There are various subtypes of verdictives, ranging from official and institutional to informal and conversational ones, each of which can in principle be the initiator of an argumentative sequence. Indeed, there is usually a verdictive at the core of argumentation, the correctness of which (truth, rightness, etc.) is focused on as being problematic.[6] In fact, a speaker performing a verdictive presents herself as willing to take on responsibility for the correctness (truth, fairness etc.) of the judgment issued and as acting on the basis of a publicly recognizable cognitive competence (Sbisà 2019: 10–13). When a verdictive is performed, at least two possible moves become available to the audience: the preferred default option is agreement, while disagreement is the dispreferred. Both of them involves different rights or entitlements and commitments or obligations for the participants in the communicative situation. As for the audience, the performance of a verdictive assigns to them a double entitlement that is describable as a "can". On the one hand, when agreement occurs,[7] the audience can (is entitled to) issue further, related verdictives on the basis of the received one, or to use its content (in whole or in part) as a premiss in reasoning or as a ground for decisions. On the other hand, since verdictives are by definition open to objections, the audience can (is entitled to) challenge the verdictive issued by asking for the speaker's grounds for it (by expressing doubts, presenting a rebuttal and so on), thereby focusing on the commitment one incurs in performing a verdictive. However, being aware that anyone in the audience can ask for her grounds for performing a verdictive, the speaker may also try to prevent disagreement by advancing reasons in support of the verdictive she is about to perform. In terms of deontic states, the commitment one incurs in performing a verdictive can be specified as an "ought": it amounts to the speaker's obligation to show that a verdictive whose correctness has been questioned (or may be questioned) by an addressee, is well-grounded. So, arguing comes into play either when the speaker's verdictive is appropriately challenged or when the

[5] Bach and Harnish (1979: 109–119) too, identify a class of verdictives but delimit its members to official findings. Conceived in this way, verdictives are part of what they call "conventional" illocutionary acts: these are acts that are performed with the intention of following a convention which bring about changes in institutional states of affairs.

[6] Also in the case of official or institutional verdictives, the judgment issued can be informally criticized as problematic, even though what they establish as being the case cannot be changed (except by means of some other official or institutional speech act).

[7] Agreement can be explicit, but most of the time it is tacit (due to the absence of any manifestation of disagreement).

speaker herself would like to try to prevent disagreement with the audience. When this happens, the speaker is expected to issue other verdictives which are claimed or supposed to vouch for the correctness (truth, fairness etc.) of the targeted verdictive. In turn, the issuance of these verdictives will affect the deontic statuses of the participants in the communicative situation at hand because they will bring about the characteristic illocutionary effect described above. It is the sequence of these speech acts that constitutes a case of argumentation. Argumentation may thus be conceived as a speech act sequence made up of two or more verdictives, where one is supported (warranted, justified or established) by the other(s). Clearly, this sequence can be intertwined with other illocutionary acts (such as challenges, objections, verdictives expressing a counter-standpoint etc.). What matters to us is that each move as an illocutionary act involved in this sequence is constrained by the illocutionary effect of the one preceding it: in turn, its illocutionary effect establishes what the legitimate options for the next move are.

It is certainly true, as van Eemeren and Grootendorst point out, that argumentation can be conceived as occurring at a "higher textual level" than its components, but it seems to me that this is not enough to treat it as an independent speech act complex. Since the unfolding of argumentation as a speech act sequence is distinguished by the characteristic effect of the verdictives composing it, its illocutionary force is inherited from them. Indeed, just as when issuing a verdictive, a speaker commits herself to the correctness (truth, fairness etc.) of her judgment, so too, a speaker giving support to a verdictive already made is committed to the correctness or soundness of the argument presented, if and when its correctness or soundness (or its underlying criteria) is appropriately challenged. When a challenge is issued, the speaker advancing the argument is bound to justify her moves by defending the verdictives the argument consists of, as well as the argument itself. In particular, when defending an argument, the speaker has to make explicit the warrant licensing the move from the verdictives presented in support of the targeted verdictive to the targeted verdictive itself, as well as the backing for the warrant (Toulmin 2003: 97–107). However, warrants are typically not explicitly asserted, but retrievable (mainly by conventional implicatures generated by expressions such as "so", "therefore", "hence" or by particularized conversational implicature) insofar as an addressee is actually interested in evaluating the justificatory link that the speaker assumes to hold between the targeted verdictive and the verdictive(s) presented in its support (see also Labinaz and Sbisà 2018: 622–623). Since the warrant is typically conveyed by means of an implicature, and since to be recognized as such, an implicature does not need to meet any felicity conditions, it cannot be equated with an implicit assertion, as Bermejo-Luque has claimed (see Section 3.1) Only once the warrant is made explicit can it be dealt with as the content of an assertion or more generally, of a verdictive. Faced with the warrant, the audience can (is entitled to) either accept

the justificatory link it establishes or challenge it by asking for the speaker's grounds for taking it to be true. In response to this challenge, the speaker is expected to articulate the backing for the warrant of her argument. It is to be noted that backings are usually not explicitly asserted, since they often lie in the speech event's common ground (see also Labinaz and Sbisà 2018: 622–623). In this case, too, its content must be made explicit and conveyed by means of a verdictive, thereby committing the speaker to its correctness.

To sum up, it might be said that argumentation as a speech act sequence can be characterized through the normative relations that hold between the speech acts composing it. It starts with the commitment a speaker incurs in performing a verdictive which constrains how she must react if someone challenges the judgment issued (even though that commitment obviously does not constrain what the verdictive should be about in the first place). If the speaker seeks to discharge this commitment, as opposed to retracting it, then she must provide an argument by performing some additional verdictives acting as reasons in support of the verdictive already made. It is clear that not all verdictives occur within an argumentative sequence, but all of them can potentially be part of one (see also Brandom 1994: 167–168). Every time an addressee challenges a verdictive, or the speaker believes that he may do so, the latter is expected to argue for the correctness of her verdictive in order to resolve the disagreement with the former and move towards the preferred communicative option, namely agreement. In this respect, as Sally Jackson and Scott Jacobs have suggested, arguing may be said to play a key role in disagreement regulation (Jacobs and Jackson 1982: 226–227; Jacobs 1986).

6. *Concluding remarks*

In this paper, I have examined whether, and if so, in what way, argumentation can be profitably described in speech-act theoretical terms. I have suggested that the two theories of argumentation that are supposed to provide the most elaborate analysis of it in speech-act theoretical terms (namely van Eemeren and Rob Grootendorst's Pragma-Dialectics and Lilian Bermejo-Luque's linguistic normative model of argumentation) both suffer from the same two flaws regarding their "illocutionary act pluralism" assumption and the lack of interest in where to place arguing in the classification of illocutionary acts. In themselves, these two flaws are not fatal, but they significantly weaken the explanatory power of the two theories, at least from a speech-act theoretical perspective. I have proposed that this derives in part from their reliance on an intention-based speech-theoretical framework. As argued in Section 5, a more promising path presents itself if we turn to a deontic framework for speech acts, which seems able to overcome the two limitations outlined above. Based on this framework, we can conceive argumentation as a speech act sequence characterized by the

conventional effects brought about by the verdictive illocutionary acts composing it: conversational moves (as illocutionary acts) involved in such a sequence are constrained by the illocutionary effects of those preceding them: in turn, their illocutionary effects establish what the legitimate options for the next move are. Since, according to the deontic framework, illocutionary force is not to be equated with the speaker's intention, there is no need to attribute multiple intentions to the speaker when accounting for the illocutionary force of argumentation, as suggested by van Eemeren and Grootendorst's and Bermejo-Luque's analyses. Furthermore, since the unfolding of argumentation as a speech act sequence depends on the characteristic conventional effects of the verdictives involved, we may say that it inherits its illocutionary force from them. This said, I am aware that, before drawing definite conclusions about the speech act status of argumentation, much more should be done to further develop this analysis in terms of deontic states and apply it to real cases of argumentation.

References

Austin, J. L. 1975. *How to Do Things with Words*. Oxford: Oxford University Press.

Bach, K., and Harnish, R. 1979. *Linguistic Communication and Speech Acts*. Cambridge: The MIT Press.

Brandom, R. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.

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

Eemeren van, F. H., and Grootendorst, R. 1983. *Speech Acts in Argumentative Discussions*. Dordrecht: Foris.

Eemeren van, F. H., and Grootendorst, R. 1992. *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective*. Hillsdale: Lawrence Erlbaum.

Eemeren van, F. H., and Grootendorst, R. 2004. *A Systematic Theory of Argumentation: The Pragma-dialectical Approach*. Cambridge: Cambridge University Press.

Eemeren van, F. H., Garssen, B., Krabbe, E., Snoeck Henkemans, F. A., Verheij, B., Wagemans, J.H. 2014. *Handbook of Argumentation Theory*. Dordrecht: Springer.

Hitchcock, D. 2017. *On Reasoning and Argument: Essays in Informal Logic and Critical Thinking*. Dordrecht: Springer.

Kissine, M. 2013. "Speech Act Classifications." In K. Turner and M. Sbisà (eds.). *Pragmatics of Speech Actions*. Berlin: de Gruyter Mouton: 173–201.

Jacobs, S. 1986. "The Management of Disagreement in Conversation." In R. Grootendorst, A. J. Blair and C. S. Willard (eds.). *Argumentation: Across the Lines of Discipline*. Dordrecht: Foris: 229–240.

Jacobs, S., Jackson, S. 1982. "Conversational Argument: A Discourse Analytic Approach". In Cox, R. J., Willard, C. A. (eds.), *Advances in Argumentation Theory and Research*. Carbondale: Southern Illinois University Press: 205–237.

Johnson, R. H. 2000. *Manifest Rationality. A Pragmatic Theory of Argument*. Mahwah: Lawrence Erlbaum.

Labinaz, P. and Sbisà, M. 2018. "Argumentation as a Dimension of Discourse: The Case of News Articles." *Pragmatics & Cognition* 25 (3): 602–630.

Sbisà, M. 1984. "On Illocutionary Types." *Journal of Pragmatics* 8: 93–112.

Sbisà, M. 2006. "Communicating Citizenship in Verbal Interaction: Principles of a Speech Act Oriented Discourse Analysis." In: H. Hausendorf and A. Bora (eds.). *Analysing Citizenship Talk. Social Positioning in Political and Legal Decision-Making Processes*. Amsterdam: John Benjamins, Amsterdam: 151–180.

Sbisà, M. 2007. "How to Read Austin." *Pragmatics* 17: 461–473.

Sbisà, M. 2009. "Uptake and Conventionality in Illocution." *Lodz Papers in Pragmatics* 5: 33–52.

Sbisà, M. 2019. "Assertion among the Speech Acts". In S. Goldberg (ed.). *The Oxford Handbook of Assertion*. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780190675233.013.7.

Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. R. 1979. *Expression and Meaning*. Cambridge: Cambridge University Press.

Searle, J. R., Vanderveken, D. 1985. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Toulmin, S. 2003. *The uses of argument*. Updated edition. Cambridge: Cambridge University Press. First edition 1958.

Walton, D. 1990. "What is Reasoning? What is an Argument?" *Journal of Philosophy* 87: 399–419.

Walton, D. 1998. *The New Dialectic: Conversational contexts of argument*. Toronto: University of Toronto Press.

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.2
Received: June 6, 2021
Accepted: November 15, 2021

Arguing about Free Will: Lewis and the Consequence Argument

DANILO ŠUSTER
University of Maribor, Maribor, Slovenia

I explore some issues in the logics and dialectics of practical modalities connected with the Consequence Argument (CA) considered as the best argument for the incompatibility of free will and determinism. According to Lewis (1981) in one of the possible senses of (in)ability, the argument is not valid; however, understood in the other of its possible senses, the argument is not sound. This verdict is based on the assessment of the modal version of the argument, where the crucial notion is power necessity ("no choice" operator), while Lewis analyses the version where the central notion is the locution "cannot render false."Lewis accepts closure of the relevant (in)ability operator under entailment but not closure under implication. His strategy has a seemingly strange corollary: a free predetermined agent is able (in a strong, causal sense) to falsity the conjunction of history and law. I compare a Moorean position with respect to radical skepticism and knowledge closure with ability closure and propose to explain Lewis's strategy in the framework of his Moorean stance.

Keywords: The consequence argument; compatibilism; (in)ability; closure; radical scepticism; Moorean stance.

> Philosophy is certainly *not* self-expression. … philosophy, of course, is argument, and you can say, well, is the conclusion true and is the argument valid? Iris Murdoch[1]
>
> How can philosophical enquiry be conducted without a perpetual *petitio principii*? Frank Ramsey[2]

[1] Quoted from Setiya (2020: 66).
[2] Ramsey (1994: 2).

1. *Introduction*

When arguing about free will, the consequence argument (CA), also called "the" argument for incompatibilism, a "master" argument, or the "unavoidability argument", is widely regarded as the best argument for the incompatibility of free will and determinism. Briefly, the remote past and the laws of nature are not up to us. However, if determinism is true, each of our present actions is a consequence of the laws of nature and the remote past. Therefore, the consequences (including our present acts) are not up to us either, and no one enjoys free will (cf. van Inwagen, 1983: v).

The evaluation of this piece of reasoning as an *argument* depends on two questions. What does it mean that one "thing" is a *consequence* of another? Do we rely on implication or something stronger, logical consequence (entailment)? And how to evaluate the *premises*: what does it mean to say that something is (not) up to us? Sheer inescapability, lack of causal control, or something different altogether? What are the starting points in this debate? What is the role of *logic* in establishing the final conclusions about CA?

I will approach these questions by focusing on Lewis's (1981) reply to van Inwagen's version of the argument. According to Speak (2011: 115), "the single most influential contribution to the overall philosophical quality of the recent free-will debate is van Inwagen's careful development of what he has dubbed the 'Consequence Argument.'" Van Inwagen, on the other hand, compliments Lewis (1981) as "…the finest essay that has ever been written in defense of compatibilism—possibly the finest essay that has ever been written about any aspect of the free will problem" (van Inwagen 2008a: 330). Well, Lewis's article is a *refutation* of CA, so, at the end of the day, whom should we praise—defense or prosecution?

The logical core of CA is rule *Beta*, the transfer of powerlessness: roughly, "It is unavoidable that p and it is unavoidable that (if p, then q); hence, it is unavoidable that q." Since the past and the laws of nature are (supposed to be) fixed and unavoidable, it follows, via determinism (the past and the laws of nature together determine everything) and a Beta-like *transfer* principle, that each of our choices is unavoidable. Principle Beta officially entered the philosophical scene with the third (modal) version of CA (van Inwagen 1983: 93–105). However, Lewis (1981) discusses the original, non-modal version of CA (van Inwagen 1975), which is silent on Beta. So, how does Lewis criticize CA? Does he object to its inferential structure, or does he deny the premises? Is he a Beta "blocker" or does he use some other "pain killer" to block CA?

Although the topic has been much discussed, these questions have not been clarified in a satisfactory way. The answer is not simple. Lewis introduces two senses of ability, and a parallel distinction between the two senses of unavoidability nowadays constitutes the core of what is called the *main compatibilistic response* to CA. Understood in one of

the possible senses ("weak"), the laws of nature are not unavoidable, which makes Lewis a "fixity finesser" (Speak 2011: 121). Understood in the other of its possible senses ("strong"), rule Beta is invalid: the premises of CA are true, yet we are still able to act otherwise. In order to properly assess the question of (in)validity, we must enter into a discussion of modal principles and practical modalities—a terrain well-traversed (cf. Kapitan 1991, 1996, 2002 and 2011). Nevertheless, I hope to add new details and emphasize points that have not so far been noted. Lewis accepts closure of the relevant (in)ability operator under *entailment* but not closure under *implication* (some of the results of a slightly more technical nature are given in the *Appendix*).

Closure under entailment has a strange corollary. It is usually taken for granted that strong ability is *causal*. Suppose I have just put my hand down on my desk and this was a free but predetermined act. I could make it the case that my hand was raised. According to Lewis, if determinism is true, then the very act of raising my hand or some event caused by this act would *directly* falsify any sufficiently inclusive conjunction of history and law. To quote Lewis on a slightly different issue (1986: 292): "A marvelous power indeed! And with so little effort!" The incompatibilists view this result as a refutation of Lewis's response to CA and compatibilism in general, while some compatibilists tend to either neglect or deny it explicitly (Rummens 2019, Perry 2004, Kapitan 2011). Why would Lewis accept such a consequence? In this paper, I seek to *defend* Lewis's original strategy. It is not easy to resolve all the tensions generated by Lewis's interpretation but some recently published material sheds new light on this debate (Lewis et al. 2020 and Lewis 2020). From the general perspective of a Moorean stance on the free will issue (it's a Moorean fact that we often have a choice what to do), the strange corollary is plausible and defensible. I end the paper with some comments on the role of *logic* and *dialectics* in establishing the conclusions about CA.

2. *CA—the modal version*

Let me start with some standard terminology and notation: '□' stands for *broad* logical necessity (including metaphysical necessity); '→' is a sign of the material conditional; 'H' stands for a true proposition about the total state of the world at some moment in the distant past[3]; 'L' designates the conjunction of the laws of nature, and 'p' an arbitrary true proposition about the present or future (typically about an action, "S raised her hand at *t*"). A special modal operator 'N' is introduced to express "power" necessity, as opposed to free will. Van Inwagen (1983: 93) defines 'Np' as: *p* is true and no one has or ever had any choice about *p*. Determinism ('DET') is the thesis that the past (a complete specification of the universe at any given instant in the past) and the

[3] Van Inwagen uses 'P_0', but to be consistent with Lewis, I will use 'H' throughout.

laws of nature together determine everything. Two principles of inference are used:

Alpha $\Box p \vdash Np$
Beta $Np, N(p \rightarrow q) \vdash Nq$

The modal argument or the third version of the consequence argument (CA_3) is then:

(1)	$\Box [(H \& L) \rightarrow p]$	// DET (premise)
(2)	$\Box [H \rightarrow (L \rightarrow p)]$	// PL, 1
(3)	$N [H \rightarrow (L \rightarrow p)]$	// Alpha, 2
(4)	N H	// No choice about the past (premise)
(5)	$N (L \rightarrow p)$	// Beta – (3), (4)
(6)	N L	// No choice about the laws (premise)
(7)	N p	// Beta – (5), (6)

If determinism is true, then it is not within anyone's power to perform any actions other than those they do perform. Van Inwagen admits that Beta is the most challenging element of the argument to defend and the "most doubtful thesis the incompatibilist must accept" (van Inwagen 1983: 222). This proved to be the case: there are counter-examples to the original Beta, acknowledged by van Inwagen, who later wrote (2017: 118):

> I mistakenly supposed that the only way in which it could be that one had no choice about the truth-value of a proposition would be for the truth-value of that proposition to be in some way so firmly "fixed" that one was unable to change it. I did not see that there is another way for one to have no choice about the truth-value of a proposition: for that truth-value to be a mere matter of chance.

If we understand having a choice about the truth value of a proposition as being able to reliably *ensure* a certain result, we face the problem of chancy events not under our control. Suppose I do *not* toss a coin but could have done so. Let *p* be "the coin does not land heads" and *q* stand for "the coin does not land tails". It is then true that 'Np' (nobody can act to ensure that the coin lands heads); it is also true that 'Nq' (nobody is able to act to *ensure* that the coin lands tails). However, it does not follow that 'N (p & q).' This conjunction, "the coin does not land heads and the coin does not land tails" is true, but by flipping a fair coin, I could ensure that the coin lands tails or heads, so I am able to render the conjunction 'p & q' false. This is a counter-example to the principle of Agglomeration (McKay and Johnson 1996: 115):

$$Np \& Nq \vdash N (p \& q)$$

With Alpha impeccable, the invalidity of Agglomeration implies the invalidity of Beta (for details, see *Appendix*). To take our previous example: let *r* stand for "the coin is not tossed." It is still true that 'Np', but so is 'N (p → r)'. Nobody can ensure that "(p & ~r)" is the case (the coin does not land heads when tossed). Yet 'Nr' is false: I was able to toss the coin (cf. Carlson 2000: 283–84).

Faced with this type of counter-example to Beta, what logical options do incompatibilists have to amend CA? One could vary the interpretation of 'N', the type of consequence relation, or both. The truth value of a proposition can be beyond one's control because it is a matter of chance ("I am not able to win a fair lottery") or because it is fixed no matter what anyone does ("I am unable to construct a perpetuum mobile"). In the first case, although there is nothing I can do that *would* ensure the result, I *might* still (by sheer luck) achieve the happy outcome. In the second case, the laws of nature prevent me from doing anything that *might* possibly result in such a machine. The incompatibilists now usually opt for the second, "no matter what" notion of unavoidability (as suggested by McKay and Johnson, 1996: 118–119). According to Huemer (2000: 538), 'Np' is to be understood as "no matter what you do (among the things that you are able to do), *p*." Van Inwagen (2008b) later interprets 'Np' as "untouchable": *p* is true and no human being is or ever has been able to act in such a way that, if he or she did act that way, *p* might be or might have been false. This interpretation blocks the McKay-Johnson counter-example ('Np' and 'Nq' are now false; the coin *might* land heads when tossed). However, laws of nature and the distant past (premises 4 and 6 of CA_3) are supposed to be untouchable, true no matter what, so CA_3 remains a sound argument.

Are there any other options to revise CA? We might use the original "no choice" as "not being able to ensure the falsity of" interpretation for 'N' and proceed *directly*:

(1) $\Box [(H \,\&\, L) \rightarrow p]$ // DET (premise)

(2) $N\,(H \,\&\, L)$ // No choice about the past and the laws (premise)

(3) $N\,p$ // Beta 2 – (1), (2)

A different (weaker) rule is now used, *Beta 2*[4], which can be expressed as:

$$Np, \Box (p \rightarrow q) \vdash Nq$$

or,

$$\Box (p \rightarrow q) \vdash Np \rightarrow Nq$$

Logicians would describe the original Beta as closure of the relevant operator ('N') under *implication* and Beta 2 as closure under *entailment* or *simple* consequence: if *q* is a consequence of *p*, then N*q* is a consequence of N*p*. A McKay-Johnson type of counter-example to Beta is inefficient against Beta 2. Rewrite the previous example in the new form: 'Np, $\Box$ (p → r), so Nr'. The second premise is now false: it is possible, in a broadly logical way, that the coin does not land heads when tossed. The counter-example to Beta fails.

How about CA? A weaker rule requires a stronger premise. And now there remains one premise only: the conjunction '(H & L)' offers

[4] As designated by Warfield and Finch (1998), introduced by Widerker (1987) as *β′* but also called *Beta box*, *alpha 2* (Huemer 2000) and *T2* (Carlson 2000).

a description of what Finch and Warfield (1998: 523) call the "broad past", the complete state of the world at a time in the distant past, including the laws of nature. They maintain that the broad past is fixed (nobody can act so as to *ensure* that it is false), so the second premise of this version of CA is true. With Beta 2 impeccable, the argument is sound and the thesis of incompatibilism established.

To summarize: the original version of CA is based on the closure of the "no choice" operator ('N') under *implication*, but closure fails for this operator. A defender of a revised CA can use a stronger interpretation of power necessity or closure of the relevant operator under *entailment*. We now turn to Lewis's treatment of CA. How does he assess these rules and CA in general?

3. *Lewis and CA*

Lewis discusses the first non-modal version of the consequence argument (CA_1), and it is not immediately obvious where to locate the relevant transfer principles. In this version, the crucial notion of (in)ability is captured by the phrase "an agent can(not) render false a proposition that, …," describing abilities in terms of the agent's power "over" the truth-values of propositions. We start with the scenario in which a certain judge J did *not* raise his hand at the moment T, but could have done so, given the usual understanding of our abilities. *H* (the proposition that expresses the total state of the world in the remote past) and *L* (the conjunction into a single proposition of all laws of physics) are used as before. Instead of 'p' (an arbitrary true proposition), we now have 'P' denoting the proposition that expresses the state of the world at T. The argument is then (van Inwagen 1975: 191):

1. If determinism is true, then the conjunction *H* and *L* entails *P*.
2. If J had raised his hand at T, then *P* would be false.
3. If (2) is true, then if J could have raised his hand at T, J could have rendered *P* false.
4. If J could have rendered *P* false, and if the conjunction of *H* and *L* entails *P*, then J could have rendered the conjunction of *H* and *L* false.
5. If J could have rendered the conjunction of *H* and *L* false, then J could have rendered *L* false.
6. J could not have rendered *L* false.
7. If determinism is true, J could not have raised his hand at T.

J's (not) raising his hand is an arbitrary action of an arbitrary agent. So, if determinism is true, then it is not within anyone's power to perform any actions other than those they do perform.

The most plausible candidate for stating a Beta-like transfer principle of inference in this version is the fourth premise. According to van Inwagen, this principle seems to be *analytic*:

> This premise may be defended as an instance of the following general principle: If S can render *R* false, and if *Q* entails *R*, then S can render *Q* false (van Inwagen 1975: 192).

In the third version, we had "having no choice about"; in the first version, the central notion is "can(not) render false a proposition." To compare them, it is convenient to use the general logical framework of *practical* modalities. According to van Inwagen (1975: 189), we can translate "He could have rendered the proposition that he did not reach Chicago by midnight false" as "He could have reached Chicago by midnight," understood as the ascription of *ability* to an agent. We might then understand, "S could have rendered false a proposition that *p*" as, broadly, "S was able to act so that *p* would or might be false" and express this locution in terms of a modal operator of ability indexed to agents as 'A_S~p'. The contemporary framework of agency-centered modalities is called the *stit* theory. The agent α *s*ees *t*o *i*t *t*hat *A* is defined as true at a certain moment *m* of the world history *h* (index), just in case the action performed by α at that index guarantees the truth of *A*. The action might result in a variety of possible outcomes, but the statement *A* must be true in each of them. I will follow Kapitan (2011: 131) and abbreviate the agent's ability to see to it that a situation *p* obtains as 'Ap,' their inability to see to it that *p* as '~Ap' and their inability to prevent *p* as '~A~p.'[5]

We can now spell out premise (4) of CA_1 in terms of the ability operator:

$$[A\sim P \,\&\, \Box\,((H \,\&\, L) \rightarrow P)] \rightarrow A\sim (H \,\&\, L)$$

This premise can be viewed as an instance of a more general logical principle for practical modalities (*Master*)[6]:

$$A\sim q, \Box\,(p \rightarrow q \vdash A\sim p$$

Recall that 'Np' says that *p* is true and that no one has had any choice about *p*. It seems natural to interpret having no choice as the inability to prevent *p*, so we get '*p* & ~A~*p*' as a plausible translation for 'Np'. There are difficulties with directly implementing this translation scheme (see *Appendix*), and different understandings of "ability" will resonate importantly in our further discussion. Still, this translation is consistent with Lewis's understanding of ability (but see discussion below). The most important result is a connection between the transfer principles for practical modalities and power necessity for our present purpose. Given the translation scheme and Lewis's understanding of ability, *Master* implies Beta 2. This is easy to see (details in the *Appendix*): If *p* entails *q*, and the ability to render *q* false implies the ability to render *p* false (*Master*), then the *in*ability to render *p* false implies

[5] Strictly speaking, we are talking about S's ability at time t, '$A_{s,t}$p', but skipping the indices will do no harm in my discussion.

[6] I borrow the name from Kapitan (2002), who discusses a variety of logical principles for practical modalities, based on "Whatever is a consequence of a possibility is itself possible," ascribed to Diodorus Chronus.

the *in*ability to render *q* false. It is then a short step to infer that the unavoidability (power necessity) of *p* implies the unavoidability of *q*. So, when *p* is unavoidable and *p* entails *q*, then *q* is also unavoidable, which is just Beta 2:

$$A\sim q, \Box(p \rightarrow q) \vdash A\sim p$$

implies[7] the validity of Beta 2 (closure of power necessity under entailment):

$$Np, \Box(p \rightarrow q) \vdash Nq$$

Let us return to Lewis and his assessment of CA_1. Lewis reformulates the argument for incompatibilism as a *reductio*: if we attribute ordinary abilities to agents in a deterministic universe (the judge J could have raised his hand), we are forced to credit them with a magical past or law-changing abilities, as well. This result is absurd, so, per *reductio*, if determinism is true, nobody has the ordinary ability to act otherwise. In his reply, Lewis launches a *distinguo* between two senses of ability (1986: 297):

> I could have rendered a proposition false in the *weak* sense iff I was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act).
> I could have rendered a proposition false in the *strong* sense iff I was able to do something such that, if I did it, the proposition would have been falsified either by my act itself or by some event caused by my act.

How does weak/strong distinction work as an antidote to CA_1? Here is a longer passage, emphasizing two points that are crucial for my reading:

> I did not raise my hand; suppose for *reductio* that I could have raised my hand, although determinism is true. Then it follows, ***given four premises that I cannot question***, that I could have rendered false the conjunction *HL* of a certain historical proposition *H* about the state of the world before my birth and a certain law proposition *L*. If so, then I could have rendered *L* false. (Premise 5.) But I could not have rendered *L* false. (Premise 6.) This refutes our supposition.
> To this, I reply that Premise 5 and Premise 6 are not both true. Which one is true depends on what van Inwagen means by "could have rendered false" (Lewis, 1986: 296). (…)
> If I could have raised my hand despite the fact that determinism is true and I did not raise it, then indeed ***it is true both in the weak sense and in the strong sense that I could have rendered false the conjunction HL of history and law***. But I could have rendered false the law proposition *L* in the weak sense, though I could not have rendered *L* false in the strong sense. So, if we take the weak sense throughout the argument, then I deny Premise 6. If instead we take the strong sense, then I deny Premise 5 (Lewis 1986: 297).

Lewis treats premise (4) as uncontroversial, but this premise is an instance of *Master*, so we can ascribe to him the acceptance of Beta 2 (implied by *Master*). Nevertheless, *reductio* fails because there is no

[7] Kapitan (2011: 134) argues for a general claim: the closure of power necessity and ability closure are equivalent. I, however, prefer a more modest claim.

uniform reading of "could have rendered false" to make all the premises true. Surprisingly, premise (6) is false according to the weak reading of ability: had I raised my (actually unraised) hand, a law would have been broken beforehand and not by my act itself or by some event caused by my act. Even more surprisingly, the implication (5) is false according to the strong reading of ability. The act of raising my (unraised) hand would directly, *by itself*, falsify any sufficiently inclusive conjunction of history and law. So, the antecedent of (5) is true.

In contrast, the consequence of (5) is false: the act of raising my (unraised) hand would not falsify laws of nature in the strong sense. It would be preceded by "the divergence miracle" that would falsify a law (Lewis 1986: 297). The laws of nature, just prior to my acting, would be slightly different from the way they are. The alternative action would take place in a different possible world: it would represent a miracle *relative* to the laws of our world but remain lawful in the possible world where I act otherwise (cf. Rummens, 2019).

4. *The main compatibilist response*

The weak/strong distinction is usually interpreted as the core of *the main compatibilist response* (or 'MCR') to CA. The notion of ability is explicated in terms of a conditional, where there are two general readings, stronger and weaker. S is *broadly* able at t to see to it that p iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) were S to do X, then p (Kapitan 2011: 135). No connection is specified between the course of action X and p. Another option is a narrower, causal reading: S is *causally* able at t to see to it that p iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) S's doing X would make it the case that p. According to this *active* reading, the agent brings about or causes p to obtain; p is the case because of what the agent does; her action leads to or results in p. Causal ability implies broad ability but not vice versa (for instance, S is broadly able to see to it that a certain tautology is true, but the tautology is not true *because* of her actions).

The notion of weak or broad ability opens the space for the denial of fixity of laws (or the past): the agent is (broadly) able to act otherwise without having the *causal* ability to break a law of nature. A free predetermined agent could have done something such that, had they done it, there would have been a difference in either law or history. "But that is not to say that the person could have brought about these conditions" (Lehrer 1980: 199). Lewis's distinction weak/strong is then interpreted along the lines of broad/causal and applied to the third, *modal* version of CA. The critical move is to accept that the agent is broadly able to act in a way that would falsify the laws of nature or the past but deny them any stronger "causal" powers. The agent could not have *brought about* the relevant differences, so premise (6) in the third version of CA is false (and so is premise (6) in the first version). Alternatively, one

might question premise (4) of CA_3, if one makes a difference between an agent who has the ability to act in such a way that they *alter* the past, as opposed to an agent who has (broad) ability to act in such a way such that, if they did so act, the past would have been different.

With two notions of ability, we now have two options for the notion of power necessity in CA_3, so let us introduce '$N_w p$' (a dual of weak ability) for, roughly, "*p* is true and no act of any person could in *any* way (weakly or strongly) falsify *p*", and '$N_s p$' (a dual of strong ability) for "*p* is true and no act of any person could strongly falsify *p*." According to MCR, premise (6) of CA_3 (and CA_1) is false for '$N_w p$', but the transfer principle (Beta) fails if we adopt the causal (strong) reading '$N_s p$':

> ... if we adopt the broad sense of ability, then, although the argument is valid, at least one of its premises is false, whereas if we adopt the causal sense of ability, then, although the premises are correct, the argument is invalid because the relevant closure principle fails. As the most prominent exponent of this line of reasoning, David Lewis, concluded, there is no one consistent reading of the consequence argument for incompatibilism critical modality that would render the Consequence Argument sound. (Kapitan 2011: 138)

Van Inwagen (2008b: 455) agrees:

> The philosopher David Lewis has contended … that our technical term 'untouchable' is ambiguous, and that if the word is understood in one of its possible senses, the Conditional Rule [i.e. *Beta*] is invalid, and, if it is understood in the other of its possible senses, *L* is not an untouchable truth.

Why is Beta not valid? Premises (4) and (6) of CA_3 are true in the *strong* sense: the laws of nature and events in the distant past are "not up to us"; they are beyond our causal control. Not so for our actions, which are up to us in the sense of being brought about by our desires, abilities, character, and beliefs (cf. Slote 1982; Mele 2006: 138). So, the conclusion is false: as free predetermined agents, we are still *causally* able to act otherwise. However, premises (4) and (6) of CA_3 are false in the weak sense of ability. As free predetermined agents, we do not possess the causal ("strong") ability with respect to the past and the laws of nature; instead, either the past or the laws of nature would have been different (would have to have been different) for our free action to take place.

So, CA is a philosophical failure according to MCR: either not valid (Beta fails) or not sound. This may be the central line of MCR, but there are immediate problems with the projection of Lewis's weak/strong onto broad/causal. First, it is difficult to directly extract the "not valid or not sound" verdict on Beta and CA_3 from Lewis (1981) on CA_1, since this version is not based on Beta. Moreover, even *weak* is defined in terms of counterfactual sufficiency (cf. Huemer 2000: 529): S can render *p* false iff S can perform some action, such that were he to do so, it would not be the case that *p*. Had I acted otherwise, a certain result would be *reliably* correlated with my alternative action, directly (strong) or indirectly (weak). Nevertheless, we cannot "ensure" the out-

come of a chancy process, so the McKay—Johnson counter-example to Agglomeration works for both, for '$N_w p$' and '$N_s p$'.

Lewis's verdict on premises (4) and (5) of CA_1 reveals his assessment of the relevant transfer principles. Slote has noted that premise (5) "If J could have rendered the conjunction of *H* and *L* false, then J could have rendered *L* false" corresponds to the principle of Agglomeration for 'N'. Given the law of contraposition and the assumption of the fixity of the past, (5) is tantamount to "if one can't falsify a law of nature *L* and can't falsify a proposition about the past *H*, one can't render false the conjunction of *L* and *H*", which is just an Agglomeration assumption for the necessity expressed by "can't render false" (Slote, 1982: 10, fn 7). With Lewis, we now have two options for the relevant notions of necessity, so let me start with Agglomeration for the dual of strong ability:

$$N_s H \,\&\, N_s L \vdash N_s (H \,\&\, L)$$

Lewis clearly *denies* the validity of this pattern. The premises are true: "… if I had raised my hand, the intrinsic state of the world long ago would have been no different", neither do we possess *strong* ability with respect to the laws of nature. However, the conclusion is false—I could have rendered false the conjunction of history and law 'H & L' in *both* senses. A denial of Agglomeration implies a denial of Beta for 'N_s' (Slote indicated this already). Nevertheless, Lewis accepts premise (4) of CA_1 in both senses (the premise he cannot question), so he accepts Beta 2 according to the translations proposed. This is consistent, since Beta 2 and Agglomeration (plus Alpha) together entail the validity of Beta, so a denial of Agglomeration will block Beta (see Appendix). One can have closure of an operator under *entailment* without the corresponding closure under *implication*.

How about weak ability and the corresponding transfer principles? Replacing practical modalities with the appropriate notion of power necessity, contraposing (5) of CA_1 and using propositional logic, we get:

$$N_w L \rightarrow N_w (H \,\&\, L)$$

Can we add '$N_w H$' as a separate conjunct in the antecedent? Is the past completely isolated from our actions in both senses of ability? Not the entire past: under determinism, the counterfactual worlds in which I act otherwise must involve a difference (a divergence miracle) in the past immediately before my action. Still, I think that Lewis would agree with the *absolute* impossibility of rendering false a true proposition about the *remote* past, so we can insert this proposition as a second conjunct in the antecedent to get:

$$(N_w L \,\&\, N_w H) \rightarrow N_w (H \,\&\, L)$$

Agglomeration for 'N_w' looks acceptable in *this special* case at least. The conditional is at least *plausible*, given the content of the propositions involved. Fischer, in defense of his conditional version of the consequence argument, pointed out that sometimes the argument is not

formally valid, but it is nonetheless reasonable to accept its conclusion given the *content* of its premises (Fischer, 1994: 228, fn 43). Since Beta 2 (accepted by Lewis in both senses) and Agglomeration together entail the validity of Beta, we might accept Beta for 'N_w' as at least "materially" valid. Of course, in this case "N_wL" is false according to Lewis, and the modal argument (CA_3) based on the notion of 'N_w', a dual of weak ability, is not sound, precisely as diagnosed by the MCR.

5. *Problems*

Have we now confirmed the canonical interpretation of Lewis and established a correspondence between Lewis's weak and strong and the notions used by the main compatibilistic reply, broad and causal? Cracks begin to appear when we reconsider premise (4) of CA_1:

$$[A\neg P \,\&\, \Box ((H \,\&\, L) \rightarrow P)] \rightarrow A\neg(H \,\&\, L)$$

According to Lewis, this premise is valid whichever notion of (in)ability we consider. And he readily accepts the result (Lewis, 1986: 297): "My act of raising my hand would (by *itself*) falsify any sufficiently inclusive conjunction of history and law." This is not easy to accept, even less so when we interpret strong ability as *causal* ability, the ability to bring something about or make changes. This seems to be suggested by Lewis's examples of *strong* ability. For instance (Lewis 1986: 294):

> Therefore, I am able to break a window, a promise, or a law only if I am able to do something such that, if I did it, my act either would cause or would be a window-, promise-, or law-breaking event.

He then goes on to deny that a free predetermined agent possesses such an ability with respect to the laws of nature (the act of raising his hand would not itself be or cause a law-breaking event, rather a law would be broken beforehand if he did it). But, according to (4), a free predetermined agent possesses such an ability with respect to the *conjunction* of history and law. He could have rendered the conjunction false in the strong sense, so he was able to do something such that, if he did it, the conjunction would have been falsified either by his act itself or by some event caused by his act. His act itself would cause or be a conjunction-breaking event.

Let me use ODD for the claim that a free, predetermined agent possesses a *strong* ability to render false the *conjunction* of the remote past and the laws of nature. ODD, in general, does not look very attractive. The incompatibilists declare ODD absurd and view this result as a refutation of Lewis's strategy. *Strong* is supposed to be the natural understanding of ability in the premises of CA for incompatibilists — "whereby they mean the ability to bring something about or cause something to be by virtue of one's actions" (Kane 1996: 50). With premise (4) impeccable, the rejection of ODD and a simple *modus tollens* re-establish the thesis of incompatibilism: given determinism, it is not the case that one can otherwise.

The MCR agrees with the negative verdict on ODD and proposes a different strategy. Once the distinction weak/strong is available, why not apply it to premise (4) of CA_1?

(4') If J could have rendered *P* false, and if the conjunction of *H* and *L* entails *P*, then he was able to do something such that, if he did it, the conjunction would have been false.[8]

While J is able to render *P* false in the strong sense, J is not able to render this conjunction false in the strong (causal) sense; he cannot bring about the difference. According to Perry (2004: 247), when we interpret "can render false" in the strong, causal sense (the proposition is false *because* of my action), then (to put *P* instead of his *Q*): "It simply does not follow from the fact that J will render *P* false that he renders false every proposition that entails *P*. What does follow is that there is no true proposition that entails *P*."

Kapitan (2011: 137–138) suggests the same move when he considers the case of a soldier (Roni) who is able to disobey his commander's instructions, but nevertheless decides to act as commanded ('P'). He claims that Roni is *broadly* able to bring about '~(H & L)', but not causally:

> If /.../ ability is construed in the causal sense, then the premises of the Consequence Argument are true; /.../ there is nothing Roni is able to do that would make it the case that either *H* or *L* does not obtain, and so, Roni is causally unable to see to it that ~(H & L). But, in that case, because Roni is able to disobey orders and by so doing would make it the case that ~P, he is causally able to see to it that ~P. Consequently, the transfer rules for causal unavoidability are invalid. Likewise, because Roni is causally able to see to it that ~P and ~(H & L) is a consequence of ~P, then the transfer rules for causal ability are invalid.

As the last sentence indicates, Kapitan clearly thinks that Beta 2 (closure under *entailment*) fails if we adopt a causal (strong) reading of (in) ability. Roni is causally *unable* to see to it that ~(H & L). The defenders of MCR in general agree that the transfer principle behind premise (4) is false. They propose to deny premise (4) with the help of *broad* and thus avoid ODD. A free predetermined agent can act otherwise in a strong ("causal") way but can falsify the conjunction in the *weak* (broad) sense only. Her action at time t will not make the conjunction untrue, it will not render it untrue. Instead, the conjunction will "be untrue" because the earlier events at time t–1 made it false (Perry, 2004: 249, see also Rummens, 2019). From the logical point of view, MCR declares both critical principles, Beta (closure under *implication*) and Beta 2 (closure under *entailment*), fallacious. But, in denying the latter, they disagree with Lewis.

One might propose a reading that makes ODD more acceptable to the wider compatibilist camp. Being able in a *strong* sense to falsify the conjunction of history and law is simply to be able to cause an event

[8] To use the "simply" weak sense of "can render false", preferred by Lewis instead of "weak" in his letter to Horgan (Lewis et al. 2020: 119).

E such that the conjunction is false in every possible world in which E occurs. And the claim that deterministic agents possess this ability is something that any compatibilist who believes in deterministic agents' ability to sometimes act otherwise, needs to accept. But this interpretation blurs the very distinction between weak and strong and it goes against the textual evidence. Consider Lewis again:

> Let us say that I could have rendered a proposition false in the weak sense iff I was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act). And let us say that I could have rendered a proposition false in the strong sense iff I was able to do something such that, if I did it, the proposition would have been falsified either by my act itself or by some event caused by my act (1986: 297).

This passage strongly suggests that, understood in the strong sense, my act itself (or something caused by my act) would *render* the conjunction of history and law untrue. The proposed reading of *strong*: "being able in a strong sense to falsify P is to be able to cause an event E such that P is false in every possible world in which E occurs" is suspiciously close to *weak* (just being able to do something such that, if S did it, P would have been falsified). Is it possible to defend a view that according to which *strong* ability with respect to a certain true proposition can be interpreted so that: (i) the proposition would have been falsified either by my act itself or by some event caused by my act; (ii) the proposition would not have been thereby *rendered* false either by my act itself or by some event caused by my act; (iii) strong does not collapse into weak; (iv) and (i) still ascribes to the agent more than just the ability to do something such that, if she did it, the proposition would have been (weakly) falsified? Difficult to say, it is not easy to resolve all the tensions generated by Lewis's distinctions. Let me try to address these issues from the general point of the ability to render a proposition false.

6. *Loaded and neutral*

Recall the general notion of ability to render a proposition false (Kapitan 2011: 135): S is able at t to see to it that $\sim p$ iff there is a course of action X such that at t (i) S is able at t to do X, and (ii) were S to do X, then $\sim p$. One way of understanding the second condition is *broad*; in most general terms, S's action would be inconsistent with the truth of p, or, from S's doing X, it may be inferred that p is false. Even logical necessities and past truths are such that one is able at t to see to it that they obtain in this sense: they are (now) true and whatever S does they (still) remain true. Let 'P' stand for my actual refraining to raise my hand, although I am perfectly able to do so. Let 'Q' be a contradiction, and 'R' an arbitrary false proposition about the past (say: "On July 20, 1969 Edwin Aldrin became the first human to step on the Moon"). The closure of the ability operator 'A' under entailment licenses the inferences: (i) "A~P, Q entails P ⊢ A~Q" and (ii) "A~P, (P & R) entails P ⊢ A~ (P & R)."

According to (i), I am able to render a contradiction false (and so a tautology true). For any course of action that I can take at *t*, it would (still) be the case that not–Q. According to (ii), I am able to render false a conjunction which was *already* false independently of my actions (Neil Armstrong was the first human to step on the Moon). For any course of action *P* that I can now take (after, say 2020), it would (still) be the case that not–R; thus, the conjunction 'P & R' would remain false.

This is the "neutral" interpretation of ability sometimes championed by modal logicians. However, this neutral, purely logical (modal) notion comes with a price: it may sound unnatural and contra-intuitive. Even some logically minded people will protest, thus Kenny (1976: 214):

> The President of the United States has the power to destroy Moscow, i.e., to bring it about that Moscow is destroyed; but he does not have the power to bring it about that either Moscow is destroyed or Moscow is not destroyed. /. . ./ The power to bring it about that either *p* or not *p* is one which philosophers, with the exception of Descartes, have denied even to God.

According to Kenny, nobody has the power to bring it about that a tautology is true, which seems equivalent to the ability to render a contradiction false. In a similar way, Perry (2004: 247) objects to (ii); nobody can render false a past falsity (the gist of his objection to premise (4) of CA_1). Contradictions and past falsities are not *made* false by the agent; they are just false independently of their (present) intentions and interventions. In a similar vein, Schneider (2004: 418) defines van Inwagen's "x can render *p* false" as "x can do something such that if x did it, *because* of that *p* would be false," treating "because" as expressing a primitive "explanatory relation." Premise (4) of CA_1 fails on this reading. Nothing that the judge can do is such that *because* of his doing so, the conjunction of *L* and *H* would be false; the conjunction is false for different reasons (although Schnieder does not say what those reasons are (Schnieder 2004: 423).

I will call the notion of ability suggested by Kenny, Perry and Schnieder "loaded" ability, roughly associated with the "active", causal power to bring something about, realize, change something, or make something false. The active component can be further strengthened with explanatory and/or agential components, perhaps cognitive and intentional. Active "causal" contribution is enough to block closure under entailment, but when we include additional constraints imposed by the general "metaphysics of agency", the invalidity of closure under entailment becomes even more apparent. To take an example by Kapitan (1996: 423): I am able to drink a cup of coffee. By so doing, I would bring about complex molecular changes in my brain, but I am ignorant of physiology, so these changes are not a reliable result of my drinking coffee. I am not able to bring them about, so closure fails.

To summarize—suppose we take "loaded" ability as an umbrella term for the notion of ability involving active, causal powers (by doing X, the agent causally contributes to the obtaining of *p* or makes

it the case that *p*). We might impose further constraints (reliability, intentionality, knowledge, skill, etc.) and thereby generate layers of practical modalities (cf. Kapitan 1996). According to MCR, causal ability in CA should be interpreted as loaded, but closure under entailment *fails* for loaded, so premise (4) of CA_1 is false. Broad ability, however, is neutral from the logical point of view. One might perhaps introduce different levels of neutrality. On a higher level, there is the *general* rule of consequence (the generalization of rule Beta or closure under implication); on a lower level, there is *simple* consequence (closure under entailment—*Master* or Beta 2, see Appendix).

Lewis denies the higher level of neutrality for *strong* ability, but he accepts the lower. This is a signal that Lewis's strong/weak distinction does not map unproblematically onto the causal/broad distinction of MCR. There are two options: (1) strong ability is not necessarily causal, or (2) the logic of strong ability is not the logic of loaded ability. One can test the first option: my act, such that I was predetermined not to perform, would itself be an event directly and reliably falsifying the broad past without thereby *making* any changes. Perhaps one "can directly bring about circumstances in which *p* is false", but "one cannot bring about the falsity of *p*" (Brown 1988: 24, fn 13). This "direct" but non-causal ability will still be stronger than "indirect" weak ability. This interpretation might further be supported by Lewis's views on causation and counterfactual dependency. By his account, causal dependence requires the appropriate patterns of counterfactual dependence among two *particular* events: *C* and *E* (*E* causally depends on *C* if and only if: if *C* had occurred, *E* would have occurred; and if *C* had not occurred, *E* would not have occurred). J can then cause the conjunction of *H* and *L* to be different only if there is some *particular* event that counterfactually depends on J's raising his hand. Nevertheless, differences in the conjunction of *H* and *L*, correlated with J's counterfactual raising of his hand, might have been realized in a variety of ways.

Still, we might be inventing distinctions where there are none. Take Lewis's letter to Lehrer where he says that they agree in distinguishing the following (Lewis et al. 2020: 94):

(1) I could have done so-and-so, and if I had, the laws (or history) would have been different,
(2) I could have brought it about that the laws (or history) would have been different.

For both for Lewis and Lehrer, (1) is true and (2) is false. According to Lewis, the second statement adds some false extra content that has to do with *causation* from my action to the law-breaking miracle (or the alternative past). The extra content, apparently characterizing *strong* ability, is causal. Moreover, Lewis's examples (Lewis 1981) and comments suggest a causal reading for strong (ability to break the window, ability to break a promise). Although some tensions will remain, the second option is more plausible: strong is causal, but the logic of strong

is not the logic of loaded. I conjecture that Lewis follows the practice of (some) logicians who accept the closure of ability under entailment with all the corollaries alluded to earlier: an agent is able to bring it about the logically unavoidable. At least in pre-*stit* logical practice, it was accepted as technically convenient and unproblematic to construe the *ability* operators as meaningfully applicable even to necessarily true sentences (Brown, 1988: 24). It is not impossible to defend such a practice; consider the comment by Chellas on Belnap, whose notion of "the agent sees to it that …" is *loaded* (not closed under logical consequence):

> Can it ever be the case that someone sees to it that something logically true is so? I believe the answer is yes. When one sees to something, one sees to anything that logically follows, including the easiest such things, such as those represented by a tautology. One should think of seeing to it that (e.g.) 0 = 0 as a sort of trivial pursuit, attendant upon seeing to anything at all. (Chellas 1992: 508)

Certain remarks made by Lewis in a different context suggest that he would accept this line of reasoning. For instance, when discussing the logic of relevance and its motivation (supposedly problematic inferences *ex falso quod libet* and *verum ex quod libet*), he explains tautology as vacuously about any subject matter and thus, one might assume, also trivially implied by and a *result* of any action the agent is able to perform (Lewis 1988: 115). There is a price to be paid as a counterexample to loaded show; it is difficult to see how logical necessities are "made true" by the agent. Nevertheless, there are also benefits—above all, generality and liability to logical investigations. Extensionality is no longer a prerequisite; van Inwagen, when faced with the properties of *loaded* ability as an objection to his official definition of "can render false", claimed that his definition at least had the consequence that "S can render *p* false" was a purely extensional context (van Inwagen 1983: 231, fn 9). In the contemporary framework of modal logic, extensionality is not sacrosanct. Even "normality" can be sacrificed (a generalization of Beta or closure under implication is required for "normal" modal logic). However, an operator that is not closed under logical consequence satisfies virtually *no* logical laws.

7. *Odd?*

I have argued that Lewis's way with closure and premise 4 of CA_1 is the way of standard (pre-*stit*) logicians on ability. Admittedly, the defenders of loaded have made some persuasive points; logical laws are not sacred, and the tools of logics have evolved (in the framework of contemporary *stit* approaches, the ability is not closed under logical consequence). Lewis's strong ability is causal but nevertheless neutral from the logical point of view. He might or might not agree with some of the proposed tweaks, but he never questioned *Master* (see a discussion about *Hasker* in the *Appendix*). In his letter to van Inwagen, he acknowledges that a compatibilist might worry that by denying prem-

ise (5), "the compatibilist has a problem explaining why he wouldn't be reversedly causing a divergence miracle" (Lewis et al. 2020: 91). By denying "strong", he solves the problem: J could have rendered antecedent, the conjunction of *H* and *L* false (in the strong, causal sense), but J could not have rendered the consequent *L* false in the strong sense. However, the compatibilist does not face the problem of explaining why he wouldn't be reversedly *causing* a difference in the conjunction of *H* and *L*. Why not?

Setting aside Lewis's somehow idiosyncratic views on causation and counterfactuals, there is a way to explain and even defend this move. Some new textual evidence comes from his recently published correspondence and previously unpublished manuscripts. Lewis in his letter to Thomas Nagel, remarks that, "I could have raised my hand" is really inconsistent with history and the law, but compatibilists should not bother, since "It's scarcely even a consequence of compatibilism — just a restatement of it" (Lewis et al. 2020: 94). I propose to extend this type of reasoning to strong ability with respect to the conjunction of *H* and *L* in CA_1. Strong ability to render false this conjunction is scarcely a consequence of the ability to do otherwise, given the truth of determinism: it is merely a restatement of it.

Horgan hinted at this line when he introduced a variation on *strong*. Under this interpretation, J can raise his hand at T and this act *itself* is now being counted as a *H* (or *L*)-falsifying event. But, according to Horgan:

> It would be outrageous, of course, to claim that J can causally influence events in the remote past. But we are saying nothing so offensive when we assert that J can render *H* false in the strong and broad sense. On the contrary, essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*. (1985: 348)

In the opposite camp, Kane, an incompatibilist, noted the same fact and, of course, duly objected to *that*. In his version of CA_1, the crucial premise (4) is renumbered as premise (3), while premise (6) reads as a denial of ODD: "It is not possible that an agent *a* at *t* can (has the power or ability to) render false the conjunction of distant past and law." In his discussion of the compatibilist opposition to *Master* (premise (3) in his reading), he writes (Kane 1996: 51):

> Premise 3 says 'if (i) you could have done other than move your hand, and (ii) your hand's moving was determined by laws and the past, then you could have rendered false a law of nature or the past' (which is the denial of 6). But to us incompatibilists, assuming at the outset that (i) 'you could have done other than move your hand' under the assumption that (ii) 'your hand's moving was determined' begs the whole question. For it means that the power to do other than move your hand that is assumed in the antecedent of 3 must be a compatibilist power—which means in turn that the argument against 3 succeeds only if one assumes at the outset a compatibilist interpretation of the power to do otherwise.

Clearly, he is targeting the position of MCR on strong (and not Lewis)—if you reject strong ability to render false the conjunction of *H* and *L* you have to object to *Master* (closure under entailment), and he finds *any* objection to Master question-begging. The very idea of a counterexample to Master is problematic; for the conditional (his premise 3, premise (4) of the original CA_1) to be false, the antecedent must be true and the consequent false. However, in the antecedent, we begin with the *assumption* that J can do something that he is causally determined not to do, which is already question-begging according to Kane. Lewis, as we saw, does not reject strong ability to render false the *conjunction* of history and law and finds no problem with *Master*. Nevertheless, he is also committed to the compatibilist interpretation of the power to do otherwise, according to Kane (1996: 223 fn 12):

> We might easily overlook this fact because the argument assumes Lewis's strong sense of "can do otherwise," which is not so obviously a compatibilist notion as is his weak sense. But, as Lewis is well aware, his strong (or causal) sense of "can do otherwise" is not necessarily incompatibilist. It can also be given a compatibilist analysis.

Let us take stock of the debate and various positions. ASSUMPTION will be the claim that a free predetermined agent can act otherwise. Given determinism and suitable transfer principles, (*Master*) ASSUMPTION implies ODD, the ability to render false the conjunction of distant past and laws of nature. Let us call WONDER the possession of *strong* ability to render false *one* of the conjuncts of ODD. Let us call PECULIAR the possession of *weak* ability to render false the conjunction of distant past and laws of nature. Finally, let us call MIRACLE the possession of *weak* ability to render false *one* of the conjuncts of ODD (laws of nature, according to Lewis's brand of local miracle compatibilism).

Incompatibilists take *Master* as beyond any doubt, but ODD is already incredible enough (Kane 1996: 50–52), so the ASSUMPTION must go. Some compatibilists (the MCR group) agree that ODD is incredible and WONDER must be avoided at all costs. Therefore, *Master* must be denied, but PECULIAR is something a compatibilist can live with. Lewis is unique: he sides with the incompatibilists in accepting *Master*, embraces ODD as a result but denies that WONDER follows from ODD. According to Lewis, ODD implies MIRACLE (with respect to the laws of nature) only, since Agglomeration fails (strong ability to render false a conjunction does not distribute over the conjuncts). Moreover, according to Lewis, both MIRACLE and ODD are something a compatibilist can live with.

This is just a skeleton; there are more nuanced positions (cf. van Inwagen: 2004, 349), and I did not discuss how to live with MIRACLE in this paper. I mainly agree with Vihvelin (2013: 164) that the miraculousness is just a result of our counterfactual speculation about what would have been the case, beforehand, if anything in a determin-

istic world had happened in any way other than the way it actually happened. If counterfactuals in a deterministic world make any sense, then so does MIRACLE. However, my main concern is how to live with ODD-ness according to Lewis. I will suggest that the acceptance of ODD can be interpreted as a corollary of a *Moorean* fact that we have free will. The compatibilists who are embarrassed by ODD underestimate the *dialectical* force of "strong."

8. *Moorean facts*

Many of Lewis's philosophical investigations start from not negotiable pieces of our ordinary picture of the world (cf. Nolan 2015). This is also true of his compatibilism (Lewis 2020: 241):

> It's a Moorean fact that we often have a choice what to do. But whether determinism holds is an unsettled question. So, having a free choice is epistemically compatible with determinism. And with indeterminism. So, it's compatible simpliciter.

We begin with a Moorean fact that we are able es to act otherwise, and an arbitrary person on the street, ignorant of philosophical technicalities, will very likely understand their ordinary ability to raise their unraised arm in causal terms. But then being able (in the ordinary, Moorean sense) to act otherwise if determinism is true amounts to being able to falsify the conjunction of history and laws of nature that determines the actual action. Consider an analogy. According to scientific essentialism, if gold exists, then it has—necessarily—atomic number 79. Possessing this property is a metaphysically necessary condition for being the kind of thing designated by the natural kind term "gold". Suppose, somehow anachronistically, that two alchemists (A1 and A2) are discussing the following statement made by A1:

PS If scientific essentialism is true and by using a philosophers' stone you are able to turn this piece of stuff which is iron into gold, then you are able to change the atomic number of this stuff from 26 to 79.

A2 might object that it is impossible to change the atomic number of an element. But A1 will reply, that, given scientific essentialism, that is just what it *means* to turn iron into gold. By turning iron into gold, you are changing the atomic number of this stuff. Of course, A2 might object to the doctrine of scientific essentialism on its own, but this is a separate issue and not an objection to PS. And A2 might think that the truth of scientific essentialism (somehow?) precludes the ability to effectively use the philosopher's stone. But then he would still agree with PS. Given scientific essentialism, the act of turning this piece of stuff, which is iron, into gold would be truly describable as changing the atomic number of this stuff from 26 to 79. And given determinism, the act of raising my actually unraised hand would be truly describable as rendering false the conjunction of distant past and the laws of na-

ture, to use the terminology of Fischer in his version of the consequence argument (Fischer 1994: 28).

Let us call sentences of the type "X would be truly describable as Y", where the truth of the proposition expressed is grounded in the entailment relation between the contents of X and Y, "quasi-analytic." It is then quasi-analytically true that being able to act otherwise given that determinism is true is just being able to falsify the conjunction of history and the laws of nature which entails the actual action. *Master* is then almost trivial—exactly as van Inwagen (1983: 72) claimed it to be. And to use the phraseology of Chellas on Belnap, one should think of seeing to it that the conjunction is false as a sort of pursuit, attendant upon seeing it to do what one was predetermined not to do. Looking for counter-examples is then almost futile; this will disturb the MCR branch of compatibilism. This will also disturb those incompatibilists who use *Master* as their main principle for deriving *consequences* inimical to compatibilism (Finch and Warfield 1998).

According to this line of reasoning, ODD is a quasi-analytic restatement of compatibilism. This is just what it means to be predetermined but still be able act freely. When ascribing to J strong ability to render false the conjunction of history and law, then essentially all we are saying is that J can do something that he is causally determined not to do; and it is no surprise to learn that the compatibilist is committed to *that*. Kane is perhaps aware of this fact when he argues that a denial of *Master* is already question-begging: not because some magical powers are assigned to the agents but because of the very ASSUMPTION (compatibilism). Of course, to reject an account of the ability to do otherwise on the sole ground that it is compatible with determinism begs the question just as well.

9. *The limits of logic*

I have explored some rather intricate details of CA and practical modalities involved (cf. also the *Appendix*). I think that logical investigations help us to systematically extract the consequences of our initial commitments. But the philosophical role of an argument and its validity is overrated by Murdoch (the initial quotation). I think that Harman is ultimately right: inference is always "inference to the best overall view." The acceptability of the premises and our starting points are an important ingredient of our overall view. Well, Lewis is unusually clear about *his* starting points. For instance:

> Apart from that [believing in the existence of concrete alternative possible worlds] I am philosophically conservative: I think philosophy cannot credibly challenge either the positive convictions of common sense or the established theses of natural sciences and mathematics. (Pyke 1995)

It is a firm conviction of common sense, a Moorean fact, that we make free choices. So compatibilism and libertarianism (free will being in-

compatible with determinism) are the only philosophical options left. I think that the respect for science is then decisive for Lewis: "we know better that we are sometimes free than that we ever escape predetermination; wherefore it may be for all we know that we are free but predetermined" (Lewis 1993: 155). For all we know, the thesis of determinism might be true, but it is up to science to (dis)confirm this thesis. A weak/strong distinction then still allows for two compatibilist options: to accept or to deny closure under entailment for strong ability.

At this point it might be useful to compare Lewis's position with respect to strong ability to act otherwise (the acceptance of closure) with a familiar Moorean position on knowledge and radical skepticism. The logical parallel of *Master* is the claim that knowledge is closed under entailment. Roughly, "If S knows that *P*, and *P* entails *Q*, then S knows that *Q*." This must be refined[9], but it will suffice for our comparison. Knowledge closure figures prominently in a much-discussed argument for skepticism. Our ordinary perceptual knowledge logically excludes radical skeptical scenarios where these scenarios are subjectively indistinguishable from a paradigm case of perception, but where one is in fact massively deceived. The radical skeptic then uses closure in the following familiar argument:

> I know I am standing; my knowing that I am standing entails that I am not dreaming; but I do not know that I am not dreaming, so I do not know that I am standing.

Moore is famous for agreeing with the sceptic in *accepting* the relevant closure principle but not giving up our ordinary knowledge; he thus argues:

> I know I am standing; my knowing that I am standing entails that I am not dreaming; therefore, I know that I am not dreaming.

Consider now the beginning of Lewis (1986: 291), paraphrased in terms of closure:

> I can raise my hand. Given the truth of determinism my doing so entails that I am able, in a strong sense, to render false the conjunction of *H* and *L*. Therefore I am able, in a strong sense, to render false the conjunction of *H* and *L*.

I know I am standing and the way in which I know is not the sort of way that is endangered by the possibility of a radical, skeptical scenario according to Moore. And I know I am free. The way in which I am (perhaps) determined not to do so is not the sort of way that counts as inability according to Lewis. A classical Moorean strategy allows one to meet the challenge of skepticism without having to deny the closure principle. Moreover, a parallel Moorean strategy allows one to meet the challenge of incompatibilism without having to deny the closure principle for (strong) ability.

[9] For instance: "If S knows that *p*, and S competently deduces from *p* that *q*, thereby forming a belief that *q* on this basis while retaining their knowledge that *p*, then S knows that *q*" (Pritchard 2016: 13).

Both a Moorean position on knowledge and a Moorean attitude on free choice have seemingly incredible consequences. Nevertheless, the cases are not quite analogous from the *argumentative* point of view. One of the crucial elements in the Moorean anti-skeptical strategy is a plausible explanation of our knowledge of the denials of skeptical hypotheses. Moore's comments are often found puzzling (Moore 1993: 169): "I have, no doubt, conclusive reasons for asserting that I am not now dreaming; I have conclusive evidence that I am awake: but that is a very different thing from being able to prove it." It is notoriously difficult to account for this evidence. (Neo)Mooreans in general endorse our common-sensical knowledge of the denial of the radical skeptical hypothesis, but few are willing to claim that the entailment "I am standing entails that I am not dreaming" is capable of *transferring* such knowledge.

Lewis never bothered with addressing the analogous problem of explaining our strong ability to render false the conjunction of history and law. I think there is an important difference between the two cases which might explain this "carelessness." Moore must provide some reasons for the claim that we know a radical skeptical hypothesis to be false. Lewis, on the other hand, must explain our strong ability with respect to the *conjunction* of *L* and *H*. According to my interpretation, ODD is quasi-analytic, merely a restatement of a compatibilist position with respect to ability to act otherwise. To use the phraseology of Chellas on Belnap, one should think of seeing to it that the conjunction is false as a sort of pursuit, attendant upon seeing it to do what one was predetermined not to do. There are no marvelous powers at issue, nothing is literally "transferred." Explaining the evidence for our (in)ability with respect to *each* of the conjuncts (A~H, A~L) and how those (in) abilities do or do not combine (Agglomeration) is where all the action is. From the purely evidential point of view, the reasons for thinking that the past is fixed for an agent are different from the reasons for thinking that the laws are fixed. The past (or distant past) is commonly understood as inevitable ("Everything that is past and true is necessary", according to Diodorus Cronus). However, the laws of nature might be Humean: as facts about regularities among events, they predict but do not constrain. In any case, van Inwagen is aware of the dialectics and carefully provides evidence for each premise of the consequence argument ('NH' and 'NL') separately. The weak/strong distinction is an attempt to refute this evidence.

And finally, to answer my original question: if compatibilism is the *accused*, whom should we praise—defense or prosecution with its CA? Well, who bears the burden of proof and who has the presumption of innocence? A Moorean stance combined with the respect for science takes compatibilism as the initial (not guilty) position. In this case, we should agree with Lewis that the prosecution with its consequence argument failed to prove the case. Still, at the end of the day, this is not a purely "logical" victory. Ramsey (initial quotation) might be right, logic just

displays the structure of your initial commitments, when you disagree with certain conclusions, you disagree with a certain set of starting points (thus you "beg the question"). But it shows this much: a certain set of starting points does not entail an absurd conclusion. Not really what Murdoch expected from a philosophical argument, but perhaps all that can be realistically expected as a *probative* role of logical tools in the case of the consequence argument.

Appendix

Lewis (1981) discusses the first version of the consequence argument where "can(not) render false", expressed here in terms of ability ('A'), is the crucial notion. Most contemporary discussions focus on the third, modal version, where the "no choice" ('N') necessity is central. In order to compare them, we must establish a translation scheme between the two notions and compare some central principles:

Alpha	$\Box p \vdash Np$
Agglomeration	$Np \,\&\, Nq \vdash N(p \,\&\, q)$
Beta	$Np, N(p \rightarrow q \vdash Nq$
Beta 2	$Np, \Box(p \rightarrow q) \vdash Nq$
Master	$A{\sim}q, \Box(p \rightarrow q \vdash A{\sim}p$

I rely, roughly, on Kapitan's proposal (2011: 131) and interpret "S can render *p* false" as: "S is able prevent that a situation *p* obtains." Having no choice about *p* or 'Np' is then naturally defined as "p & ~A~p" or "*p* is true and nobody is able to prevent *p* from being true." Caution is required, however. Tautologies are logically necessary truths, so it follows (via rule Alpha), that no one has, or ever had, any choice about whether a tautology is true. But let 'R' stand for the proposition that expresses the fact that J did not raise his hand at T and let 'Q' denote an arbitrary contradiction. We assume that J could have rendered *R* false, *Q* entails *R* (a contradiction entails anything), so, via *Master*, J could have rendered *Q* false?! This result follows even from van Inwagen's official definition of "S can render *p* false" (van Inwagen 1983: 68):

> It is within S's power to arrange or modify the concrete objects that constitute his environment in some way such that it is not possible in the broadly logical sense that he arrange or modify those objects in that way and the past have been exactly as it in fact was and *p* be true.

Well, whatever S does—it is impossible for S to arrange or modify their environment and a contradiction be true (cf. Schnieder 2008: 106). Let us take an arbitrary contradiction 'Q' and its negation '~Q' which is, assuming classical logic, a tautology. According to the translation schemes proposed *Master* licenses a strange result. No one has, or ever had, any choice about whether a tautology is true ('N~Q'), but one *can* prevent a contradiction from being true, which seems equivalent to being able to see to it that the *very* tautology in question ('A~Q') obtains. This looks strongly counterintuitive.

However, let me first note that 'Np' is defined as a *conjunction*: *p* is true *and* nobody can prevent *p* from being true. We get a better match between the "no choice" of CA_3 and "can(not) render false" of CA_1 when we include the first conjunct in the closure principle governing 'A' (*Hasker*[10] instead of *Master*):

$$A\sim q, \Box(p \rightarrow q), p \vdash A\sim p$$

The principle restricts the scope of one's ability to the ability to prevent a (contingently) true proposition from being true. Van Inwagen himself hinted at such a restriction when discussing certain odd consequences of his official definition of the ability to render a proposition false (S can render false an arbitrary falsity about the past). However, the first argument for the incompatibility of free will and determinism involves *true* propositions only (van Inwagen 1983: 68).

Different understandings of "ability" resonated importantly in our discussion of *Master*. The ability to see to it that a tautology obtains is strange when this ability is associated with "active", causal power to bring something about. But "loaded" (causal, active) modal notions satisfy virtually no logical laws, so there is not much to say about the principles governing them. I have argued that Lewis would accept the claim that one is *able* to render a contradiction false. The conjunction "N~Q & A~Q" is odd but not "abominable." When both locutions are translated in terms of neutral 'A' we get an asymmetry. One is unable to *prevent* a tautology (which is, of course, true) from being true, but one is *able* so to act that a contradiction is false.

Given these explanations, we can show that *Master* (and a revised *Hasker*) implies Beta 2. I will assume standard propositional logic (PL) and the fact that broad logical necessity is *alethic* (also compare Kapitan 1991: 335):

1. Np	hypothesis
2. $\Box (p \rightarrow q)$	hypothesis
3. $\neg A \neg p$ & p	1 definition of 'N'
4. $A\neg q \rightarrow A\neg p$	2 Master
5. $\neg A \neg p \rightarrow \neg A \neg q$	4 PL
6. $\neg A \neg p$	3 PL
7. $\neg A \neg q$	5, 6 PL
8. $p \rightarrow q$	2 modal logic
9. p	3 PL
10. q	9, 8 PL
11. $\neg A \neg q$ & q	10, 7 PL
12. Nq	11 definition of 'N'

[10] Hasker (1989: 112) defends the principle (PEP 5): If it is in S's power to bring it about that P, and "P" entails "Q" and "Q" is false, then it is in S's power to bring it about that Q. (PEP 5) is equivalent to *Hasker*.

What is the relation between various closure principles (Master, Beta, Beta 2)? Beta 2 and Agglomeration together entail the validity of Beta, just for the record:[11]

1. N p	Premise
2. N (p → q)	Premise
3. N [p & (p → q)]	1, 2 Agglomeration
4. □ {[p & (p → q)] → q}	Necessity of a logical truth
5. N q	3, 4 Beta 2

A counter-example to Beta is therefore a counter-example to the combination of Agglomeration and Beta 2. One might accept Beta 2 as valid but still deny Beta because Agglomeration fails for the relevant operator; this is Lewis's position, according to my interpretation. Lewis accepts premise (4) of CA_1, an instance of *Master*, and we have established that this principle implies Beta 2. He explicitly denies Agglomeration for the dual of *strong* ability, and we also know that he challenges Agglomeration for the "no choice" operator in general.

Lewis (1993) discusses the operator "it is true that, and such-and-such agent never had any choice about whether …" abbreviated as "Unfree". According to Lewis, the best argument for incompatibilism (CA, apparently) rests on a plausible principle that "Unfree" is closed under implication. Suppose that some premises imply a conclusion, and we prefix "Unfree" to each premise and to the conclusion. According to the closure principle, the prefixed premises imply the prefixed conclusion. Here is the full text (Lewis 1993: 169–170, fn 11):

> The closure principle is a generalization of the 'Rule Beta' that plays a leading role in Peter van Inwagen's defence of incompatibilism in *An Essay on Free Will* (1983); it first appears on page 94. The closure principle says that the logic of 'Unfree' is a 'normal' modal logic, see Brian Chellas (1980: 114–115). We can see from Chellas's Theorem 4.3(4) that the closure principle is equivalent, *inter alia*, to this combination of four principles:
>
> RE: if 'A if B' is valid, so is 'Unfree A if Unfree B',
> N: 'Unfree T' is valid, where T is an arbitrary tautology,
> M: 'Unfree (A & B)' implies 'Unfree A and Unfree B,' and
> C: 'Unfree A and Unfree B imply 'Unfree (A & B)'.
>
> The compatibilist must therefore challenge one of the four, most likely C, and Michael Slote (1982) has done so.

Let us embed this remark in the more general framework of modal logic. Let 'O' stand for an appropriate modal operator ("(in)ability", "(un) avoidability", etc.). The general logical principles governing this operator are then (where 'T' is tautology):

[11] Carlson (2000: 288, fn 12) gives credit to Krister Bykvist for this derivation. However, Chellas (1980: 122) in his 4.5.b already indicates this result.

RE.	From $\vdash A \leftrightarrow B$ infer $\vdash OA \leftrightarrow OB$
N.	OT
M.	$O(A \& B) \rightarrow (OA \& OB)$
C.	$(OA \& OB) \rightarrow O(A \& B)$

The four principles jointly result in the *general* rule of consequence:

RK. From $\vdash (A_1 \& A_2 \& \ldots A_n) \rightarrow A$ infer $\vdash (OA_1 \& OA_2 \& \ldots OA_n) \rightarrow OA$

Beta is then just a special case of RK for 'N':

From $\vdash (A \& (A \rightarrow B)) \rightarrow B$ infer $\vdash (NA \& N(A \rightarrow B)) \rightarrow NB$

Rule RE together with M yields:

RM. From $\vdash A \rightarrow B$ infer $OA \rightarrow OB$

This rule expresses a weak or *simple* consequence: if *B* is a consequence of *A*, then OB is a consequence of OA. We can easily obtain *Master* from RM ("p entails q" is equivalent to "~q entails ~p", so, according to RM, 'A~p' is a consequence of 'A~q').

References

Brown, M. 1988. "On the Logic of Ability." *Journal of Philosophical Logic* 17: 1–26.

Carlson, E. 2000. "Incompatibilism and the Transfer of Power Necessity." *Noûs* 34: 277–290.

Chellas, B. F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Chellas, B. F. 1992. "Time and Modality in the Logic of Agency." *Studia Logica* 51: 485–517.

Finch, A. and Warfield, T. A. 1998. "The *Mind* argument and libertarianism." *Mind* 107: 515–28.

Fischer, J. M. 1994. *The Metaphysics of Free Will*. Oxford: Basil Blackwell.

Hasker, W. 1989. *God, Time and Knowledge*. Ithaca: Cornell University Press

Horgan, T. 1985. "Compatibilism and the Consequence Argument." *Philosophical Studies* 47: 339–356.

Huemer, M. 2000. "Van Inwagen's Consequence Argument." *The Philosophical Review* 109: 525–544.

Kane, R. 1996. *The Significance of Free Will*. New York: Oxford University Press.

Kane, R. (ed.). 2002. *The Oxford Handbook of Free Will* 1st. Oxford: Oxford University Press.

Kane, R. (ed.). 2011. *The Oxford Handbook of Free Will* 2nd. Oxford, New York: Oxford University Press.

Kapitan, T. 1991. "How Powerful Are We?" *American Philosophical Quarterly* 28: 331–338.

Kapitan, T. 1996. "Modal Principles in the Metaphysics of Free Will." *Noûs* 30: 419–445.

Kapitan, T. 2002. "A Master Argument for Incompatibilism?" In Kane, R. (ed.). 2002: 127–157.

Kapitan, T. 2011. "A Compatibilist Reply to the Consequence Argument." In Kane, R. (ed.). 2011: 131–150.

Kenny, A. 1976. "Human Abilities and Dynamic Modalities." In Manninen, J. and Tuomela, R. (eds.). *Essays on Explanation and Understanding*. Springer: 209–232.

Lehrer, K. 1980. "Preferences, Conditionals and Freedom." In van Inwagen, P. (ed.). *Time and cause: essays presented to Richard Taylor*. Dordrecht: D. Reidel: 187–201.

Lewis, D. 1981. "Are We Free to Break the Laws?" *Theoria* 47, 113–21. Reprinted in Lewis, D. 1986. *Philosophical Papers II*. New York and Oxford: Oxford University Press: 291–298.

Lewis, D. 1988. "Relevant Implication." *Theoria* 54: 161–174.

Lewis, D. 1993. "Evil for Freedom's Sake?" *Philosophical Papers* 22: 149–172.

Lewis, D. K. 2020. "Outline of 'Nihil Obstat: An Analysis of Ability.'" *The Monist* 103: 241–44.

Lewis, D. K., Beebee, H. and Fisher, A. R. J. (eds.) 2020. *Philosophical Letters of David K. Lewis: Volume 1: Causation, Modality, Ontology*. Oxford, New York: Oxford University Press.

McKay, T. and Johnson, D. 1996. "A Reconsideration of an Argument against Compatibilism." *Philosophical Topics* 24: 113–22.

Mele, A. R. 2006. *Free Will and Luck*. Oxford: Oxford University Press.

Nolan, D. 2015. "Lewis's Philosophical Method" In Loewer, B., & Schaffer, J. (eds.). *A Companion to David Lewis*. John Wiley & Sons: 25–39.

Perry, J. 2004. "Compatibilist options." In Campbell, J. K., O'Rourke, M. and Shier, D. (eds.). *Freedom and Determinism*. Cambridge: MIT Press: 231–254.

Pyke, S. 1995. *Philosophers*. London: Zelda Cheatle Press.

Pritchard, D. 2016. *Epistemic angst: Radical skepticism and the groundlessness of our believing*. Princeton: Princeton University Press.

Ramsey, F. P. 1994. *Philosophical Papers*. Cambridge: Cambridge University Press.

Rummens, S. 2019. "The Counterfactual Structure of the Consequence Argument." *Erkenntnis*, https://doi.org/10.1007/s10670-019-00117-2.

Schnieder, B. 2008. "On what we can ensure." *Synthese* 162: 101–115.

Setiya, K. 2020. "Philosophy as Self-expression." *The Philosopher* 108 (3): 65–69.

Slote, M. 1982. "Selective Necessity and the Free-Will Problem." *The Journal of Philosophy* 79: 5–24.

Speak, D. 2011. "The Consequence Argument Revisited." In Kane, R. (ed.). 2011: 115–130.

van Inwagen, P. 1975. "The Incompatibility of Free Will and Determinism." *Philosophical Studies* 27: 185–99.

van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.

van Inwagen, P. 2004. "Freedom to Break the Laws." *Midwest Studies in Philosophy* 28: 334–350.

van Inwagen, P. 2008a. "How to think about the problem of free will." *Journal of Ethics* 12: 327–341.

van Inwagen, P. 2008b. "The Consequence Argument." In van Inwagen, P.

and Zimmerman, D. (eds.). *Metaphysics. The Big Questions* 2nd. Oxford: Blackwell: 450–456.

van Inwagen, P. 2017. *Thinking about Free Will*. Cambridge, New York: Cambridge University Press.

Vihvelin, K. 2013. *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. New York: Oxford University Press.

Widerker, D. 1987. "On an Argument for Incompatibilism." *Analysis* 47: 37–41.

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.3
Received: July 16, 2021
Accepted: November 15, 2021

Implicitness, Logical Form and Arguments

MARTINA BLEČIĆ
University od Rijeka, Rijeka, Croatia

In the paper I suggest that a loose notion of logical form can be a useful tool for the understanding or evaluation of everyday language and the explicit and implicit content of communication. Reconciling ordinary language and logic provides formal guidelines for rational communication, giving strength and order to ordinary communication and content to logical schemas. The starting point of the paper is the idea that the bearers of logical form are not natural language sentences, but what we communicate with them, that is, their content in a particular context. On the basis of that idea, I propose that we can ascribe logical proprieties to what is communicated using ordinary language and suggest a continuum between semantic phenomena such as explicatures and pragmatic communicational strategies such as (particularized) conversational implicatures, which challenges the idea that an implicatum is completely separate from what is said. I believe that this continuum can be best explained by the notion of logical form, taken as a propriety of sentences relative to particular interpretations.

Keywords: Logical form; impliciture; conversational implicatures; context.

1. *Introduction*

Philosophers have always been interested in the patterns of correct reasoning, that is, in what differentiates valid from invalid inferences. The idea was that good reasoning shows patterns that can be schematically characterized by abstracting from the specific content of the involved premises and conclusions, thus revealing a general form com-

mon to many other good inferences. Arguments constructed following these schemas are considered valid. A significant complication is that the context is important in ordinary language as to which proposition is expressed by a particular sentence. Still, ordinary language and correct reasoning are interconnected, and it is hard to talk of one without the other or to reduce correct reasoning to abstract schemas for which content and context are irrelevant. Doing so seems like engaging in a logical game devoid of any real-life implications. Perhaps this is one of the reasons there are two senses in which we use the term "logical form"—one is logical and the other is linguistic. Still, the two are sometimes used interchangeably, which poses the question of whether there is a unified account of logical form.

In this paper, I will use the idea, proposed by A. Iacona, that the bearers of logical form are not natural language sentences, but what we communicate with them, that is, their content in a particular context. This means that we can ascribe logical proprieties to what is communicated in real-life situations, using ordinary language. I will use this idea to propose a continuum between phenomena such as explicatures, in which the content of an utterance is assessed in relation to the context and the speaker, even when language is used directly and literally, and indirect and implicit communicational strategies such as (particularized) conversational implicatures. Sometimes, sentences will need a contextual supplement to express a full proposition or the specific content the speaker has in mind. This expanded sentence can then be evaluated for its truth or falsity. All of this happens on the level of explicit communication. On the implicit level, we can also provide supplements to what is said to reach the message intended by the speaker, but in this case, it will not amount only to a few words, but entire additional premises will be needed to reach the intended conclusion. I present the idea that we can look at (particularized) conversational implicatures as arguments in which the speaker provides reasons for the conclusion and does not expect an answer from the interlocutor. They are also distinguished from other forms of arguments by the fact that only one premise is explicitly expressed in them, and the conclusion (the content of the implicature, the so-called implicatum) is built jointly by the speaker and the hearer in a given conversational situation. I suggest that this shows that a loose notion of logical form can be a useful tool for assessing or evaluating everyday language, from explicit to implicit sentences. Reconciling ordinary language and logic provides formal guidelines for rational communication, giving a sense of order to ordinary communication and content to logical schemas.

2. *Logical form*

When we talk about logical form, we can have in mind two applications of the term. The first one is related to logic, where we use the notion of logical form, for example, to evaluate the validity of arguments. The

second one is said to underlie the structure of sentences and it is used in theories of meaning.

The beginnings of the study of logical form are found in the observation of patterns of inference that can be determined through the schematization of certain expressions. Aristotle, the Stoics and medieval logicians used paraphrases of natural language sentences made with the tools provided by the same language, but which aimed to emphasize patterns of correct reasoning (for example: All A are Bs, All Cs are A: All Cs are B). Much later, philosophers such as Frege, Wittgenstein and Russell aimed to clarify the logical properties of a natural language sentence by formalization in the perfect language of logic, using formal quantifiers and logical connectives[1]. A basic feature of a logically perfect language, which makes it different from natural language, is the following: every sentence has definite truth-conditions that are determined by its semantic structure and reflected in its syntactic structure (Iacona 2018: 22).

Natural language cannot satisfy this requirement. Just think of features of natural language like vagueness ("Jack is bold"), ambiguity ("Jack is going to the bank") or context sensitivity ("Jack is here"). Frege (1879, 1891), Russell (1905, 1984, 1998), and Wittgenstein (1992, 1993) shared three main claims about logical form: logical properties depend on logical form; logical form may not be visible in surface structure, and logical form is exhibited in a logically perfect language. They did not regard natural language as intrinsically interesting; they believed that natural language is incapable of being studied rigorously and systematically. This attitude towards natural language prevailed until the middle of the twentieth century. A first step in the opposite direction was prompted by the program of generative grammar in the 1950s when N. Chomsky (1976, 1995) proposed the idea of a Universal Grammar, a system of syntactic rules that underlies all natural languages. The next big step was the idea of compositionality proposed by Frege and Davidson. The idea is fairly simple: since the language has a finite vocabulary but an infinite number of sentences, it must be assumed that the meaning of a sentence depends on the meanings of its parts. According to Davidson's account, to give the logical form of a sentence is to describe its semantically relevant features against the background of a theory of truth. The central idea of Davidson's program is that meaning is a matter of truth conditions: to know the meaning of *s* is to know the conditions under which *s* is true: „Snijeg je bijel" is true if and only if snow is white: "What should we ask of an adequate account of the logical form of a sentence? Above all, I would say, such an account must lead us to see the semantic character of the sentence—its truth or falsity—as owed to how it is composed, by a finite number of applications of some of a finite number of devices that suffice for the language as a whole, out of elements drawn from a finite

[1] For a good historical overview see Pietroski 2016.

stock (the vocabulary) that suffices for the language as a whole. To see a sentence in this light is to see it in the light of a theory for its language, a theory that gives the form of every sentence in that language" (Davidson 1968: 131).

Later, Montague (1970, 1973) provided the first proper formal treatment of English based on Tarski's method. An important extension of Montague's methods concerns the treatment of indexicals and demonstratives, that is, context-sensitive expressions such as 'I', 'you', 'now', 'this', 'that' or 'there'. As suggested by Lewis (1970) and Kaplan (1977), a sentence containing indexes or demonstratives can be formally described as a sentence that has true conditions with respect to parameters, understood as sets of coordinates that provide appropriate semantic values for indexes or demonstratives appearing in the sentence. This is important because it shows that we can ascribe logical proprieties to highly contextual sentences.

Today, there is an additional element added to the conception of logical form proposed by Frege, Russell and Wittgenstein, creating this picture: logical properties depend on logical form, meaning depends on logical form, logical form may not be visible in surface structure, logical form is exhibited in a perspicuous language (see Iacona 2018: 35). Here, the logical view of logical form, which is concerned with the formal explanation of logical properties and logical relations, such as validity or contradiction and the semantic view, concerned with the formulation of a compositional theory of meaning, come together. Following Iacona, we can say that semantic theorists such as the beforementioned Davidson, Montague, Lewis and Kaplan, but also Neale (1993), Stanley (2000), and Borg (2007) share the assumption that the notion of logical form that appears in semantic theory is also able to explain the logical properties and relationships between sentences.

The question is whether the semantic notion of logical form can fulfill this logical role, for instance, when it comes to sentences involving context sensitivity. Iacona claims that it cannot. For example, consider the following simple example (Iacona 2018: 49). When you say "I am a philosopher" and I say "You are not a philosopher," in a particular context, our disagreement is not just about the difference between the way we judge the truth of each of those sentences. There is a logical relationship between them, namely, contradiction, so it cannot be the case that both sentences are simultaneously true. A semantic approach that ascribes a logical form to each sentence separately, and out of context, cannot explain the logical relations dependent on context. According to Iacona, we need a new notion of logical form. The bearers of logical form are not natural language sentences, but what we communicate with them, that is, their content in a particular context. Sometimes the logical form of a sentence will not be straightforward: "This means that there is no such thing as "the" logical form of a sentence. Sentences have logical form relative to interpretations, because they have logical form in virtue of the content they express" (Iacona 2018: 60).

An interesting idea explored by Iacona is the question of the relation between form and expressions such as non-standard quantifiers or vague terms. Consider "More than half". The semantic theorist will deny that we can explain the validity of arguments involving these expressions in terms of logical form. But there is an intuitive sense in which such arguments are good because of their form. If there are four professors of philosophy in a certain department we would not be wrong if we take the sentence "more than half of the professors in the philosophy department are happy" to yield the conclusion "more than two professors in the philosophy department are happy". When there are only three professors in the philosophy department this conclusion would be false. Still, there would be some other inference that would seem valid. It would not be wrong to conclude "more than one professor in the philosophy department is happy". The goodness of each of these conclusions, it seems, depends on the form of what the sentences express. But since the semantic theorist ascribes the same logical form in all contexts to the sentence "more than half of the professors in the philosophy department are happy", she is unable to explain these conclusions as logical.

Iacona believes that such inferences are good by virtue of their logical form. According to him, even though "more than half of" is not first-order definable, it is "first-order expressible"[2]. This means that in any interpretation the logical form of "more than half of the professors in the philosophy department are happy", can be captured by some first-order formula that embodies all the relevant logical properties that the original sentence has in that interpretation. Thus, in any interpretation there would be some cardinal number, such that "more than half of the professors in the philosophy department are happy", is true in that interpretation for the intersection of the professor and the set of lucky ones which is greater than that number. This general fact explains why we have the intuition that "more than half of the professors in the philosophy department are happy", entails other sentences because of its logical form. Iacona concludes that there have to be two distinct notions of logical form, which are equally legitimate, but serve two distinct purposes.

Iacona's lesson for us will be that the bearers of logical form are not natural language sentences, but what we communicate with them; their content in a particular context. The most important idea is that sentences have a logical form relative to interpretations because they have a logical form in virtue of the content they express. His idea that there are logical proprieties hidden in sentences containing non-standard quantifiers will also be useful since it points to the possibility that

[2] Generalized quantifiers have been studied extensively, see for example Barwise and Cooper (1981), Clark (2011), Keenan and Paperno (2012) and for an overview see Westerståhl (2019).

there is much logically describable content to be discovered in everyday communication.[3]

3. *Explicit and implicit communication*

Now we ask what is the relation between the content expressed by a sentence and the form that leads correct reasoning? Can we reduce semantics to syntax by providing a logical form for every natural language sentence as Davidson wanted? We will follow Soames (1989) in the idea that truth-functional semantics cannot explain understanding and that it is the wrong way to study semantics. But how do we proceed from here? What is the relation between meaning and context? Consider the sentence "It's raining". What proposition is expressed here? Where does it rain, and is this information important? According to Bach (2014), this sentence does not express a complete proposition, and for Recanati (2004) there is a hidden indexical that specifies the location ("in X"). The options are many. Now we will briefly explore some of them.

According to Carston (2002), the gap between the coded sentence and what is said is a fundamental feature of natural language. Consider the sentence "The pot is black". We do understand it, but can we say what is expressed by it? In which sense is the pot "black"? Is it dirty, burnt, or made from a black material? To get to this sentence's concrete content, we need to explicate it.

According to Carston, the sentence is not the truth-bearer, this role is played by the proposition that is expressed by the speaker's utterance. And this proposition is called explicature. Here is how she explains the simple sentence "She hasn't called":

> On any normal occasion of use, this will be understood as expressing a complete proposition in which it is predicated of a particular female that it is not the case that she has called (in some specific sense of 'call') some other particular person within some relevantly delimited time span up to the time of utterance. However, the sentence form itself encodes something much less specific, a non-propositional (non-truth-evaluable) logico-conceptual

[3] The idea that logical form is contextual lead some authors to be vary of the use of the term. As Dutilh Novaes writes: "(…) psychologists typically rely on a fairly naive understanding of the concept: the logical form of a sentence or argument would straightforwardly be 'read off' from its surface structure. But, at least since Russell, most philosophers are well aware of the fact that the 'logical form' of a sentence, if there is indeed such a thing, is often not straightforwardly correlated with its surface, grammatical structure. More importantly, what the experiments with the postal conditional illustrate is that logical form is not something that a sentence or argument has, in an independent, quasi-metaphysical sense; rather, logical form is at best something that speakers attribute to sentences or arguments by means of an interpretation. (…) I maintain that it is best simply to stop using the concept of 'logical form' as a property of sentences and arguments. Even understood as something that is attributed to sentences and arguments by speakers, the concept evokes too many infelicitous connotations. Rather, it seems more appropriate to speak more generally of the semantic interpretation given to a sentence by a speaker, so as to avoid conceptual muddle" (Dutilh Novaes 2012: 122).

> structure, an 'assumption schema', which functions as a template for the construction of fully propositional (truth-evaluable) logico-conceptual structures. It is this schematic logical form that the initial (purely linguistic) phase of understanding delivers and which is the input to the pragmatic processes aimed at constructing the propositional form intended by the speaker, or one similar enough to it to have the intended effects. (Carston 2002: 59)

But explicatures are not the only enrichment option for a sentence, there are also Gricean conversational implicatures. We can illustrate the difference between the two with the beforementioned pot example. The sentence we consider is "The pot is black", a possible explicature is the content that the pot is burnt. A possible implicature, derived from this particular explicature is: "Use another one". Both explicatures and implicatures depend on the intention of the speaker, but there are differences. According to Carston: "An assumption (proposition) communicated by an utterance is an 'explicature' of the utterance if and only if it is a development of (a) a linguistically encoded logical form of the utterance, or of (b) a sentential subpart of a logical form" (2002: 124). As Carston states it: "the Gricean schema for figuring out a speaker's conversational implicature(s) from what she has said is a pure piece of (…) personal-level practical belief/desire reasoning; it is conscious, rational and normative" (2002: 7). Explicatures are to be found in every sentence, which is not the case with implicatures. We can say that explicatures are necessary, and conversational implicatures are optional in the sense that the same sentence on certain occasions carries with its utterance an implicature and doesn't' carry it in others. F. Recanati has argued for two quite distinct kinds of pragmatic processes, subpropositional associative primary processes, driven solely by cognitive effort considerations, and properly inferential propositional secondary processes, guided by the standard Gricean maxims and not explainable at a sub-personal computational level. Theories along the lines of this idea have been called dual pragmatic theories. According to these theories, pragmatic processes are capable of acting twice: once before the delivery of a complete proposition expressed (that is, before determining the truth-conditional content of the sentence as uttered in a given context) and then once again to yield any implicatures of the utterance (i.e. any further, indirectly conveyed, propositions).

But not all theorists agree with this contextual conception of semantics. According to Borg, the answer to a better understanding of semantic meaning lies in what is called minimal semantics:

> A minimal semantic theory, on my terms, is a theory which seeks to give the literal meaning of types of words and sentence-types (relativized to a context of utterance) in a given natural language, and that's pretty much it. Specifically, such a theory keeps its nose out of a range of related explananda, such as how we succeed in communicating with one another using language, how we come to know about objects in the world around us, and, in general, how properly linguistic information comes to interact with the vast range of other information an agent possesses. (Borg 2004: 54)

According to Borg, what lies within reach of our formal semantic theory must involve only those features which can be recovered by simple deductive operations on the syntactic content of a sentence, "nothing which requires abductive reasoning (like mindreading) can be treated as a proper part of the semantic theory. So the semantics/pragmatics divide becomes a division between information in, or generated by, our purely formal, computational language faculty and information in, or generated by, other cognitive domains" (Borg 2004: 261).

As opposed to the contextualist idea that context is necessary to determine the content of a sentence, proponents of minimal semantics believe that syntactical features are enough to provide us with a proposition. According to them, we can determine the proposition expressed by every sentence in this formal way. But consider the sentence "Ann cannot continue", what proposition, out of any context, does it express? As I mentioned earlier, Bach believes that in some situations, sentences just do not express propositions, at least not on a purely sintactico-semantic level. Bach proposes the notion of conversational impliciture, which he considers to be closely related, but not the same as explicature[4]. He distinguishes two types of impliciture, depending on whether the hearer must do some conceptual filling in of a propositional radical or fleshing out of a minimal proposition to ascertain what the speaker means. For example, sentences like "Steel isn't strong enough" or "Willie almost robbed a bank", "though syntactically well-formed, are semantically or conceptually incomplete, in the sense that something must be added for the sentence to express a complete and determinate proposition (something capable of being true or false)" (Bach 1994: 127). On the other hand, sentences like "You're not going to die" and "I have eaten breakfast" require a degree of completion, not because they do not express a proposition, they do, but this proposition is not what is communicated by the speaker. "You are not going to die" said to a child with a cut finger does not mean that the child is immortal, but that she is not going to die from that particular cut. Likewise, someone saying that they did not have breakfast is probably not saying that they never ate breakfast.

This idea that in order to get to a complete proposition, even at the explicit level of communication we need to establish what is being communicated is in accordance with Iacona's view of logical form. Once we have, with the use of pragmatic additions, determined the content of a particular sentence, we can explore the logical form underneath it.

4. *Explicit and implicit: A matter of degree*

Levinson (2000) had discussed the relationship between logical form and implicatures. According to him, the Gricean notion of "what is said", which is the proposition expressed by the use of a sentence or the truth-conditional content of the utterance, that depends on reference

[4] See Bach (2010).

resolution, indexical fixing, and disambiguation, is too restricted. He points out that "implicatures can be seen paradoxically to play a role in the establishment of what is said" (Levinson 2002: 172). It is important to note that Levinson is focused on generalized conversational implicatures, that can fall within the scope of logical operators and other higher-level processes of semantic composition. A generalized implicature is a conversational implicature that is inferable without reference to a special context. Expressions with the form an X usually imply that X is not closely related to the speaker or subject, as in "John walked into a house yesterday". This expression implies that the house is not John's house. In this paper, I tried to broaden the role of conversational implicatures in the discussion of logical form focusing on particularized conversational implicatures, for which context is crucial, but their interpretation is still governed by patterns of correct reasoning.

I will now return to the differences between conversational implicatures and explicatures and/or implicitures. The idea is that there is a sharp distinction between the fleshed-out logical form attributed to an utterance, that is, its explicature, and fully implicit, pragmatically conveyed propositions, that is, implicatures (see Sperber and Wilson 1986: 181, 182). When we consider sentences like "It is raining", to get to a content that we can assess as true or false, we will have to specify the location we have in mind since we do not want to convey the information that it is raining somewhere, in an unspecified part of the word (or even further). This information is, in this sense, necessary. This is an impliciture, for some authors, and for others, it is an explicature. Besides, there are other communicational layers the speaker could wish to convey to the hearer. For example, if the utterance of "It's raining" is a response to the question "Have you mowed the grass?" then it would mean something like: "No, I didn't mow the grass (because it's raining)". This is a clear case of conversational implicature. For the hearer to reach the intended message there must be an adequate connection between what is said/explicated and what is implicated. Even though this should not be problematic, it seems that some formulations of the distinction between explicatures/implicitures and implicatures point to a complete separation between the implicature and the syntactico-semantic layer of the sentence.[5] Consider how Bach presents the distinction between implicitures and implicatures:

> An implicatum is completely separate from what is said and is inferred from it (more precisely, from the saying of it). What is said is one proposition and what is communicated in addition to that is a conceptually independent proposition, a proposition with perhaps no constituents in common with what is said. (…) In contrast, implicitures are built up from the explicit content of the utterance by conceptual strengthening or what Sperber and

[5] Levinson (2000) presents six criteria that are used for distinguishing explicatures from implicatures and rejects all of them and claims that they "fail to make any clear distinction between explicature and implicature" (196). Again, it should be noted that Levinson focuses on generalizes conversational implicatures.

> Wilson (1986) call 'enrichment', which yields what would have been made fully explicit if the appropriate lexical material had been included in the utterance. Implicitures are, as the name suggests, implicit in what is said, whereas implicatures are implied by (the saying of) what is said." (Bach 1994: 141)

Bach explicitly states that implicatures are completely separate from what is said. According to him, what is communicated is a proposition conceptually independent from what is said, so different from it that it does not need to have any constituents in common with it. This seems like a position that could make the understanding of implicatures seem like a mysterious process. Returning to our previous example, if the implicature in question is "I didn't mow the grass" then there must be a connection between the initial sentence "It is raining", its impliciture "It is raining in Rijeka" and its implicature. In order for a person to be able to conclude that her interlocutor did not mow the grass, she must be able to correctly link what was said and the implicature—there has to be a meaningful connection. This connection is not some just abstract relation between these communicational levels, it is a relation that can, at least to a certain extent, be formalized. To properly understand an implicature, its reconstruction will have to start at the semantic level. During this process, the initial constituents will be used in such a way to lead to the desired communicational outcome. That means that the difference between implicitures (or explicatures) and conversational implicatures is not a difference in type, but degree. There is a continuum, both formal and communicational, between the starting sentence, its potential explicatures, and its potential implicatures. In order to get to a truth-evaluable explicature, sometimes we will need to add something to its syntactico-semantic base, and sometimes we will also need to add other elements, whole premises, to get to an intended conversational implicature.

To make this clearer, we need to look at the reconstruction of implicatures (which is, just to briefly clarify, a post facto enterprise, not a real psychological interpretation) as a process of argument reconstruction. Looking at the reconstruction of conversational implicatures as a reconstruction of arguments gives epistemic strength to the belief that the listener creates based on indirectly conveyed content, it provides reasons for their conclusion. Walton and Macagno have argued that conversational implicatures should be analyzed as implicit arguments involving inference patterns that lead from a given premise to a conclusion (2013: 211). Consider the following example:

Oliver: Are you ready for the cinema?
Ana: I'm tired.

According to the systematization provided by Walton and Macagno, the dialogue can be explained as an argument from cause, that is, as a type of causal argument in which an event is associated with the cause that led to it. The general scheme is as follows:

Major premise: In general, we can say that B will happen (or that it can happen) if A happens.
Minor premise: In this case, A happened (or it can happen).
Conclusion: Therefore, in this case B will happen (or can happen)

In the example given, Ana replies that she is tired instead of providing a direct answer to the question of whether she is ready to go to the movies. The purpose of her utterance is not to inform Oliver of her psychophysical condition, but to lead him to conclude from cause to effect. Fatigue is, at least in principle, incompatible with going to the movies: if someone is tired, then they should stay home. If someone stays at home, they cannot watch a movie at the cinema. Such a causal relationship is presented as a choice. Either A or B, not A, hence B.

Conversational implicatures can be considered as special cases of enthymemes, that is, instances of arguments with unstated premises or conclusions (see Blečić 2018). Enthymemes are reconstructed on the basis of their explicit elements using deductive, inductive or abductive forms of reasoning. The missing premises are generally taken to be assumptions that are needed to make the argument valid. Because of that, we should say that enthymemes are not the same as the reconstructed arguments based on them. We can say that the reconstructed argument represents the original one. The same goes for conversational implicatures. The reconstructed argument is not the same as the utterance, but it can be a representation of it and of its underlying logical structure. According to Gilbert (1991), incomplete arguments should be supplemented by assumptions that are convincing to the target audience and that fit into the position advocated by the speaker, of course, if there is no evidence to the contrary. Applying these ideas to conversational implicatures as reasonable arguments, we can say that the missing premises must be filled with assumptions that are plausible to the listener and that, at least seemingly, fit the speaker's attitude and the message he wants to convey. Arguments that we can say are made by implicatures, or that can be reconstructed on the basis of them, will always contain a dose of uncertainty, that is, they will not have the power of deduction, which does not allow exceptions. Yet such a thing is quite common in everyday reasoning, as noted by Mercier and Sperber (2017: 163 and 164). We can consider conversational implicatures as reason-giving arguments in which the speaker addresses a hearer who does not need to reply. In those cases, the speaker is not trying to convince the hearer to accept his position, as in the case of disputational arguments, but is explicitly stating a reason in support of the intended message. By grasping the intended message, the hearer intuitively accepts the reasons provided for it and can also reconstruct the argumentative path that leads from an explicit reason, qua premise, to the intended conclusion. A competent language speaker confronted with a statement that potentially carries a conversational implicature will attribute rationality to the sender of the mes-

sage and begin the search for their communicative intent. Attributing this intention and the general knowledge of the speech community of which he is a part will be essential elements in his argumentative reconstruction of the speaker's message, and thus in justifying the belief he has created. Such a reconstruction will have an objective character: it does not depend solely on the speaker's intention, at least not if we do not view the speaker as a competent member of the speaking community who respects a whole range of established linguistic and social conventions. To be understood, conversational implicatures must have an argumentative basis. Here, the idea of a logical form, at least if broadened to encompass a looser, non-deductive sense, comes into play, providing good reasoning and communication patterns. Good reasoning shows patterns that can be schematically characterized by abstracting from the specific contents of the involved premises and conclusions, thus revealing a general form common to many other good conclusions. In the case of implicatures, such patterns will rarely be deductive, but they will nevertheless be crucial for their correct interpretation.

5. *Concluding remarks*

In order to assess the content of a sentence, we need to consider it contextually. Some sentences will need additional elements to convey a full proposition or, if they already carry a complete proposition, to modify it in order to suit what the speaker had in mind. This view of language is compatible with the idea that the bearers of logical form are not natural language sentences but their content in a particular context. Sentences have a logical form relative to interpretations because they have a logical form in virtue of the content they express. This idea reconciles what is sometimes seen as distinct areas of human interest—logic and pragmatics. Logic deals with formal principles of reasoning, and pragmatics with the use of language. How can we reconcile the two? Perhaps we should not start the exploration of the logic that governs communication at the syntactico-semantic level but the pragmatic level. We can assess the truth-value of a sentence only after we have explicated its content. Sometimes, we will be in the position to search for truth-values of the implicatum. We can always ask whether one implicates is true or false. But, before that, we will have to get to the right communicational content. To do that, we must look at conversational implicatures as implicit arguments. Implicatures are a rational communicational strategy, and rules are governing their communicatively cooperative production and reception. I suggest that these are the rules that govern the creation of good arguments, regardless of their status as deductive, inductive or abductive arguments. Of course, these kinds of arguments are governed by very different principles, but we need to consider them all if we want to address all possible pragmatic communicational possibilities.

We create meaningful connections between what is said and the

environment (linguistic and non-linguistic), and to do so, these connections must be logically (in a broad sense) appropriate. Logical proprieties lead rational and meaningful communication; both implicit and explicit, making the notion of logical form useful for pragmatics.

References

Bach K. 2010. "Impliciture vs Explicature: What's the Difference?" In B. Soria and E. Romero (eds.). *Explicit Communication*. Palgrave Studies in Pragmatics, Language and Cognition. Palgrave Macmillan, London. https://doi.org/10.1057/9780230292352_8

Bach, K. 1994. "Conversational Impliciture." *Mind and Language* 9 (2): 124–162.

Bach, K. 2014. "Context Dependence." In M. Garcia-Carpintero i M. Kolbel (ed.). *The Bloomsbury Companion to the Philosophy of Language*. Bloomsbury Publishing: 153–184.

Barwise, J. and Cooper, R. 1981. "Generalized Quantifiers and Natural Language." *Linguistics and Philosophy* 4(2): 159–219.

Blečić, M. 2018. "Do Conversational Implicatures Express Arguments?" *Croatian Journal of Philosophy* 11 (2): 335–350.

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

Chomsky, N. 1976. "Conditions on rules of grammar." *Linguistic Analysis* 2: 303–351.

Chomsky, N. 1995. *The minimalist program*. Cambridge: MIT Press.

Clark, R. 2011. "Generalized Quantifiers and Number Sense." *Philosophy Compass* 6 (9): 611–621. doi:10.1111/j.1747-9991.2011.00419.x

Davidson, D. 1968. "On saying that." *Synthese* 19: 130–146.

Dutilh Novaes, C. 2012. *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge: Cambridge University Press.

Frege, G. 1967 [1879]. "Concept script, a formal language of pure thought modelled upon that of Arithmetic". In J. van Heijenoort (ed.). *From Frege to Godel: A sourcebook in mathematical logic*. Cambridge: Harvard University Press: 5–82.

Frege, G. 1980 [1891]. "Function and concept." In P. Geach and M. Black (eds.), *Translations from the philosophical writings of Gottlob Frege*. Oxford: Blackwell.

Gilbert, M. A. 1991. "The Enthymeme Buster: A Heuristic Procedure for Position Exploration in Dialogic Dispute." *Informal Logic* 13 (3): 159–166.

Iacona, A. 2018. *Logical Form: Between Logic and Natural Language*. Cham: Springer International Publishing.

Kaplan, D. 1977." Demonstratives." In J. Perry, J. Almog and H. Wettstein (eds.). *Themes from Kaplan*. New York: Oxford University Press: 481–563.

Keenan, E. L. and Paperno. D. (eds.). 2012. *Handbook of Quantifiers in Natural Language*. Dordrecht: Springer Netherlands. doi:10.1007/978-94-007 2681-9

Levinson, S. C. 2000. *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge: MIT Press.

Lewis, D. 1970. "General semantics." In *Philosophical papers* (Vol. I, 189–229). Oxford: Oxford University Press.

Macagno, F. and Walton, D. 2013. "Implicatures as Forms of Argument." In A. Capone et al. (eds.). *Perspectives on Pragmatics and Philosophy*. Berlin/New York: Springer: 203–224.

Mercier, H. and Sperber, D. 2017. *The Enigma of Reason*. Cambridge: Harvard University Press.

Montague, R. 1970. "Universal grammar." In *Formal philosophy*. New Haven: Yale University Press.

Montague, R. 1973. "The proper treatment of quantification in ordinary English." In *Formal philosophy*. New Haven: Yale University Press.

Neale, S. 1993. "Logical form and LF." In C. Otero (ed.). *Noam Chomsky: Critical assessments*. London: Routledge: 788–838

Pietroski, P. 2016. "Logical Form." *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2016/entries/logical-form/>.

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

Russell, B. 1905. "On denoting." *Mind* 14: 479–493.

Russell, B. 1984. "Theory of knowledge." In *Collected papers* (Vol. VI). London: Allen and Unwin.

Russell, B. 1998. *The philosophy of logical atomism*. La Salle: Open Court.

Soames, S. 1989. "Semantics and Semantic Competence." *Philosophical Perspectives*, Vol. 3, *Philosophy of Mind and Action Theory*: 575–596.

Sperber, D., and Wilson, D. 1986. *Relevance*. Oxford: Blackwell

Stanley, J. 2000. "Context and logical form." *Linguistics and Philosophy* 23: 391–434.

Westerståhl, D. 2019., "Generalized Quantifiers", *The Stanford Encyclopedia of Philosophy* (, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/generalized-quantifiers/>.

Wittgenstein, L. 1992. *Tractatus logico-philosophicus*. London: Routledge.

Wittgenstein, L. 1993. "Some remarks on logical form". In J. Klagge and A. Nordmann (eds.). *Philosophical occasions*. Indianapolis: Hackett: 29–35

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.4
Received: March 16, 2021
Accepted: December 1, 2021

On Formalizing Logical Modalities

LUIGI PAVONE
University of Palermo, Palermo, Italy

This paper is in the scope of the philosophy of modal logic; more precisely, it concerns the semantics of modal logic, when the modal elements are interpreted as logical modalities. Most authors have thought that the logic for logical modality—that is, the one to be used to formalize the notion of logical truth (and other related notions)—is to be found among logical systems in which modalities are allowed to be iterated. This has raised the problem of the adequacy, to that formalization purpose, of some modal schemes, such as $\mathbf{S}_4$ *and* $\mathbf{S}_5$*. It has been argued that the acceptance of* $\mathbf{S}_5$ *leads to non-normal modal systems, in which the uniform substitution rule fails. The thesis supported in this paper is that such a failure is rather to be attributed to what will be called "Condition of internalization." If this is correct, there seems to be no normal modal logic system capable of formalizing logical modality, even when* S_5 *is rejected in favor of a weaker system such as* S_4*, as recently proposed by McKeon.*

Key words: Logical truths; logical formality; uniform substitution; Kripke semantics; Carnap-style semantics.

1. *Introduction*

Kripke (1959, 1963) defined general validity for extending the notion of a valid formula to intensional languages, providing a modal semantics that has proved very useful in clarifying a wide range of modal notions. This paper addresses the problems that arise under a special interpretation of modal operators in terms of logical modalities, i.e., reading "*L*" as "it is logically necessary that…", and "*M*" as 'it is logically possible that…'. To begin with, can Kripkean general validity capture the meaning of a valid formula in a first-order modal language when modal operators are interpreted in such a way? A line of thought, which can be traced back to Pollock (1966), says it cannot. This view invokes

an alleged intuitive truth about logical possibility—let us call it "Necessary Logical Possibility" (NLP)—according to which true sentences about what is logically possible are logically true. For example, consider the following logically possible sentence, (i) "Jones passes the exam". Since (ii) "possibly, Jones passes the exam", obtained from (i) by prefixing the logical possibility adverb "possibly", is true, NLP states that (ii) is logically true as well. Let us call "logical possibility sentences" all true sentences about the logical possibility of states of affairs.

If NLP is right, there are intuitive modal logical truths, logical possibility sentences, such as (ii), that do not fall under the Kripkean concept of general validity. The point against Kripke semantics is that the purely formal (or logical) sense of modal operators is not captured if they are allowed to range over arbitrary non-empty sets of (logically) possible worlds. As Cocchiarella (1975) puts it, such a construction of modal operators,

> by allowing the exclusion of some of the worlds (models) of a logical space, imports material conditions into the semantics of modal operators. This exclusion, however appropriate for the representation of non-logical […] modalities, is quite inappropriate for the representation of what are purported to be merely formal or logical modalities. (1975: 13)

The aim of construing an alternative semantics for logical modality was pursued in different (but convergent) ways in Cocchiarella (1975) and Hanson and Hawthorne (1985), by using a fixed collection of logically possible worlds. As we will see in detail in Section 3, their proposal can be thought of as a version of Carnap's modal logic (1947) based on extensional models. Let us call it "Carnap-style semantics."

The main problem with this alternative semantics raises on the basis of what can be called the Quinean sense of logical formality, on which logical formality is closely related to the principle of Uniform Substitution (US): "a logical truth […] is definable as a sentence from which we get only truths when we substitute sentences for its simple sentences" (Quine 1970: 50). As noticed by Makinson (1966), and more recently by McKeon (2005), US fails in Carnap-style semantics, in which the following puzzling situation arises: the notion of a valid formula respects the intuitions of NLP, but at the cost of abandoning the formality of logic conceived in terms of substitutivity.

Section 2 shows why Kripke semantics should be considered inadequate to represent logical modality. Section 3 shows how proponents of Carnap-style semantics propose to solve the problems affecting Kripke semantics. Section 4 shows that Carnap-style semantics is not without problems in its own right, since it violates US, thereby undermining the notion of logical formality understood in terms of substitutivity.

McKeon (2005) has argued that the main reason for the failure of US is the acceptance of the modal scheme $\mathbf{S}_5$, i.e. "$M\varphi \rightarrow LM\varphi$", as a valid logical scheme in formalizing logical modality. In defense of Kripke semantics, the author maintains that the correct logic for logical modal-

ity must be a system that is "at least as strong as S_4, but not as strong as S_5" (2005: 313). In contrast, I will argue that substitutivity fails due to a more general *desideratum* for modal systems that aim to represent logical modality—what I will call "Condition of Internalization" (CI). If this is correct, McKeon's proposal to preserve US within the framework of Kripke semantics fails, due to the normality of the formalization he proposes, where by "normality" it is meant the property of being a logic in which the propositional tautologies, the rule of Necessitation (N), i.e. if φ is logically true so is $L\varphi$, and the Distribution axiom, i.e. $L(\varphi \rightarrow \psi) \rightarrow (L\varphi \rightarrow L\psi)$, hold. Section 4.1 considers Schurz's proposal (2001) to define logical formality in terms of "semantically isomorphic substitutions". I will argue that such a definition is circular. Finally, in Section 4.2 I will argue against a further argument proposed by McKeon in favor of Kripke semantics, what I will call "the argument from analogy". That argument appears to be flawed if one takes into account all the implications of CI when it is applied to the property of being context-insensitive that is unanimously attributed to all valid first-order formulas.

2. *Kripke semantics*

Let $\mathcal{L}$ be a first-order language without individual constants. $\mathcal{L}$ is the same as that for the standard first-order logic (with the symbol of identity) but supplemented with the modal letter "L" and the corresponding formation rule: if φ is a formula, so is $L\varphi$. One can add the modal letter "M" for the notion of possibility, by definition: $M\varphi$ means $\neg L\neg\varphi$. The further details are assumed to be defined. Kripke possible worlds semantics can be introduced as follows.

Let $w = \langle d_0, d_i, V\rangle$ be an outer/inner sub-model such that d_0 is a possible empty outer domain, and d_i is a non-empty inner domain.[1] These two sets are disjoint and their union counts as the domain of discourse of w, that is $d(w)$. V is an interpretation function assigning extensions to the predicate letters of $\mathcal{L}$, so that for any n-place predicate letter $\boldsymbol{P}$, $V(\boldsymbol{P}) \subseteq d(w)^n$.

Let φ be an atomic formula $\boldsymbol{P}^n(\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)$, and υ an assignment from individual variables into $d(w)$, we say that υ satisfies[w] (satisfies at w) φ iff $\langle \upsilon(\boldsymbol{x}_1), \ldots, \upsilon(\boldsymbol{x}_n)\rangle \in V(\boldsymbol{P})$. If φ has the form $\boldsymbol{x}_1 = \boldsymbol{x}_2$, υ satisfies[w] φ iff $\upsilon(\boldsymbol{x}_1) = \upsilon(\boldsymbol{x}_2)$. If φ has the form $\neg\psi$, υ satisfies[w] φ iff υ does not satisfies[w] ψ. If φ has the form $\psi \vee \gamma$, υ satisfies[w] φ iff υ satisfies[w] ψ or υ satisfies[w] γ. First-order quantifiers are restricted to inner domains, so that υ satisfies[w] $\forall\boldsymbol{x}\psi(\boldsymbol{x})$ iff for every υ' different from υ at most for $\boldsymbol{x}$ such that $\upsilon'(\boldsymbol{x}) \in d_i$, υ' satisfies[w] $\psi(\boldsymbol{x})$.

A Kripke model is a four-tuple $\mathfrak{M} = \langle W, @, R, D\rangle$, where W is an arbitrary non-empty set of sub-models, @ is a member of W representing

[1] Intuitively, d_i is the set of all objects existing in w, and d_0 is the set of all merely possible object with respect to w.

the actual state of affairs, R is the customary accessibility relation defined on W. For any $w \in W$, $d(w) = D$. Note that W is allowed to vary from model to model: if $\mathbf{W}$ is the set of *all* possible worlds, $W \subseteq \mathbf{W}$. A semantically autonomous subset of Kripke models—called "constant domains semantics"—consists of those models $\langle W, @, R, D\rangle$ such that for every sub-model $w \in W$, $d_i(w) = D$.

The notion of general validity can be defined in terms of truth[$\mathfrak{M}$] (truth in $\mathfrak{M}$) in a number of steps. Let φ be an atomic formula, and υ an assignment from individual variables into D, we say that υ satisfies[$\mathfrak{M}$] (satisfies in $\mathfrak{M}$) φ at a world $w \in W$ iff υ satisfies[w] φ. If φ has the form $\boldsymbol{x}_1 = \boldsymbol{x}_2$, υ satisfies[$\mathfrak{M}$] at a world $w \in W$ iff $\upsilon(\boldsymbol{x}_1) = \upsilon(\boldsymbol{x}_2)$. Satisfaction[$\mathfrak{M}$] conditions for propositional connectives are defined in the standard way. First-order quantifiers are restricted to inner domains, so that υ satisfies[$\mathfrak{M}$] $\forall \boldsymbol{x}\psi(\boldsymbol{x})$ at a world $w \in W$ iff for every υ' different from υ at most for $\boldsymbol{x}$ such that $\upsilon'(\boldsymbol{x}) \in d_i(w)$, υ' satisfies[$\mathfrak{M}$] $\psi(\boldsymbol{x})$ at w. Modal operators are semantically defined as follows: υ satisfies[$\mathfrak{M}$] $L\varphi$ at a world $w \in W$ iff υ satisfies[$\mathfrak{M}$] φ at every worlds $w' \in W$ such that $(w, w') \in R$. A formula φ is said to be true[$\mathfrak{M}$] at a world $w \in W$ iff for any υ, υ satisfies[$\mathfrak{M}$] φ at w. φ is said to be true[$\mathfrak{M}$] iff φ is true[$\mathfrak{M}$] at @. Finally, a formula φ is said to be generally valid iff φ is true[$\mathfrak{M}$] for any model $\mathfrak{M}$.

If the accessibility relation R is reflexive, transitive, and symmetrical, $\mathbf{S}_5$ turns to be generally valid. Under the logical interpretation of the modal operators, $\mathbf{S}_5$ formalizes NLP. Indeed, under that interpretation, that scheme says: if a sentence φ is logically possible, then it is logically necessary that it is so. As pointed out by McKeon, the acceptance of NLP entails the acceptance of the validity of $\mathbf{S}_5$ (cf. 2005: 308). Therefore, those who accept NLP will be committed to formalize the notion of logical modality within modal systems in which $\mathbf{S}_5$ is valid. Such a correspondence between NLP and $\mathbf{S}_5$, due to the fact that $\mathbf{S}_5$ is the formalization in the object language of the meta-theorem NLP, is just a case of a more general *desideratum* for modal systems that aim to represent logical modality—let us call it "Condition of Internalization" (CI). CI can be stated as follows: whenever Γ is a sentence in the meta-language of a logical system and φ is the formalization of it in the object-language of that system, Γ is a meta-theorem of the system iff φ is valid. For example, if Γ is the meta-theorem (of a logical modal system) that *all true sentences are logically possible*, the scheme $\varphi \rightarrow M\varphi$, which formalizes Γ in the object-language (of that system) under the logical reading of the modal operators, must be valid; if it was not, the following puzzling situation would arise: there would be a model $\mathfrak{M}$ such that some sentence of the form $\varphi \rightarrow M\varphi$ would not be true[$\mathfrak{M}$], in contrast with Γ, which states that it is not logically possible that something true is not logically possible.

The main problem pointed out by the critics of Kripke semantics is that it is unable to account for the intuitive truth expressed by NLP.

Although $\mathbf{S}_5$, *qua* formalization of NLP, can be assumed as generally valid (with the appropriate conditions on the accessibility relation), yet NLP cannot be a meta-theorem of the system S_5. In fact, NLP requires that all sentences of the form $M\varphi$ be logically true, for any consistent first-order sentence φ, but that is not the case for consistent first-order sentences that are not valid in first-order logic. Let φ be a first-order sentence of that sort, for example (i) again. Since there is a sub-model w such that the contradictory of (i), i.e. (iii) "Jones does not pass the exam", is true at w—the existence of such a sub-model is guaranteed by the fact that (i) is not of a valid first-order logical form –, a Kripke model $\mathfrak{M} = \langle W, @, R, V\rangle$ can be formed out of w in such a way that $W = \{@\}$, and $@ = w$. Given that (iii) is true at @, and @ is the only possible world in W, the corresponding modal sentence (ii), of the form $M\varphi$, will be false in $\mathfrak{M}$, contrary to what NLP states: that all logical possibility sentences are logically true. In contrast to NLP, there are counter Kripke models for $M\varphi$, for any consistent but not valid first-order sentence φ. In other terms, the constraint CI turns out to be violated with respect to the relationship between NLP and $\mathbf{S}_5$.

The Kripkean notion of general validity turns out to be undermined by some others difficulties identified by Pollock (1966). According to Pollock, Kripke semantics does not capture a fundamental feature of logical truths, i.e. the fact that they are not *context-sensitive*: logical truths should be true regardless of their domain of discourse.[2] For example, sentences of the form $\exists\boldsymbol{x}\psi(\boldsymbol{x}) \wedge \exists\boldsymbol{x}\neg\psi(\boldsymbol{x})$ should not count as logical truths, because they are true only if the intended domain of discourse contains more than one object. Those sentences are *context-sensitive*, because they are *domain-sensitive*. Now, if "L" means "it is logically true that…", the scheme $\neg L(\exists\boldsymbol{x}\psi(\boldsymbol{x}) \wedge \exists\boldsymbol{x}\neg\psi(\boldsymbol{x})$ formalizes in the object language the meta-theorem that sentences of the form $\exists\boldsymbol{x}\psi(\boldsymbol{x}) \wedge \exists\boldsymbol{x}\neg\psi(\boldsymbol{x})$ are not logically true. Therefore, it should be valid, and the scheme $L\neg L(\exists\boldsymbol{x}\psi(\boldsymbol{x}) \wedge \exists\boldsymbol{x}\neg\psi(\boldsymbol{x}))$ as well in virtue of the axiom $\mathbf{S}_5$, but this is not the case in Kripke semantics (cf. 1966: 316). Again, the condition CI is not obeyed.

3. *Carnap-style semantics*

In formalizing logical modalities, those who accept NLP are committed to a modal logic system in which $\mathbf{S}_5$ is valid and NLP is a meta-theorem of that system, i.e. a logic in which all true sentences of the form $M\varphi$ are logically true. This aim was pursued in the framework of Carnap's modal logic (1947) by using a fixed space of all logically possible worlds.

Carnap's modal language had an infinite collection of individual constants. The notion of logical necessity, relative to an appropriate domain of discourse, was defined as truth in every state-description,

[2] Indeed, the domain of discourse of a sentence can be considered as a part of a larger context that also includes other elements, such as the speaker, time, etc.

where a state-description is a set S such that for any atomic sentence φ, just one of φ or $\neg\varphi$ is in S. φ holds at S iff $\varphi \in S$. Given an assignment function υ from individual variables into *intensions*, that is, assignments "of exactly one individual constant to each state-description" (Carnap 1947: 181), Carnapian satisfaction conditions can be defined as follows. If φ is an atomic formula $\boldsymbol{P}^n(\boldsymbol{x}_1 \ldots \boldsymbol{x}_n)$, υ satisfies[S] (satisfies at S) φ iff $\upsilon(\boldsymbol{x}_1) = \boldsymbol{i}_i$, where $\boldsymbol{i}_i$ is an intension, $\boldsymbol{i}_i(S) = \boldsymbol{a}_i$, and the sentence obtained from φ by appropriate substitutions, i.e. $\boldsymbol{P}^n(\boldsymbol{a}_1 \ldots \boldsymbol{a}_n)$, holds at S. If φ has the form $\boldsymbol{x}_1 = \boldsymbol{x}_2$, υ satisfies[S] φ iff $\upsilon(\boldsymbol{x}_1) = \boldsymbol{i}_1$, $\upsilon(\boldsymbol{x}_2) = \boldsymbol{i}_2$, and $\boldsymbol{i}_1(S) = \boldsymbol{i}_2(S)$. If φ has the form $\neg\psi$, υ satisfies[S] φ iff υ does not satisfies[S] ψ. If φ has the form $\psi \vee \gamma$, υ satisfies[S] φ iff υ satisfies[S] ψ or υ satisfies[S] γ. If φ has the form $\forall\boldsymbol{x}\psi(\boldsymbol{x})$, υ satisfies[S] φ iff for any υ' different from υ at most for $\boldsymbol{x}$, υ' satisfies[S] $\psi(\boldsymbol{x})$. Finally, if φ has the form $L\psi$, υ satisfies[S] φ iff υ satisfies[S'] for every state-description S'.

Carnap's modal logic validates both $\mathbf{S}_4$ and $\mathbf{S}_5$. NLP is ensured by the uniqueness of the logical space of possibilities. If φ is a satisfiable sentence whatsoever, then there is a state-description S such that φ holds at S. Since S is in the single fixed logical space of possibility, $M\varphi$ will hold at all state-descriptions.

This approach to validity for modal languages was adopted using extensional models, instead of state-descriptions, by Montague (1974), Beth (1960), and more recently by Cocchiarella (1975) and Hanson and Hawthorne (1985). A simple way to state the Carnapian notion of validity in terms of extensional models—call it C-validity—is to use our language $\mathcal{L}$ and constant domains semantics. One can state that a Carnap-style model is a constant domains model $\langle \mathbf{W}, @, R, D\rangle$ such that R is reflexive, transitive, symmetrical, and $\mathbf{W} = \{w : d_i(w) = D\}$, i.e. the set of all sub-models of the same domain of discourse. Following Carnap, the number of the objects of the domain of discourse can be infinite. Thus a sentence will be valid iff it is true in all models of the same logical space of possibility.[3] This time, the sentence (ii) turns out to be logically true, for it instantiates a C-valid formula, in accordance with the *desiderata* of NLP.

4. *Formality of logics and the condition of substitutivity*

Carnap-style modal semantics ensures NLP and the validity of $\mathbf{S}_5$ by using a single logical space including all logical possibilities, but at the cost of abandoning Quine's sense of the formality of logic, according to

[3] An alternative extensional version of the Carnapian notion of validity—call it C-validity*—can be defined by stating that a formula is C-valid* iff it is true in all models for *any* domain of discourse. However, this choice is open to some problems, because not all logical possibility sentences are C-valid*. All logical possibility sentences of the form $M\exists\boldsymbol{x}\exists\boldsymbol{y}(\boldsymbol{x} \neq \boldsymbol{y})$ are not C-valid*, although $\exists\boldsymbol{x}\exists\boldsymbol{y}(\boldsymbol{x} \neq \boldsymbol{y})$ is a logically satisfiable scheme. On the other hand, those of the form $M\exists\boldsymbol{x}\forall\boldsymbol{y}(\boldsymbol{x}=\boldsymbol{y})$ are C-valid*, but not C-valid. Thus NLP would hold in general for all logical possibility sentences whose truth conditions do not require constraints on the domains of discourse in addition to that of being non-empty.

which logical formality is closely related to US: if φ is valid, so is $s(\varphi)$, where $s(\varphi)$ is obtained from φ by a substitution function s from predicative letters into formulas of the object-language. The weak point of this sort of modal semantics is that the class of its valid formulas is not closed under US. Let φ be a sentence of the form $M\psi$, where ψ is an arbitrary atomic sentence. φ is C-valid in virtue of the Carnap-style semantic rules, but if s is a substitution function such that $s(\psi) = \psi \wedge \neg\psi$, $s(\varphi)$ will be obviously false, and even logically false. Why does substitution fail in Carnap-style modal semantics?

According to Makinson (1966), this is because "schematism" is neglected when modal formulas are understood in terms of what the author calls "naive modal rules", which are exemplified by Carnap's modal semantics. By schematism he means the thesis that a logical scheme is *acceptable* iff all its instantiations by uniform substitutions are logically true in terms of the naive modal rules (cf. 1966: 334). Based on this, Makinson argued that the scheme $M\varphi$ is not acceptable, because some of its instantiations, e.g. those of the form $M(\varphi \wedge \neg\varphi)$, are not logically true. However, in Makinson's proposal, US preserves the property of being acceptable (referred to logical schemes), but not the property of being logically true referred to sentences, thereby leaving open the possibility to get false sentences from logical truths by uniform substitution. This amounts to rejecting Quine's understanding of logical formality in terms of substitutivity. In fact, consider the true sentence (ii), of the form $M\varphi$. Following Makinson, (ii) is logically true, in virtue of the naive modal rules, although the modal logical scheme from which it is formed out is not acceptable. This begs the question: in what sense should the truth of (ii) be a matter of form or structure? There seems to be no straightforward answer in Makinson's combination of schematism and naive modal rules.

The failure of US is ascribed to the scheme $\mathbf{S}_5$ by McKeon (2005). For the safety of substitutivity, McKeon proposed to reject NLP, which obliges us to accept the validity of $\mathbf{S}_5$. According to McKeon, the correct formalization of the logic of logical modality is a modal system that is "at least as strong as S_4, but not as strong as S_5". The acceptance of the validity of $\mathbf{S}_4$ can be motivated by some considerations. First, once iteration is admitted, $\mathbf{S}_4$ seems to be more intuitive than $\mathbf{S}_5$. Unlike $\mathbf{S}_5$, $\mathbf{S}_4$ does not violate US. Moreover, $\mathbf{S}_4$ formalizes N, which characterizes all normal modal systems. In fact, under the logical reading of the modal operators, $\mathbf{S}_4$ says that if a sentence is logically true, it is logically true that it is so. On the grounds of CI, it is not possible to accept N without accepting the validity of $\mathbf{S}_4$, just as we cannot accept NLP without accepting the validity of $\mathbf{S}_5$. Since N is a meta-theorem of normal modal logics, and $\mathbf{S}_4$ formalizes it (under the logical reading of the modal operators) in their object-language, the choice of a system that is "at least as strong as S_4", as suggested by McKeon, is needed if one wishes to formalize logical modality in a normal modal system.

My thesis is that McKeon's defense of Kripke semantics, whose aim is to preserve US, cannot work. In fact, the modal logical system proposed by McKeon would require a semantics validating the formalization in the object-language not only of N, i.e. $\mathbf{S}_4$, but also of US, in accordance with CI. The correct logic for logical modality should be such that the following principle holds: if $L\varphi$ is true, so is $Ls(\varphi)$, where $s(\varphi)$ is an instantiation of φ by uniform substitution. Accordingly, the following logical modal scheme should be valid: $L\varphi \rightarrow Ls(\varphi)$. Let us call it "**US**", which is the schematic internalization of US in the object-language. Under the logical reading of its modal elements, **US** says that if φ is logically true, so is any instantiation of it by US. That lands us with a version of the same old problem: US will be violated in Kripke semantics for any hypothetical normal modal logic in which its schematic internalization is valid.

Let φ be an interpreted formula of the form $M\psi$, where ψ is atomic. Two possible cases can be hypothetically considered: when φ is logically true, when it is not. If φ is logically true, then all instantiations of it by US will be logically true, but this contrasts to the evidence that some of them, those of the form $M(\psi \wedge \neg\psi)$, are logically false. On the other hand, if φ is not logically true, then the contradictory of it, of the form $L\neg\psi$, will be (at least) logically possible, i.e. there must be a logically possible state of affairs at which $L\neg\psi$ is true. Therefore, it must be the case that for some Kripke model $\mathfrak{M} = \langle W, @, R, D\rangle$, ψ fails to be true in every possible $w \in W$, from which the truth[$\mathfrak{M}$] of $L\neg\psi$ is obtained. Under the logical reading of modal operators, that $L\neg\psi$ is true[$\mathfrak{M}$] means that it is true[$\mathfrak{M}$] that $\neg\psi$ is logically true. By the following instantiation of **US**, i.e. $L\neg\psi \rightarrow Ls(\neg\psi)$, and *modus ponens*, it must also be the case that for any substitution function s, $Ls(\neg\psi)$ be true[$\mathfrak{M}$], for some $\mathfrak{M}$. This means that if $s(\neg\psi) = \neg(\psi \vee \neg\psi)$, then $L\neg(\psi \vee \neg\psi)$ must be true[$\mathfrak{M}$], for some $\mathfrak{M}$. Since the scheme $\psi \vee \neg\psi$ is a propositional tautology, there is no logical possible state of affairs at which $\psi \vee \neg\psi$ will fail to be true, hence it is not possible that $\neg(\psi \vee \neg\psi)$ is logically true, and so there is no Kripke model $\mathfrak{M}$ such that $L\neg(\psi \vee \neg\psi)$ is true[$\mathfrak{M}$]. Such a contradictory circumstance can be described as follows: in the logic of logical modality, one will inconsistently get both $ML\neg(\psi \vee \neg\psi)$, namely it is logically possible that $\neg(\psi \vee \neg\psi)$ is logically true, and $\neg ML\neg(\psi \vee \neg\psi)$, namely it is not logically possible that $\neg(\psi \vee \neg\psi)$ is logically true.

This suggests that US fails in every normal modal logic in which its internalization **US** is assumed to be valid in accordance with CI. If this is right, McKeon's proposal to combine US with a normal modal system "at least as strong as S_4, but not as strong as S_5" fails because of the normality of the modal system proposed.

4.1 *Semantically isomorphic substitutions*

According to Schurz (2001), Carnap's logic is "the only complete modal logic" if modal operators express logical modalities (2001: 365). He addressed the problem of logical formality with the C-valid scheme $M\varphi$ (with φ logical consistent) by redefining the notion of logical formality in terms of what the author call "semantically isomorphic substitutivity".

He proposes to distinguish between semantically isomorphic substitutions and semantically homomorphic substitutions, arguing that logics must be closed under the former, but not (necessarily) under the latter. A substitution function from simple sentences into formulas of the object-language counts as semantically isomorphic "iff it preserves the semantical freedom of interpretations",[4] it is homomorphic otherwise (cf. 2001: 371). A substitution S preserves the semantical freedom of interpretations iff for any primitive φ, $s(\varphi)$ is such that if φ is a logically contingent sentence with a certain degree of semantical freedom, then so is $s(\varphi)$. More technically, if $\langle \mathbf{W}, @, R, D\rangle$ is a Carnap-style constant domains model, the substitution function s is said to be semantically isomorphic iff $\mathbf{W}_s = \mathbf{W}$, where $\mathbf{W}_s$ is the set of all sub-models for $s(\varphi)$, i.e. the set of all interpretations of sentences assigned by function s to primitives.[5] Schurz argued that closure under semantically isomorphic substitutions is "the strongest reasonable requirement concerning closure under substitutions" (2001: 371). Although Carnap-style modal semantics is not closed under semantically homomorphic substitutions, yet it is under semantically isomorphic substitutions, and that would be enough to ensure formality.

According to Schurz, all true sentences of the form $M\varphi$ are formal logical truths. Can Schurz's semantically isomorphic substitutivity replace US? Reasonably, it cannot without renouncing to define logical truths formally. To paraphrase Quine (1970: 50), Schurz's proposal to define logical truths in terms of formality should be something like: a logical truth is definable as a sentence from which we get only truths when we *isomorphically* substitute sentences for its simple sentences. Yet, since the notion of semantically isomorphic substitution is defined in terms of semantic modal notions, such as *semantical freedom*, *logical possibility*, *logical contingency* etc., logical truth cannot be formally captured in terms of semantically isomorphic substitutivity without circularity.[6]

[4] In informal words, by "semantical freedom of interpretations" with respect to a primitive sentence φ Schurz means the space of all possible variations of interpretation of φ. For example, a substitution function s from φ to $\varphi \wedge \neg\varphi$ produces a decrease in the semantical freedom of φ in that φ can be true or false, while $\varphi \wedge \neg\varphi$ can only be false.

[5] For more details see Schurz (2001: 371 et seq.)

[6] Before speaking about semantically isomorphic substitutions, the author introduces syntactically isomorphic substitutions, which "replace primitives [...] by other primitives in a unique way" (Schurz 2001: 370). Syntactically isomorphic substitu-

4.2. *McKeon's argument from analogy*

In this Section I will argue against what I have called McKeon's argument from analogy, according to which the notion of logical modality can be accommodated into Kripke semantics in virtue of an analogy between how first-order quantifiers and modal operators semantically work. Indeed, possible worlds semanticists usually interpret modal operators as quantifiers on possible worlds.

As we have seen in Section 2, first-order logical truths are said to be formal only if they are not context-sensitive. First-order logical truths context-insensitiveness entails their domain-insensitiveness: for any first-order sentence φ, for any ψ obtained from φ by (arbitrarily) restricting the quantifiers occurring in φ, if φ is logically true, so is ψ. The property of being context-insensitive can be considered in one of its consequences, that is: if φ is a first-order logical truth, φ is compatible, so to speak, with all possible domains of discourse.

According to McKeon, such property has a straightforward transposition within modal logics, in virtue of the analogy between how first-order quantifiers and modal operators work: just as first-order logical truths are such for any domain of discourse, so modal logical truths should be such for any logical space of logical possibilities, that is, for any set of logical possible worlds modal operators are restricted to. McKeon's argument from analogy rests on these premises. I will not proceed by questioning them, but rather by considering some of their consequences within a modal system representing logical modalities.

First-order domain-insensitiveness entails that for any first-order sentence φ, if φ is logically true, then the sentence of the form $\varphi \wedge \exists \boldsymbol{x} \forall \boldsymbol{y}(\boldsymbol{x} = \boldsymbol{y})$ will be logically possible. That sentence says that the following conjunction is true: that φ (e.g. "John passes the exam or John does not pass the exam"), and that there is exactly one object. If it were not, the sentence of the form $\neg \exists \boldsymbol{x} \forall \boldsymbol{y}(\boldsymbol{x} = \boldsymbol{y})$, or equivalently $\forall \boldsymbol{x} \exists \boldsymbol{y}(\boldsymbol{x} \neq \boldsymbol{y})$, which expresses the proposition that there are at least two things, would be a logical consequence of φ, whose truth would depend on the contingent fact that there is more than one thing. Accordingly, it comes to be essential to the logic of logical modality to assume the following meta-theorem: for any first-order sentence φ, if $L\varphi$ is true, so is $M(\varphi \wedge \exists \boldsymbol{x} \forall \boldsymbol{y}(\boldsymbol{x} = \boldsymbol{y}))$. At this point CI will require that the logical scheme $L\varphi \rightarrow M(\varphi \wedge \exists \boldsymbol{x} \forall \boldsymbol{y}(\boldsymbol{x} = \boldsymbol{y}))$ be valid for any first-order sentence φ, which is not the case in Kripke semantics under the logical reading of modal operators.

In fact, let φ be a first-order logical truth. Let $\langle W, @, R, V \rangle$ be a Kripke model $\mathfrak{M}$ with the following characteristics: $W=\{@\}$, $d_i(@) = \{o_1, o_2\}$. $L\varphi$ will be true in $\mathfrak{M}$, but $M(\varphi \wedge \exists \boldsymbol{x} \forall \boldsymbol{y}(\boldsymbol{x} = \boldsymbol{y}))$ will not, and that means: at

tions are defined purely syntactically, but they are only a proper subclass of semantically isomorphic substitutions. Thus what closure under syntactically isomorphic substitutions can at most do is to provide necessary but not sufficient conditions for the notion of logical truth.

the logically possible state of affairs represented by @, the first-order logical truth φ requires for its truth a domain of discourse consisting of more than one object.

We can conclude that McKeon's argument from analogy fails because it does not take into account the consequences of formalizing in the object-language (of the selected modal system) the first-order logical truths property of being context-insensitive, which is required by CI.

5. *Conclusion*

The formalization of logical modality in modal logical systems raises some serious problems for the formality of logic, when logical formality is syntactically understood in terms of uniform substitutivity. On the one hand, Carnap-style semantics violates the rule of uniform substitution. Schurz's proposal to redefine logical truth in terms of semantically isomorphic substitutions seems to be inappropriate, because of its circularity. On the other hand, Kripke semantics for the system S_5 is afflicted by inconsistency: in violation of what we have called "Condition of internalization" (CI), $\mathbf{S}_5$ is valid, but NLP is not a meta-theorem of S_5, since for any consistent but not valid first-order sentence φ, there is a counter Kripke model for $M\varphi$. In this paper, some reasons have been provided to extend the criticism of Kripke semantics for the system S_5 to every normal modal system aiming to represent logical modality. Such a system should contain both $\mathbf{S}_4$ (i.e. the internalization of N) and **US** (i.e. the internalization of US) as valid, but they are jointly inconsistent.

References

Beth, E. W. 1960. "Extension and Intension." *Synthese* 12 (4): 375–379.

Carnap, R. 1947. *Meaning and Necessity*, 4th edition. Chicago: The University of Chicago Press.

Cocchiarella, N. 1975. "On the primary and secondary semantics of logical necessity." *Journal of Philosophical Logic* 4: 313–317.

Hanson, W. H. and J. Hawthorne 1985. "Validity in Intensional Languages." *Notre Dame Journal of Formal Logic* 26 (1): 9–35.

Kripke, S. 1959. "A Completeness Theorem in Modal Logic." *The Journal of Symbolic Logic* 24 (1): 1–14.

Kripke, S. 1963. "Semantical considerations on modal logic." *Acta Philosophica Fennica* 16: 1–14.

Makinson, D. 1966. "How meaningful are modal operators." *Australasian Journal of Philosophy* 44 (3): 331–337.

McKeon, M. 2005. "A defense of the Kripkean account of logical truth in first-order modal logic." *Journal of Philosophical Logic* 34: 331–337.

Montague, R. 1974. "Logical Necessity, Physical Necessity, Ethics, and Quantifiers." IR. n Thomason (ed.). *Formal Philosophy*. New Haven and London: Yale University Press: 71–83.

Pollock, J.L. 1966. "Model Theory and Modal Logic." *Logique et Analyse* 9: 313–317.

Quine, W. V. O. 1970. *Philosophy of Logic*. Harvard: Harvard University Press.

Schurz, G. 2001. "Carnap's Modal Logic." In W. Stelzner and M. Stöckler (eds.). *Zwischen traditioneller und moderner Logik. Nichtklassische Ansätze*. Paderborn: Mentis: 365–380

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021
https://doi.org/10.52685/cjp.21.63.5
Received: July 8, 2021
Accepted: November 7, 2021

The Intuition behind the Non-Identity Problem

MATEJ SUŠNIK
Institute of Philosophy, Zagreb, Croatia

This paper examines a well-known non-identity case of a mother who chooses to conceive a blind child instead of a sighted one. While some people accept the non-identity argument and claim that we should reject the intuition that the mother's act is morally wrong, others hold onto that intuition and try to find a fault in the non-identity argument. This paper proposes a somewhat middle approach. It is argued that the conclusion of the non-identity argument is not necessarily in conflict with our intuitive response to this case.

Keywords: Blameworthiness; choice; non-identity argument; non-identity problem; permissibility.

1. *Introduction*

In non-identity cases, one's present actions determine the life's quality of people who do not yet exist. David Boonin (2014: 2), modifying the example proposed by Derek Parfit (1982: 118), thinks of a woman named Wilma who, due to her present health condition, faces a choice between conceiving a blind child now or conceiving a sighted child a couple of months later. Wilma finds it inconvenient to wait with conception until her condition is cured, so she decides to conceive now. Nine months later, she gives birth to a blind girl named Pebbles. This kind of case generates the non-identity problem (NIP): although most people have the intuition that it was wrong of Wilma to conceive Pebbles, it is difficult to explain what makes her act morally wrong.[1]

[1] Boonin's book presents the most extensive and detailed discussion of the non-identity problem in contemporary literature. Even though I focus on Boonin's central

That this intuition may not be firmly grounded is established by the non-identity argument (NIA).[2] While Pebbles's disability negatively affects her life's quality, it does not affect it to the extent that it makes her life not worth living. Also, had Wilma conceived later, she would have given birth to a different child. It follows that Pebbles could have either existed as blind or not existed at all. Therefore, Wilma's act of conceiving Pebbles did not make Pebbles worse off than she would have been, and hence it was not morally wrong of Wilma to bring her into existence.[3]

As Boonin remarks, the NIP arises only for those who find implausible the NIA's conclusion that "Wilma's act of conceiving Pebbles is not morally wrong" (2008: 130; 2014: 27).[4] As he points out, they can solve the NIP either by abandoning the intuition that gives rise to it or by rejecting some of the NIA premises.

While I share the doubt regarding that intuition's reliability, my reasons for such a doubt differ in one crucial respect. The reason why the intuition about the Wilma case should be rejected does not lie in the force of the NIA, but in the fact that it is mistakenly described as the belief that Wilma's act of conceiving Pebbles is morally wrong.

2. *Opposing intuitions*

Consider the following case:

> *Incompetent physician:* Susan wants to conceive a child and thus makes an appointment with a physician. Since the physician is incompetent, he does not inform Susan that her child will be born blind if she conceives now. He also does not inform her that this can be prevented if Susan postpones conception for a couple of

example of Wilma, the argument developed in this paper applies equally to the other cases of this sort, such as Gregory Kavka's "case of a slave child" (Kavka 1982: 100). It does not apply to other non-identity cases, such as those that involve a choice between two different social policies (for example, see Parfit 1984: 361–62).

[2] Boonin carefully distinguishes between the "non-identity problem" and the "non-identity argument". The non-identity argument, he notes, "gives rise to the non-identity problem" (Boonin 2014: 3).

[3] Boonin formulates the NIA in the following way:

"P1: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been

P2: If A's act harms B, then A's act makes B worse off than B would otherwise have been

P3: Wilma's act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles

P4: If an act does not harm anyone, then the act does not wrong anyone

P5: If an act does not wrong anyone, then the act is not morally wrong

C: Wilma's act of conceiving Pebbles is not morally wrong" (Boonin 2014: 27).

[4] For example, Wilma's act will probably not seem morally impermissible to those who believe that blindness and similar impairments are not bad. Moreover, some of those people may even think that such impairments are good. For example, see Cooper (2007).

> months until she cures her health condition by taking a certain medication. Not being aware of all this, Susan conceives now and nine months later gives birth to a blind girl named Kelly.

I presume that most people, confronted with this case, will not have the intuition that Susan did something morally wrong. However, this is puzzling. If our intuition in the case of Wilma is that her act of conceiving Pebbles was morally wrong, how is it that we do not have the same intuition in the case of Susan? Is the act of conceiving a blind child not the same in both cases?

Perhaps not. It could be replied that Wilma and Susan performed different act-types. More specifically, one could point out that our intuition in the case of Wilma is not that her act of *conceiving a blind child rather than a sighted one* has the property of being wrong, but that her act of *deliberately and knowingly conceiving a blind child rather than a sighted one* has the property of being wrong. Thus, it all depends on how the act-type in question is described.[5] Since the latter act-type is instantiated only by Wilma, the answer might go, there is nothing strange about our having different intuitions regarding the two cases.

However, here is where the trouble starts. This answer suggests that to account for the different moral reactions to the two cases, one must focus on the agent and the circumstances under which the act was performed. While Wilma's act was performed deliberately and in full knowledge of the outcome, Susan lacked the relevant information about her options.

But why would one feel that these features make Wilma's *act* of conceiving Pebbles morally wrong? To think so is to assume that what appear to be the properties of an agent (e.g., Wilma *knowingly* and *deliberately* conceiving Pebbles) can sometimes affect the rightness or wrongness of the performed act. Although something like this assumption has been endorsed by a number of authors—most notably in the context of the discussion about the relevance of intentions to moral permissibility—it has also been forcefully criticized (e.g., Thomson 1991; Scanlon 2008). And some of that criticism is that while the intentions with which an agent acts sometimes alter the nature of the act, they do not alter the *moral* nature of the act. For example, someone who deliberately and out of malice kills another person commits the act of murder. And the act of murder is distinguished from the act of manslaughter, which does not require the presence of a bad intention. Therefore, murder and manslaughter are different acts, but they are still both wrong (see, for example, Scanlon 2008: 12–13).

In that regard, it remains unclear why the fact that Wilma knowingly and deliberately conceived a blind child should make us react differently to the two cases. Perhaps we might think that Wilma's act of conceiving Pebbles is *worse* than Susan's act of conceiving Kelly, but

[5] For a discussion of the different ways in which act-types can be described, see Wedgwood (2011) and FitzPatrick (2012).

to say of two acts that one is worse than the other is not to say that one is wrong while the other is not.

In response to this, one could say that the accurate description of an act-type instantiated by Wilma should also include the reasons behind her choice. Then it would become apparent that it is the objectionable reason behind Wilma's choice that makes us respond differently to the two cases. Thus, while Susan instantiated an act-type *conceiving a blind child rather than a sighted one*, Wilma instantiated an act-type *deliberately and knowingly conceiving a blind child rather than a sighted one for the reasons of convenience*.

But this suggestion does not seem to avoid the problem mentioned above: it assumes that the reason for which Wilma performed her act has the power to affect its rightness or wrongness. However, it is difficult to explain how the reason behind Wilma's choice could possess such power. If it is plausible to say that the moral permissibility of an act is determined by its effects on others, endorsing the above suggestion would mean that one should explain how the reason for which Wilma conceived Pebbles negatively affected Pebbles. Furthermore, the NIA also stands in the way of accomplishing this task since it seems to establish that Pebbles is neither wronged nor harmed by Wilma's act.

The NIA applies to Susan's case just as it applies to Wilma's case.[6] It establishes that neither of them did anything morally wrong. However, the NIP arises only in Wilma's case. This is because our intuition about Susan's case is compatible with the conclusion of the NIA: her act *does not seem* morally wrong. But Wilma's act, on the other hand, *seems* morally wrong. This difference in our intuitive responses requires explanation.

In the remainder of this paper, I suggest that the problem disappears once we realize that our moral reaction to the Wilma case has nothing to do with the alleged wrongness of her act at all.

3. *Two assessments of Wilma's act*

When introduced to the Wilma case, most people will say that there is something morally objectionable about her act. But the term 'morally objectionable' seems to be ambiguous between 'morally wrong' and 'morally bad'. To take the former interpretation is to direct the discussion to the morally relevant features that make an action right or wrong. In particular, it is to focus on the (possible) harmful effects Wilma's act of conceiving Pebbles has (will have) on Pebbles.

But that is not the case if the term 'morally objectionable' is taken to mean 'morally bad'. Here one does not assess Wilma's act of conceiving Pebbles, but rather her acting on a *choice* to conceive Pebbles. Of course, her choice manifests itself in the act of conceiving Pebbles, but that does not mean that it cannot be assessed independently of that

[6] This can be easily seen if one replaces "Wilma" with "Susan" and "Pebbles" with "Kelly" in the NIA as it is formulated by Boonin.

act. Her acting on that choice reveals her judgment that conceiving a blind child was the best course of action in the given circumstances.[7]

With that in mind, we can focus all our attention on the reasons on which her judgment was based and assess her decision-making process. Such an assessment would essentially be backward-looking and as such should be distinguished from a forward-looking assessment that primarily focuses on the results of her choice. This latter assessment is, we might say, concerned with permissibility. And assuming that the NIA is sound, we should say that Wilma's act does not negatively affect Pebbles and is thus morally permissible. The former assessment, on the other hand, is not concerned with permissibility but rather with Wilma's subjective point of view and her deliberative process that resulted in her choosing to conceive a blind child. In that regard, one might say, the attribution of blame would be appropriate if her reasons for choosing to conceive a blind child turned out to be morally unacceptable. Thomas Scanlon describes the distinction between these two kinds of assessments in the following way:

> [I]t is the distinction between the permissibility of an action and a special kind of agent assessment, in which what is being assessed is not the agent's overall character but rather the quality of the particular piece of decision making that led to the action in question. (Scanlon 2008: 27–8)

Now, most people firmly hold onto their intuition that there is something morally objectionable about Wilma's act even after they learn about the NIA. Of course, it is quite possible that they are mistaken and that the explanation for this, as Boonin remarks, lies in our moral intuitions not being "sufficiently fine tuned to respond at every instance to what, at a purely intellectual level, we understand to be the case" (Boonin 2008: 148). But there is a simpler explanation available, and that is that our intuition in the Wilma case is not about the alleged wrongness of her act all.

In support of this suggestion, it should be noted that no argument has been provided in favor of the claim that our initial moral response to this case should be described as the belief that Wilma's act of conceiving Pebbles is morally wrong. Quite the contrary, this is merely assumed. And indeed, there are at least two reasons to think that this response should be described differently, namely as the belief that Wilma *acted badly*.[8] First, we would then be able to explain why we hold different intuitions about the cases of Wilma and Susan (Susan does not act badly). Second, we would also be able to explain why we cannot easily reject our intuition about the Wilma case despite the force of the NIA (the NIA is concerned with the permissibility of Wilma's act).[9]

[7] The relation between judgment and choice is discussed by Holton (2009).

[8] I borrow the phrase "acts badly" from FitzPatrick (2012). FitzPatrick develops a similar view as the one defended in this paper, but in an entirely different context.

[9] Perhaps it might be objected that interpreting our intuition in this way implausibly implies that one can be blameworthy even if one has not done anything wrong. But not everyone finds this implication controversial (see, e.g., Capes 2012; Scanlon 2008).

4. *Why Wilma acts badly*

But why think that Wilma acts badly when she chooses to conceive Pebbles? To answer this question, we should focus on her deliberative process and look deeper into the reasons behind her choice. And, as some authors have already pointed out (Wasserman 2019: 65–7), once we focus on the fact that her choice is exclusively determined by her convenience, there is room to argue that her act manifests insensitivity to the harmful condition of her future child.

Before Wilma has acted on her choice to conceive Pebbles, as David Wassermen notes, all she knew about her future child was that the time of his/her conception would determine whether he/she would be born blind or sighted. This point is crucial. For, as Wasserman convincingly argues, it enables us to see that, insofar as we consider things from Wilma's perspective, her choice came down to choosing, not between two different children, but between the condition of blindness and the condition of sightedness. And the moment Wilma acts on her choice to conceive Pebbles, her act signals her indifference and insensitivity to the vast difference between these two conditions (Wasserman 2019: 75–81).

However, Wasserman does not seem to realize that Wilma's insensitivity does not make her act wrong, but rather that it makes her act badly. And this allows Boonin to quickly dismiss the insensitivity charge as unfounded. In his recent paper on the NIP, Boonin offers three possible responses to the claim that Wilma's act manifests insensitivity to the condition of blindness (Boonin 2019: 146–147). But if we keep in mind that our intuitive reaction to the Wilma case is not about her act's wrongness, it becomes easier to see that none of them are successful.

Boonin first considers a case in which you are faced with a choice between saving a blind or saving a sighted child from drowning. Since the sighted child is further away, you find it more convenient to save the blind child. Now, it seems that most people would agree, Boonin says, that saving the blind child out of convenience in this case does not make you insensitive to the harm of blindness. And if so, why is the case of Wilma any different? (Boonin 2019: 146–147)

However, there is a crucial difference between the two cases. Namely, we do not believe that the badness of blindness is a reason not to choose to save the blind child from drowning, but we do believe that the badness of blindness is a good reason not to choose to bring a blind child into existence. This point has been first made by David Benatar, who points out that we do not apply the same standards when determining whether a "life is worth continuing" and whether a "life is worth creating" (Benatar 2006: 23). Thus, while acting out of convenience in the drowning case is not a sign of insensitivity, the same cannot be said about Wilma. Her finding it more convenient to conceive a blind child makes her insensitive to the forceful reason that speaks against such a choice.

There is a second way to rule out the insensitivity objection. Even if one grants that Wilma's act reveals insensitivity, Boonin remarks, "this would not show that her act was wrong, only that her motive was objectionable" (Boonin 2019: 146).

However, as it should already be clear, this does not present a problem for the view endorsed here. One could argue that Wilma acts badly and thus manifests insensitivity even if she has not done anything wrong. The key point here is that the judgment about Wilma's insensitivity is entirely independent of Pebbles coming into existence. Wilma's act, I suggest, would manifest an equal degree of insensitivity even if, by some coincidence, it did not result in the creation of a blind child. Although her acting on a choice to conceive Pebbles led to her actually conceiving Pebbles, our intuitive moral reaction is a reaction to the former, not the latter. And her merely acting on that choice makes it sufficient for us to say that her act manifests an inappropriate parental attitude to her future child's bad condition.

Boonin's third response to the insensitivity charge involves slightly changing the original example. He now asks us to imagine that Wilma is familiar with the NIA and finds it convincing. Since she now has a good reason to believe that Pebbles will neither be harmed nor wronged by being brought into existence, it cannot be correct to say that her acting on a choice to conceive Pebbles displays insensitivity (Boonin 2019: 146).

While this may seem like a promising route to take, I believe that the initial difficulty remains. The problem with this response is that it fails to allow for the possibility that Wilma's acting badly may be compatible with her act being morally permissible. The fact that Wilma is now aware of the considerations that justify her act only shows that she can defend it against the impermissibility charge. But it does not show that she is not insensitive if she performs it. As Joshua Gert notes in a somewhat different context, "[i]n many cases justifying considerations are ones that it would be morally better to *ignore*" (Gert 2004: 35).

Wilma's act, even if permissible, is in tension with our understanding of the nature of the relationship between parents and children. Following the lead of Scanlon, we might hold that such a relationship is, just like any other, "constituted by certain attitudes and dispositions" (Scanlon 2008: 131). And these attitudes and dispositions, Scanlon correctly points out, can be deemed more or less appropriate, depending on how far they are from the ideal standard. If it is plausible to say that prospective parents are, just like actual parents, expected to care for the welfare of their (future) children, then there is a reason to think that Wilma's act is not in line with such expectations. It is not enough to stipulate that Wilma believes that the NIA justifies her act in order to rule out the charge that she does not sufficiently care *about the condition* her future child will be born with.

5. *Conclusion*

Various solutions to the NIP have been proposed. Some people (David Boonin being the most prominent example) argue that the NIA forces us to abandon the intuition that Wilma's act is morally wrong. And those who, on the other hand, believe that Wilma's act is morally wrong try to find a fault in the NIA. I have proposed a somewhat middle approach to the NIP. What gives rise to the NIP (at least in cases like that of Wilma) is not the NIA but rather the way our intuitive moral response is described. After being introduced to the Wilma case, we do not form the belief that her act is morally wrong. Instead, we form the belief that Wilma acts badly. I conclude that even if the NIA is sound, it leaves our intuitive reaction to the Wilma case unscathed.[10]

References

Boonin, D. 2014. *The Non-Identity Problem and the Ethics of Future of People*. New York: Oxford University Press.

Boonin, D. 2019. "Solving the Non-Identity Problem: A Reply to Gardner, Kumar, Malek, Mulgan, Roberts and Wasserman." *Law, Ethics and Philosophy* 7: 127–156.

Benatar, D. 2006. *Better Never to Have Been. The Harm of Coming into Existence*. New York: Oxford University Press.

Capes, J. A. 2012. "Blameworthiness Without Wrongdoing." *Pacific Philosophical Quarterly* 93 (3): 417–437.

Cooper, R. 2007. "Can It Be a Good Thing to Be Deaf?" *Journal of Medicine and Philosophy* 32 (6): 563–583.

FitzPatrick, W. J. 2012. "Intention, Permissibility, and Double Effect." In M. Timmons (ed.) *Oxford Studies in Normative Ethics Vol. 2*. Oxford: Oxford University Press: 97–127.

Gert, J. 2004. *Brute Rationality: Normativity and Human Action*. Cambridge: Cambridge University Press.

Holton, R. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.

Kavka, G. 1982. "The Paradox of Future Individuals." *Philosophy and Public Affairs* 11 (2): 93–112.

Parfit, D. 1982. "Future Generations: Further Problems." *Philosophy and Public Affairs* 11 (2): 113–172.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Scanlon, T. M. 2008. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.

Thomson, J. J. 1991. "Self-Defense." *Philosophy and Public Affairs* 20 (4): 283–310.

Wasserman, D. 2019. "Chasing the Elusive Wrongdoing Intuition." *Law, Ethics and Philosophy* 7: 63–83.

Wedgwood, R. 2011. "Defending Double Effect." *Ratio* 4: 384–401.

[10] I thank the two anonymous reviewers for their helpful remarks. My thanks also go to Neven Sesardić who commented on an earlier version of the paper. This article is an output of the project "Harm, Intentions and Responsibility" supported by Croatian Science Foundation (Grant No. UIP-2017-05-4308).

Croatian Journal of Philosophy
Vol. XXI, No. 63, 2021

Book Reviews

Jonathan Gilmore, Apt Imaginings, Feelings for Fictions and Other Creatures of the Mind*, Oxford: Oxford University Press, 2020, 276 pp.*

In his award-winning new book, *Apt Imaginings, Feelings for Fictions and Other Creatures of the Mind*, Jonathan Gilmore (recently named co-editor of the *Journal of Aesthetics and Art Criticism*) tackles the following question: is the way that we engage with fictional objects analogues to the way we engage with objects in real-life? We enjoy *The Sopranos,* taking pleasure in Tony's evading the police and outsmarting other criminals, but in reality, we very much want dangerous criminals like Tony Soprano to face justice. Such invariance between our emotional responses and rational and moral evaluations across fictions and real life is defined as continuity question and it is the central problem of the book.

In the introduction, Gilmore discusses the notion of continuity in two distinct senses: descriptive and normative. In the descriptive sense, the question concerns whether our moral and rational engagements with fictions are the same (in the relevant explanatory way) as our moral and rational engagements with real people and real life. In the normative sense, we ask whether the kinds of reasons that we use to justify those engagements are the same i.e., invariant across fictions and real life. Gilmore's main point is to affirm the continuity thesis in a descriptive sense and to deny it in a normative sense, thus defending what he calls the discontinuity thesis.

In the second chapter, Gilmore sets the bedrock of his argumentation: a cognitive theory of imagination. He defines the concept of *imagination* as an irreducible mental state which is type-identified in virtue of its functional role, not the representational content. In other words, imagination, like other mental states, is not identified by its content—propositional or perceptive—but by the functional role it plays. According to Gilmore, there are good reasons to accept the cognitive theory of imagination. Firstly, imaginings are usually constrained by our will while other similar mental states are not. We can imagine a cold sunset on our favorite beach at will, but we cannot believe that the Earth is flat without good reasons or evidence. Secondly, beliefs are usually context-independent, while imaginings are not. Thirdly, beliefs aim at the truth, while imaginings have no such normative constraint. Having defined the concept of imaginings, Gilmore elaborates on the wide body of empirical evidence that points to the conclu-

sion that there is a continuity between the way we imagine visual imagery and the way we perceive visual imagery in real life. Gilmore concludes that imaginings are invariant across fictions and real life.

In the third chapter, Gilmore discusses another crucial aspect of his account: the evaluative theory of emotions. In his words: "…an emotion is not merely identified with a bodily feeling or behavioral tendency caused by an encounter with some object (event, state of affairs, and so on) but has an intentionality, an "aboutness," vis-à-vis the thing that elicits it. (…) the emotion instantiates an appraisal of the value (to oneself or what one cares about) of that object" (45).

For example, when I am afraid of a dog, I acknowledge that the dog is dangerous. When I am proud of my son's achievements in school, I appreciate him being curious and hard working. Gilmore presents a lot of empirical evidence which suggests that emotions that we feel for fictional characters are the same as the emotions that we feel for real people, primarily with respect to their neurophysiological basis. However, there are interesting asymmetries between the emotions we feel for fictional characters and the emotions we feel for real people. Our behavioural reactions to fictions are different from our reactions to comparable situations in real life. For example, when we encounter a frightening scene in a scary movie, we do not immediately start running from danger. Despite such asymmetries Gilmore concludes that emotions that we feel are the same kind of emotions in fiction and in real life—continuity thesis stands.

In chapters four to eight, Gilmore defends discontinuity thesis, extending his scope to different domains: affective responses, truth in fiction, conative engagement (tragedy and desire) and moral evaluations. In chapters four and five Gilmore argues that some of our affective responses to fiction have no equivalent to our responses in real life. For example, we can be moved by movie's score, poetry's rhyme or novel's writing style and these things can create mood and ambient unique to engagement with fiction.

In chapter six, Gilmore states that the way we discover what is true in fiction is analogues to how we discover what is true in real life: the same kind of deductive and inductive reasoning enables us to form predictions about the events. However, some inferences valid in fictions are invalid in real life. For example, in *Oliver Twist* we can infer from the fact that Fagin is physically grotesque, that he is morally corrupt.

In the following chapter Gilmore discusses the paradox of tragedy, which states that when we engage with works such as tragedy our pleasure seems to be internally related to our distress. After considering and rejecting three prominent accounts that offer a solution to the paradox of tragedy, Gilmore offers his own: the contradictory desires are rational in the light of a third—higher-order desire.

He offers the following argument:

"(1) a desire that a work be such that something, S, occurs in it;
(2) a desire that S not occur; and,
(3) a desire that one have both (1) and (2)" (173)

For example, in *King Lear*, we desire that a work be such that Cordelia dies at the end of the play, since such an ending is demanded by the genre. We

also desire that Cordelia does not die since we recognize her innocence and morality. Finally, we have a higher-order desire to have the first and the second desire, which is why we engage with tragedy. Gilmore affirms the discontinuity thesis with a claim that we engage rationally with tragedy in a way that has no equivalent in real life.

In the penultimate chapter, Gilmore argues that there is a discrepancy in how we morally value fictional characters and real people. His examples range from Dostojevski's Raskoljnikov and Shakespeare's Richard III to Tony Soprano, Hannibal Lector and Tom Ripley all the way to running over pedestrians in *Grand Theft Auto* and enjoying pornography. Gilmore explores theories of simulation, mirroring and contagion and concludes that discrepancy i.e., discontinuity holds. That being said, Gilmore's argumentation in this chapter seems incomplete and lacks the persuasive force present in his other chapters.

In the last chapter, Gilmore pursues separate issue, focusing shifting his intention to artistic function and moral value.

To conclude, this is an amazing book, brilliantly written and a joy to read. It is not a surprise that the book received a prestigious *Outstanding Monograph Prize* by The American Society for Aesthetics. Gilmore uses numerous empirical research to build up his arguments and defend his thesis. He works within the argumentative framework from philosophy of rationality and philosophy of mind, which strengthens his argumentation and brings a new and fresh outlook to the philosophy of art and aesthetics. The book will undoubtedly be of interest to anyone working in these domains.[1]

DAVID GRČKI
University of Rijeka, Rijeka, Croatia

[1] This work has been supported in part by Croatian Science Foundation under project number UIP-2020-02-1309.

Wolfgang Huemer and Ingrid Vendrell Ferran (eds.), Beauty: New Essays in Aesthetics and the Philosophy of Art, *München: Philosophia, 2019, 434 pp.*

Discussing the many complexities of beauty and demands that aesthetic theory of beauty should address, the great Roger Scruton wrote:

> We discern beauty in concrete objects and abstract ideas, in works of nature and works of art, in things, animals and people, in objects, qualities and actions. As the list expands to take in just about every ontological category (there are beautiful propositions as well as beautiful worlds, beautiful proofs as well as beautiful snails, even beautiful diseases and beautiful deaths), it becomes obvious that we are not describing a property like shape, size, or color, uncontroversialy present to all who can find their way around the physical world. For one thing: how could there be a single property exhibited by so many disparate types of things?[1]

Beauty, edited by Wolfgang Huemer and Ingrid Vendrell Ferran, addresses precisely the issues Scruton emphasizes as the most perplexing in relation to beauty. It provides new paths for philosophical explorations of beauty, extracting it from the traditional aesthetic theories and offering new perspectives on how it invades our lives when and where we least expect it. This is not to say that the (history of) aesthetics and aesthetic thinking about beauty are ignored. Quite the contrary, the introduction (and several chapters) offers a succinct but illuminative account of the development of philosophical understanding of beauty and the role beauty had in philosophical theories on (the value of) art. The focus here is on the shift that took place over the past century, in which beauty was dethroned from the aesthetic hierarchy. As the editors argue, "it is likely that we still lack the necessary historical distance to analyze [the reasons for such a change]" (8). That may well be the case, but *Beauty* certainly brings us a mile closer to appreciating how beauty is coming back into our philosophical exploration.

To be sure, beauty always was, and continues to be, "an anthropological constant of our human condition" (14), but within aesthetics, this focus came at the price of losing sight of other values, aesthetic and artistic alike. The reduction of art to beauty (and art theory to beauty theory) invoked a strong reaction not only among the philosophers, but among the artists alike. As the editors explain, within aesthetics, Jerome Stonitz's focus on aesthetic attitude was confronted by George Dickie who argued that aesthetic appreciation is not a matter of assuming a particular kind of attitude. In a similar vein, artists themselves broke free of the art for art's sake agenda. As evident by Dadaism, or abstract paintings, art was no longer at the service of creating beauty, but was dedicated to expressing moral, political and social ideas. The abandonment of beauty was further evident in development of institutional theories of art, and in creation of artworks which rejected harmony and symmetry.

Over the last couple of years several prominent books on beauty were published, bringing beauty slowly back to the philosophical and aesthetic spotlight. This is a valuable theoretical move, claim Huemer and Vendrell Ferran, but a care must be taken to avoid the "old trap of reductionism"

[1] Scruton, Roger, *Beauty*, Oxford: Oxford University Press, 2009:1.

(13). Instead, our contemporary theories should focus on the manner in which exploration of beauty is inseparable from the exploration of our cognitive, emotional and other psychological properties on the one hand, and, on the other, on other values that we acknowledge in our artistic engagements, such as works' cognitive, ethical or political value. In light with that, the editors intend their book to re-explore and re-evaluate the nature of beauty, so as to enable us to come up with a more profound understanding of the manner in which it impacts our lives and our artistic practices. It is my impression that they achieved this goal, and with flying colors.

Due to the lack of space, in what follows, I will provide only a rough sketch of the book, focusing more on broader theoretical concerns than on the details of argumentation developed in individual chapters.

Contributions by Sonia Sedivy, Hanne Appelqvist, Elisabeth Schellekens, Maria Elisabeth Reicher, Maria Jose Alcaraz Leon, Catrin Misselhorn and Otto Neumaier will be primarily relevant to those interested in the core aesthetic concerns, such as aesthetic judgments of beauty and taste, aesthetic properties, aesthetic emotions and the connection between beauty and emotions, the question of sensory as opposed to intelligible beauty, the domain of aesthetics and its relation to beauty, and the like. Prominent here are discussions of some of the leading aesthetic figures, such as Kant, Wittgenstein, Bell, Beardsley, Isenberg, Mothersill and Sibley, and the issues revolve around objectivity and subjectivity of aesthetic judgments, of passivity and activity of aesthetic experiences, of the parallels between aesthetic and perceptual judgments and the limits of aesthetics as a field of study. Elisabeth Schellekens points to the limits of the perceptual model of understanding aesthetics, most notably, its inability to account for intelligible beauty. She discusses three challenges to the notion of intelligible beauty, focusing her discussion on the relation between beauty and understanding. Analyzing how the sense of beauty is related to cognitive gains, she proposes that "aesthetic pleasure can occur … in cognitive process albeit not strictly in its resolution". This account of aesthetic pleasure captures "our intuitions about aesthetic delight as linked to the way in which it may generate new ideas and connections" (87).

Alcaraz Leon also voices a criticism of the analogy between aesthetic experience and perception. She expands the discussion of aesthetic judgments, suggesting they should be understood more broadly than allowed for by analogy between aesthetic experience and perception, making our aesthetic reactions passive and reactionary. Instead, we should recognize the "agential dimension of aesthetic judgments" and focus on "the idea that aesthetic judgment is a practical matter: something that we do!" (126). More to the point, Alcarez Leon emphasizes the fact that "there is a connection between aesthetic judgment and being a particular person, between our taste and our personality" (130); a connection which is extremely important in our lives, but lost, if we conceive of aesthetic judgments as solely a capacity to respond to certain aspects of our world.

Another traditional topic, the beauty of landscape, is addressed by Allen Carlson, who defends a position he calls cognitive landscape composition. On this view, the proper appreciation of landscape is available to appreciators who "must focus thoughtful contemplation on the cognitive resources

relevant to the composition of the landscape in question, which is knowledge about the particular origin and nature of the land from which it is composed" (343). In that sense, landscape is a "creation of human thought and imagination by which certain aspects of land are deemed salient and thereby given order, unity and coherence, and, by this mean, also given beauty" (347). A particularly interesting aspect of Carlson's essay is the account of the cultural landscape, i.e. heritage landscapes, the appreciation of which requires knowledge of anthropology, history, sociology, economics, architecture and history.

Contributions by Noël Carroll, Richard Eldridge, Davide Dal Sasso and Peter Lamarque discuss beauty in relation to art—the idea that beauty is central to art, its relation to conceptual, modern art, and poetry. Carroll first challenges the primacy of beauty in art, drawing on the historical examples spanning medieval to avant-garde periods, of artworks not designed to elicit the sense of beauty. What primarily interests him here is the role of beauty in criticism, given that "it makes no sense to bring it to bear upon a work of art that is legitimately intended (…) to oppose the pursuit of beauty for the sake of some conflicting intellectual and/or emotional purposes." (176). Central in this respect is his discussion (and modification) of Danto's theory, starting with Danto's emphasis on indiscernibles.

Danto is also lurking behind John Gibson's essay, dedicated to an exploration of the particular way in which some works of art function as a metaphor for life, whereby life is "transfigured" in the experience of the work. Though beauty is not Gibson's primary concern, the essay is illuminative in showing how our artistic experiences contribute to the sense of having understood the world better via artistic engagements. Grounded in aesthetic cognitivism (the view, roughly, that art gives us knowledge), and drawing on theories of metaphor, Gibson offers an account of how "art opens up a particular kind of window on the real, by providing a frame that transfers features of a work onto the aspects of the world that it casts as its subject." (302)

Lamarque brings the issue of beauty into discussions on poetry, with the aim of exploring the nature and role of aesthetic experience in response to a poem. As he explains, the central element of such an experience is appreciation, which is not reducible to the textual features of the poem. Rather, it is a trained response, which incorporates a "kind of attention to understanding" how the textual features are used to achieve certain artistic and aesthetic ends, and is concerned with the "pleasures of reading" (312).

Lisa Katharin Schmalzried focuses on the beauty of human beings. She analyses two conceptions: the characterological, on which one's beauty depends on one's physical appearance and on one's expressive features, i.e. "expressions of a person's character and mind" (353), and the dualist, according to which one's outer and inner beauty are mutually independent. This conception is Schmalzried's primary interest and she focuses on analyzing character traits and cognitive abilities underlying inner beauty. Relevant here is Plato-inspired virtue analysis, on which one is inwardly beautiful if one is virtuous, where the notion of being virtuous is grounded in Kant's account of moral duty and further modified in consequence to the analysis of Schiller's linking of one's virtuousness and inner beauty. In ad-

dition, taking inspiration from Aristotle, Schmalzried analyses the eudaimonist conception of inner beauty, which depends on one's intellectual and ethical virtues. The model she ends up defending equates inner beauty with relational virtuousness, grounded in Burke and Reid's accounts of inner beauty, which center on virtues that inspire love, affection and attraction.

A wider, social context within which issues of beauty arise is discussed by Stephen Davies, whose contribution focuses on the history and beautifying function of cosmetics in the context of sexual politics, social expectations, personal preferences and evolution. Informative on the cultural variations in the kinds of cosmetics and the manners of its production and use, the essay brings together two things we are "obsessed as species (…), adornment and decoration of ourselves, our possessions, our environment" (407). A wonderful achievement of Davies is revealing just how much influence these obsessions exert over our lives, in manners most often unthought-of and with consequences rarely considered.

To conclude. Informative, challenging and thought-provoking, *Beauty* is bound to expand philosophical discussion of beauty in directions rarely explored before in such depth and with such insightfulness. It will change our understanding of beauty and the value we attach to it, not only with respect to how beauty relates to other aesthetic categories we praise and cherish, but also with respect to emphasizing just how profoundly beauty, in its numerous instantiations, impacts all the aspects of our lives, society and environment. Rarely has beauty been discussed in relation to our ethical and epistemic agency within analytic philosophy, and rarely have these discussions managed to show the centrality of our aesthetic endeavors for who we are. Insights offered in individual chapters give more than 'a promise of happiness', as Nehamas might put it, in that they can be put to the service of making us more appreciative in our artistic endeavors, as both creators and appreciators, in helping us become better aesthetic agents, more responsive to the beauty around us and better equipped to use it for our sense of happiness and wellbeing. The book is a must-read for everyone interested in aesthetics and art, for everyone amazed by beauty and determined to keep it in sight.[2]

IRIS VIDMAR JOVANOVIĆ
University of Rijeka, Rijeka, Croatia

[2] This work has been supported by the University of Rijeka, UNIRI grant no. umjpo-20-2.

Ivan Cerovac, Epistemic Democracy and Political Legitimacy*, Cham: Palgrave Macmillan, 2020, 242 pp.*

Ivan Cerovac (2020) re-examines the source of democracy's potential to create legitimate decisions. He defends an epistemic conception of democratic legitimacy and raises questions such as whether political decisions can be false or true, wrong or correct, and whether certain individuals in a political community can know better than others what the correct political decision is.

Cerovac adheres to a non-monistic position that locates democracy's potential to produce legitimate decisions both in its moral and in its epistemic qualities. He improves the existing debate by providing a thorough analysis and an elaborate critique of current theories of epistemic democracy. In addition, the book discusses the suitable division of political and epistemic labor and presents epistemic arguments for property-owning democracy.

There are six chapters in the book. The book is divided into two main parts: the theoretical framework (the first five chapters) and the institutionalization of epistemic democracy (the final chapter). Cerovac begins the first chapter with Rawls' liberal principle of legitimacy, which states that one cannot be legitimately compelled to do something until sufficient reasons are provided, reasons that do not violate one's reasonable moral convictions. He continues with an epistemic account of political legitimacy that typically revolves around three tenets. The truth tenet says that there is truth in politics, the knowledge tenet says that there are experts in politics, and the authority tenet says that experts should rule. Cerovac's strategy is to endorse the first two, but to resist the third tenet.

Cerovac begins the topic in Chapter 2 by criticizing and rejecting the pure epistemic proceduralism approach, which is based on Fabienne Peter's notion that democratic decision-making procedures have the capacity to produce legitimacy due to some moral and intrinsic epistemic qualities. Cerovac, on the other hand, contends that instrumental epistemic value is required to analyze and enhance our epistemic practices. He then moves on to Thomas Christiano's pure deliberative proceduralism, in which Christiano claims that an instrumental representation of democratic legitimacy is impossible since it would need public agreement on the quality of results. Cerovac disagrees with Christiano, stating that when he promotes a discussion about aggregative democracy, that is, when he says that a state with more prosperity is preferable than a state with less welfare, he is using instrumental arguments.

In the third chapter, in the context of the knowledge tenet, Cerovac illustrates the difference between pragmatist deliberative democracy and second-personal epistemic democracy. Pragmatic deliberative democracy proponents say that political decisions can be correct or incorrect and that the evaluation system should be assessed on its ability to produce correct decisions. We should support debating processes in a liberal society that fosters freedom of thought, expression, and the press because the public vote is the best way to make the proper judgments. Cerovac opposes this viewpoint, stating that while it is successful in defending the epistemic value of public deliberation, it is not always successful in defending the epis-

temic value of democracy. Then he returns to Fabienne Peter's new theory, second-personal epistemic democracy, claiming that she bases her account on the concept of epistemic peerhood, where epistemic peers are defined as people who are equally likely to make the correct (or incorrect) judgment. Due to a lack of public consensus on who the experts are, he considers Peter's notion of epistemic equality unconvincing. Cerovac wraps off the third chapter by suggesting that the knowledge tenet be validated.

In the next chapter, he argues that epistocracy, as the rule of those who know, cannot meet the liberal legitimacy criteria. Cerovac believes that we can't expect all reasonable citizens to consider the same group of individuals as (moral) political experts, and that any group's authority would be rejected by at least some reasonable citizens. He then moves on to Mill's scholocracy, a method of decision-making in which everyone gets at least one vote, but those with greater education have several votes. Cerovac also opposes scholocracy, claiming that it is fair to expect certain epistemically harmful characteristics (biases) in a group that is granted more political power. After discussing epistocracy and scholocracy, Cerovac proposes that the authority tenet be rejected.

The fifth chapter refers to the theory of Marquis de Condorcet, who claims that aggregative democratic procedures have superior epistemic quality than their deliberative counterparts if all political decisions can be expressed as binary choices, citizens act and vote independently (no pre-voting deliberation), and citizens make decisions at least a little better than random procedures. Cerovac claims that none of the three conditions are met in contemporary societies because not all political decisions can be effectively expressed as binary choices, citizens are not independent (they receive information from the same sources), and citizens are worse than random procedures, at least in some political issues and situations.

Cerovac's uniqueness and the point of his primary argument are best represented in the book's last chapter. Cerovac depicts a democracy in which voters elect political representatives (legislative government) after determining that the candidates are experts. His main claim is that experts exist in the case of descriptive issues and natural sciences, and we can publicly appoint such experts because reasonable citizens can agree on who the experts in these areas are. On the other hand, in the field of value sciences (politics, morals, and ethics), we cannot publicly decide because we disagree about who the experts are, and Cerovac proposes democratic procedure as the most impartial form of procedure in which citizens elect their representatives, who they think are experts. It is crucial that the experts do not rule on the grounds of their expertise, but because they are authorized by a democratic procedure that is then publicly justified. Cerovac's key assertion is that, in the case of descriptive theses and natural sciences, experts exist, and we may select them publicly since reasonable individuals can agree on who these experts are. On the other hand, in the field of value sciences (politics, morals, and ethics), we cannot publicly decide because we do not agree about

who the experts are. Cerovac's thesis is based on the epistemic division of labor presented in the book. As a result, Cerovac proposes that the process should not be one-way: experts should be able to assist citizens in selecting achievable and coherent goals, and citizens should be able to assist experts in developing policies and making choices.

Cerovac's theory, in my opinion, may encounter some difficulties. The first challenge concerns obviously unreasonable attitudes in crises. Take, for example, the current pandemic and the antivaccination movement. Here I will use Ingrid Robeyns argument about the capability to be protected against infectious diseases. According to Robeyns, not only do you need access to a vaccination, but you also need enough other people to choose to get vaccinated, because protection needs a specific minimum number of individuals to be vaccinated. In other words, my capability to be protected from the debilitating effects of *COVID-19*[1] will depend on others' choice to exercise that capability (to get vaccinated). Due to the lack of addressing experts in crisis situations and relying on the media, there is a lack of reasoning in democratic procedures. In a situation where the only prevention or reduction of a pandemic is responsible human behavior and vaccination, growing democratic decision-making can lead to dangerous situations such as the greater and faster spread of the virus. I agree with Baccarini's conclusion that the scientific community's voices form a succession of valid public judgments based on the complicated application of scientific results and methodologies. As a result, it appears that voices advocating for theses that contradict scientific results and methodologies can be legitimately dismissed. Therefore, it seems that in a crisis, communication between experts and citizens, as proposed by Cerovac, is difficult to implement.

The second challenge concerns the field of art. When we look at instances of publicly popular art or works that are considered as established art in democratic decision-making, Cerovac's argument becomes difficult. In Croatia, there is a well-known case of the *Kiklop* literary award given to Nives Celzijus, a Croatian singer, and her book, *Naked Truth*. While, on the other hand, the reading of literary works that are considered literary classics, such as Miroslav Krleža's works, are less and less quoted. It is common for valuable works to go unnoticed, while popular works that do not have to be valuable will be affirmed in society.

Brian Barry's liberal-egalitarian argument can also be used. Barry thinks art is very valuable. The state should support higher levels of art and valuable things so that they can be available to everyone, not just the privileged elite. What is valuable due to the egalitarian principle of distribution should be promoted and made available for everyone. If art were completely left to the market, then quality art would be available only to those who could afford it (in the case of tickets for theatres and museums). Determining valuable art on the basis of a democratic public voice is problematic because situations often arise in which irrationality in the choice of political values and irrationality in works of art are combined. I will mention just a few examples of such cases. The first example is Bertolt Brecht, a German

[1] Robeyns does not mention COVID as an example. I only use her argument about the ability to vaccinate.

writer who advocated anti-militarism and anti-nationalism. In his most renowned plays, he emphasized his opposition to the National Socialist and Fascist. The following example is Zoran Žmirić's "Patient in Room 19," an anti-war novel about the Homeland War between Croatia and Serbia. The idea is that someone who is strongly right-wing will be an opponent of these writers' values, since they will believe that their art is not something that should be supported by the public in terms of funding their performances/work. Today's entire production is based solely on interest in economic goods, and the problem is that citizens do not have access to valuable content. Thus, examples show the poor outcome of leaving decision-making to unfettered value pluralism of the Cerovac type.

To conclude, I loved Cerovac's book and found it quite valuable because it provides a thorough examination of existing epistemic democracy ideas. Nonetheless, using the examples of irrational attitudes in crises and Brian Barry's egalitarian argument for encouraging higher value art, I intended to draw out some potential challenges to Cerovac's thesis on experts.

IVA MARTINIĆ
University of Rijeka, Rijeka, Croatia

Acknowledgement to Referees

The Editorial Board would like to thank our colleagues listed below who served as referees of manuscripts submitted to the *Croatian Journal of Philosophy* during the past two years.

Kenneth Aizawa, *Rutgers University, Newark, USA*
María José Alcaraz León, *Universidad de Murcia, Spain*
Adam Anderzeyewski, *University of Warsaw, Poland*
Zvonimir Anić, *Institute of Philosophy, Croatia*
Karan August, *TU Delft Architectural Theory Journal, Netherlands*
Carola Barbero, *Università di Torino, Italy*
Christopher Bartel, *Appalachian State University, USA*
Elvio Baccarini, *University of Rijeka, Croatia*
Boran Berčić, *University of Rijeka, Croatia*
Andrei Bespalov, *University of Warwick, Coventry, UK*
Majid D. Beni, *Nazarbayev University, Kazakhstan*
Giulia Bistagnino, *Università degli Studi di Milano, Italy*
Martina Blečić, *University of Rijeka, Croatia*
Sandy Boucher, *University of New England, USA*
Zdenka Brzović, *University of Rijeka, Croatia*
Stipe Buzar, *Libertas International University, Croatia*
Ivan Cerovac, *University of Rijeka, Croatia*
Anthony Cross, *Texas State University, USA*
Taylor Cyr, *Samford University, USA*
Chiara Destri, *Sciences Po, France*
Matthew DeCoursey, *Education University of Hong Kong, PRC*
Michael Devitt, *City University of New York, USA*
Joe Dewhurst, *University of Munich, Germany*
Laura T. Di Summa, *William Paterson University, USA*
Dušan Dožudić, *Institute of Philosophy, Croatia*
Víctor Durà-Vilà, *University of Leeds, UK*
Carlo Gabbani, *Liceo "XXV Aprile" Pontedera (Pisa), Italy*
Matjaž Gams, *Jožef Stefan Institute, Slovenija*
Molly Gardner, *Bowling Green State University, USA*
John Gibson, *University of Louisville, USA*
Paweł Grabarczyk, *IT University of Copenhagen, Danemark*
Filip Grgić, *Institute of Philosophy, Croatia*
Simone Grigoletto, *University of Padova, Italy*
Alex Jackson, *Boise State University, USA*
Daria Jadreškić, *University of Klagenfurt, Austria*

Tomislav Janović, *University of Zagreb, Croatia*
Tvrtko Jolić, *Institute of Philosophy, Croatia*
Dunja Jutronić, *University of Split, Croatia*
Marko Jurjako, *University of Rijeka, Croatia*
Srećko Kovač, *Institute of Philosophy, Croatia*
Marek Kuźniak, *University of Wrocław, Poland*
Davor Lauc, *University of Zagreb, Croatia*
Ronald Loeffler, *Grand Valley State University, USA*
Thornton Lockwood, *Quinnipiac University, USA*
Erik Magnusson, *University of Manitoba, Canada*
P. D. Magnus, *University at Albany, State University of New York*
Derek Matravers, *The Open University, UK*
Jorge Alberto Manero Orozco, *Czech Academy of Sciences, Czech Republic*
Luka Malatesti, *University of Rijeka, Croatia*
Per Milam, *University of Gothenburg, Sweeden*
Dana Miller, *Fordham University, USA*
Nenad Miščević, *University of Maribor, Slovenia*
Takeshi Nakamura, *Osaka University of Health and Sport Science, Japan*
Dawa Ometto, *University of Leipzig, Germany*
Deborah Orr, *York University, Canada*
Gianfranco Pellegrino, *Luiss University, Italy*
Héctor J. Pérez, *Universidad Politécnica de Valencia, Spain*
Maël Pégny, *CNRS, Paris, France*
Timothy Perrine, *Wuhan University, PRC*
Nicola Perullo, *University of Gastronomic Sciences in Pollenzo, Italy*
Gualtiero Piccinini, *University of Missouri, St. Louis, USA*
Adam C. Podlaskowski, *Fairmont State University, USA*
Kalle Puolakka, *University of Helsinki, Finland*
Aleta Quinn, *University of Idaho, USA*
Joel Michael Reynolds, *University of Massachusetts Lowell, USA*
Ivan Restović, *Institute of Philosophy, Zagreb*
Sarah Roberts-Cady, *Fort Lewis College, USA*
Mathew Schrepfer, *Brown University, USA*
Severin Schroeder, *University of Reading, UK*
Joachim Schulte, *University of Zurich, Switzerland*
Merel Semeijn, *University of Groningen, Netherlands*
Nenad Smokrović, *University of Rijeka, Croatia*
Stephen Snyder, *Bogazici University, Turkey*
Cristian Soto, *Universidad de Chile, Chile*
Uwe Steinhoff, *University of Hong Kong, PRC*
Goran Sunajko, *University of Zagreb, Croatia*
Gabriela Świtek, *University of Warsaw, Poland*
Antonin Thuns, *Université Libre de Bruxelles, Belgium*
Connell Vaughan, *TU Dublin School of Creative Arts, Ireland*

Iris Vidmar Jovanović, *University of Rijeka, Croatia*
Christopher Vogel, *Shandong University, PRC*
Alberto Voltolini, *University of Turin, Italy*
George Wrisley, *University of North Georgia, USA*
James Young, *University of Victoria, Canada*
Nebojša Zelič, *University of Rijeka, Croatia*

Table of Contents of Vol. XXI

Articles

Book Reviews

Croatian Journal of Philosophy is published three times a year.
It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is indexed in
The Philosopher's Index, PhilPapers, Scopus, ERIH PLUS
and in *Arts & Humanities Citation Index* (*Web of Science*).

Payment may be made by bank transfer

SWIFT PBZGHR2X
IBAN HR4723400091100096268

Croatian Journal of Philosophy is published with the support
of the Ministry of Science and Education
of the Republic of Croatia.

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, A. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, T. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

The publication of a manuscript in the Croatian Journal of Philosophy is expected to follow standards of ethical behavior for all parties involved in the publishing process: authors, editors, and reviewers. The journal follows the principles of the Committee on Publication Ethics (https://publicationethics.org/resources/flowcharts).

ISSN 1333-1108
9 771333 110001