dialectica

International Journal of Philosophy

The Formalization of Arguments

edited by Robert Michels

Contents

European Society for Analytic Philosophy

Vol 74 : 2

June 2020

dialectica

International Journal of Philosophy
Official Organ of the European Society of Analytic Philosophy

founded in 1947 by Gaston Bachelard, Paul Bernays and Ferdinand Gonseth

dialectica vol 74 issue 2

The Formalization of Arguments

edited by Robert Michels

June 2020

Contents

The Formalization of Arguments

An Overview

Robert Michels

The purpose of this introduction is to give a rough overview of the discussion of the formalization of arguments, focusing on deductive arguments. The discussion is structured around four important junctions: i) the notion of *support*, which captures the relation between the conclusion and premises of an argument, ii) the choice of a *formal language* into which the argument is translated in order to make it amenable to evaluation via formal methods, iii) the question of *quality criteria* for such formalizations, and finally iv) the *choice of the underlying logic*. This introductory discussion is supplemented by a brief description of the genesis of the special issue, acknowledgements, and summaries of each article.

1 The Formalization of Arguments

An argument in the philosophical sense is a set of sentences consisting of (at least)[1] one sentence stating a conclusion and (at least) one sentence stating a premise which is or are supposed to support the conclusion.[2] Arguments are of central importance to philosophy not only as a subject of systematic study, but also methodologically as the means to criticise or support philosophical claims and theories. More generally, arguments are an indispensable part of

1 Most arguments discussed by philosophers involve only one conclusion and some have argued against admitting multiple conclusions (see e.g. Steinberger 2011), but there are systematic developments of multiple conclusion logics. See e.g. Shoesmith and Smiley (1978). For the sake of simplicity, I will focus on single conclusion arguments throughout most of this text.

2 Note that throughout this paper I will mostly refer to natural language sentences instead of e.g. utterances of them. I will ignore related metaphysical questions including e.g. questions about what sentences are or about propositions and their relation to natural language sentences and sentences of formal languages. The focus on sentences is both in line with at least significant parts of the literature on formalization and moreover also serves to simplify and homogenize the discussion of different views. I hope that the presentational advantages outweigh the costs of imprecision and a sometimes dangerously liberal use of the term "sentence."

any responsible rational discourse; to give an argument for a claim is to give a reason for it and to set out this reason for oneself and for others to scrutinize.

The analysis, development, and critique of arguments are some of the most important tasks performed by contemporary philosophers working in the analytic tradition. The process of formalization is an important step in any one of these tasks since it makes arguments amenable to the application of formal methods, such as those of model theory or of proof theory. These methods give us precise and objective quality-criteria for arguments, including in particular criteria for their logical validity.

Assuming that we have identified the premises and conclusion of an argument, its formalization will require us to make a number of choices, including those captured by the following four interrelated questions:

1. Which kind of inferential support do the premises lend to the conclusion of the argument?
2. Into which formal language should we translate the argument's premises and conclusion?
3. What makes such a translation into a particular formal language adequate?
4. Which formalisms can be used to evaluate the quality of the argument?

The remainder of this introduction is structured around these four questions about the formalization of arguments. It starts out with a brief discussion of each of these questions in the following four sections, briefly discussing some answers given in the literature and providing some references for further reading. The main aim of this introductory part of this paper is to give readers who are not familiar with the relevant literature a partial look at the more general discussion to which the papers collected in this special issue contribute. This overview is neither comprehensive, nor authoritative. The last two sections of the introduction contain some information about the genesis of the special issue and the editor's acknowledgements and a brief overview of the content of the papers published in this special issue.

2 Inferential Support

A standard classification of arguments individuates kinds of arguments in terms of the kind of *inferential support* which its premises lend to an argument's conclusion. We may accordingly distinguish between, among others, abductive, statistical, inductive, deductive arguments and arguments from analogy. The sort of arguments we encounter in everyday life, e.g. in discussions with neighbours and friends or in political debates, rarely fit into just one of these categories. Rather, they might consist, for example, of an abductive argument for a conclusion which in turn serves as a premise among others in a deductive argument, whose conclusion in turn is used to argue for another claim by analogy, and so on. They may of course also involve particular forms of reasoning which do not neatly fit into the classificatory scheme which one finds in philosophy books, e.g. because they draw on particular non-verbal aspects of a particular discussion, or positively contribute to a debate in a particular context, even though they have the form of a logical fallacy (e.g. an appeal to authority). One might hence argue that theoretical engagement with "real world" arguments require different, perhaps more permissive approaches than those covered in introductory books and courses on logic and critical thinking.[3] Still, many such arguments, or at least parts of them, can be broken down into smaller segments which exemplify one of the canonical argument types.

Deductive arguments enjoy a special status in philosophy due to the particularly strict way in which the premises of a deductive argument supports its conclusion. Consider for example the following argument:

(1) If the train runs late, its passengers will miss their connections.
(2) The train runs late.
(3) ∴ Its passengers will miss their connections.

The conclusion of this argument, which in schemas of this sort will be marked by the prefixed symbol "∴" throughout this text, like that of any valid deductive argument, is logically entailed by its premisses. But what is logical entailment? In contemporary logic, there are two fundamental accounts of what it means for a sentence to be logically entailed by another. The first is the syntactic account which characterizes logical entailment proof-theoretically in terms of derivability or provability in a logical system. Considering the formal language of first-order logic, the core idea of this account is that a sentence s of language is logically entailed by a set of sentences Δ of the same language if, and only if, there is a proof of s which can be constructed in a formal calculus, e.g. using the introduction- and elimination-rules of the logical constants in case of the natural deduction calculus, and taking at most the sentences in Δ as

3 See e.g. Betz (2010, 2013). See also Groarke (2021) for an overview of the field of informal logic.

doi: 10.48106/dial.v74.i2

hypotheses.[4] The second account is the semantic account, which characterizes entailment in model-theoretic terms. Its core idea is that, focusing again on the language of first order logic, a sentence s (i.e. a well-formed formula of that formal language) is logically entailed by a set of sentences Δ if, and only if, for all models $\mathfrak{M}$ for this language, if all sentences in Δ are true in $\mathfrak{M}$, s is true in $\mathfrak{M}$, where a model is a set-theoretical construction used to semantically interpret all well-formed sentences of the language.[5] As is well-known, the two relations characterized by these accounts coincide for sound and complete logics, such as classical first-order logic, in the sense that they render exactly the same entailments valid. The term "logical consequence" is usually reserved for the latter, semantic notion and I will follow this convention in the remainder of this section.

It is important to distinguish the question of the validity of an argument from that of its *soundness*. An argument is sound if, and only if, it is both valid, i.e. if its conclusion is logically entailed by its premises, and if its premises are true. Neither the proof-theoretic, nor the model-theoretic approach just described is concerned with the truth of an argument's premises. Both approaches target the notion of validity.

The proof-theoretic characterization of deductive entailment is intrinsically linked to particular formal systems which characterize logical expressions like that of negation, conjunction, or the quantifiers in terms of introduction- and elimination-rules which tell us under which conditions we can either introduce or eliminate formulas containing such an expression in the context of a proof. The totality of these rules fix what is provable in such a system and a fortiori give us the sort of syntactic characterization of logical entailment which interests us in the current context. One important philosophical question about introduction- and elimination-rules in a formal system concerns the relation between the two kinds of rules. It was forcefully raised in Prior (1960), who argued against the idea that the meaning of logical expressions is completely fixed by their introduction- and elimination-rules by introducing the connective "tonk" whose associated pair of rules permit us to derive absolutely any sentence from any sentence. An influential idea for how the problem raised by "tonk" and similarly problematic connectives can be avoided is that such connectives violate a harmony-constraint which is

4 The two standard systems in the contemporary discussion (natural deduction and the sequent calculus) were introduced in Gentzen (1935); see Plato (2014);Schröder-Heister (2018) for more general introductions to proof-theory.

5 The key historical text is Tarski (2002); see Beall, Restall and Sagi (2019) for an introduction.

supposed to govern the relation between a logical expression's introduction- and its elimination-rules.[6] But even if it turned out that such a constraint can be formulated, Prior's argument could still be taken to show that, as Prawitz puts it, "ordinary proof theory has nothing to offer an analysis of logical consequence" (2005, 683).[7] A suitable notion of harmony may give us a way of guarding a formal system against incoherence and a fortiori allow us to accept its harmonious introduction- and elimination-rules as constitutive of the meaning of its logical expressions within that system. Even so, there still would remain an explanatory gap between a formal-system-relative harmonious notion of provability and the general, formal-system-independent notion of logical consequence. One proposal for a way to close this gap is due to Dummett and Prawitz, who argue that logical consequence can be characterized using proof-theoretic means and the notion of canonical proof (see e.g. Dummett 1976; Prawitz 1974, 2005).

Concerning the semantic characterization, many contributors to the recent literature have focused on two different properties which might be used to characterize or define logical consequence, that of being *necessarily truth-preserving* and that of being *formal*.

That logical consequence is closely linked to necessity is a well-established idea in analytic philosophy.[8] In the contemporary debate, this connection is usually spelled out in terms of necessary truth-preservation: If a sentence *s* is a logical consequence of a set of sentences Γ, then it is necessary that if the sentences in Γ are true, so is *s*. Or, to put it differently, it is impossible for the sentences in Γ to be true, but for *s* not to be.

The property of being necessarily truth preserving distinguishes deductive from inductive arguments, such as the following:

(4) Every dog which has been observed up until now likes to chase cats.
(5) Bella is a dog.
(6) ∴ Bella is a dog who likes to chase cats.

6 See e.g. Dummett (1991, ch. 9), and Tennant (1987), Steinberger (2011), and for a recent criticism, Rumfitt (2017).

7 This quote echoes the approach taken by Tarski (1956b, 412f), and followed by many contributors to the recent literature, who motivates his semantic definition of logical consequence by arguing against the syntactic approach.

8 See e.g. Wittgenstein's claim that deductive inferences have an inner necessity in §5.1362 of his *Tractatus* (1922).

doi: 10.48106/dial.v74.i2

Clearly, the fact that every dog observed up until now likes to chase cats does not guarantee that absolutely every dog, including (possibly unobserved) Bella, likes to chase cats. The truth of the premises of this argument, and of those of any inductive argument in general, does not necessitate the truth of its conclusion.[9] The focus of this special issue and of the following parts of this introduction is on deductive arguments.

While necessary truth preservation plausibly gives us a necessary condition for an argument's being deductive, i.e. for its conclusion to be a logical consequence of its premises, there are reasons to doubt that the notion of logical consequence can be adequately explained, characterized, or defined in terms of this property. An important open question in this regard is what kind of necessity the property of necessarily preserving truth involves. The seemingly obvious claim that it is the notion of logical necessity would lead us into an explanatory circle, since logical necessity is plausibly explainable in terms of logical consequence. It is furthermore not clear whether other kinds of necessity, such as for example analyticity, a priority, or metaphysical necessity, can serve this purpose (see Beall, Restall and Sagi 2019, sec. 1).

The second property which is much discussed in the literature on logical consequence is the notion's *formality*. Intuitively speaking, this property distinguishes logical inferences from material entailments such as:

(7) The ball is red.
(8) ∴ The ball is coloured.

Or:

(9) Some dog sees some cat.
(10) ∴ Some cat is seen by some dog.

While these arguments reflect intuitively correct inferences, their conclusions are not logical consequences of their premises. This is because the entailments from (7) to (8) and from (9) to (10) obtain due to the material content of these sentences, i.e. due to what the sentences are about, not due to their form: That (8) is entailed by (7) is guaranteed by the meanings of "is red" and of "is

9 Since both arguments by analogy and statistical arguments can be considered special kinds of inductive arguments (see Salmon 1963, ch. 3), the same holds for them. Abductive arguments also fail to be necessarily truth-preserving, but it can be argued that abduction is not just a special case of induction (see Douven 2021).

coloured" and that (10) is entailed by (9) is guaranteed by the meanings of "sees" and "is seen by."

The validity of a deductive argument in contrast depends solely on the logical form of its premises and conclusion.[10] The logical form of a sentence in turn is determined by the logical expressions it contains and the way they combine with the contained non-logical expressions. That deductive logic is formal in this sense is uncontroversial, but it is hard to say what "formal" means without just defining it ostensively by referring to examples of sentences which we assume to share the same logical form. Can we define the notion of formality in other terms, giving us a systematic criterion to distinguish between the logical and the non-logical expressions of a language? There are several answers to this question two of which will now be briefly introduced.[11] Before this is done, it should be noted that while the focus in the current section is on the notion of logical consequence, most of the discussion of formality focuses on the use of this notion to distinguish logical from non-logical expressions of languages.[12] There is a direct connection between these two loci of formality, since the logical expressions in a sentence determine its logical form and it is in turn the logical form of sentences which ensure that they stand in the relation of logical consequence.[13]

One approach to formality proposed in the literature says that formality can be understood in terms of topic neutrality (see e.g. Ryle 1954, 115ff; Haack 1978, 5–6). The idea is that logical entailments hold irrespective of what the entailed and the entailing sentences are about. What distinguishes the logical expressions of a language is that they, unlike predicates like "is red" and "is coloured" or individual constants, are not about any thing in particular, but that their meaning is rather tied to certain schematic patterns of application which are universally applicable. This criterion for formality gives us a simple and plausible explanation of why the entailment from (7) to (8) is not formal and thus not logical. The main problem noted even by those like Haack who

10 While this clearly holds for the notion of validity one gets e.g. from classical first-order logic, one might see relevance (also: *relevant*) logic as an exception. The core idea of relevance logic is that certain intuitively paradoxical inferences, which are valid in classical logic, can be ruled out as invalid by imposing a relevance constraint to the effect that the conclusion of an argument (or the consequent of a conditional) should not be on a different topic than its premises (the conditional's antecedent). This constraint is however implemented via a formal principle. See Mares (2020) for an overview.

11 For discussions of further answers, see e.g. MacFarlane (2000), Dutilh Novaes (2011).

12 See e.g. Tarski (1986), Sher (1991), Bonnay (2008).

13 See, however, Sagi (2014) for an alternative view.

are sympathetic to it is that topic neutrality only gives us a vague criterion for demarcating logical from non-logical expressions: Why could we for example not count the inference from (9) to (10) as formal? After all, it might appear that we can extract a schematic pattern of the following form from this entailment:

(11) x Φs y.
(12) ∴ y is Φed by x.

Putting complications about surface grammar aside which the schema ignores (e.g. "sees" and "is se*en* by"), one may on the one hand argue against its formality by pointing out that the correctness of the inference seems to depend on the seemingly material fact that "Φs" and "is Φed by" are converse relations. On the other hand, one might argue that the two converses are really identical (see Williamson 1985) and then claim that (11) and (12) are just the same sentence in different guises. After stripping away these guises, the inference would really just be a trivial inference from one sentence to itself, instantiating an inference schema which holds irrespective of what the sentence involved means. The point here is of course only that as a criterion for logicality, topic neutrality leaves room for disagreement about particular cases, giving us at best a vague account of what formality is.

The second account of formality is provided by Tarski's classical permutation-invariance-based characterization of logicality (see 1986). This account could be seen as a way to make the topic-neutrality-based account of formality more precise. Its core idea is that the distinguishing feature of logical expressions is that their meaning is invariant under all permutations of the domain of objects of a model. A *model* in the model-theoretic sense is a set-theoretical construction based on a domain of objects which is designed to enable us to semantically interpret sentences of a formal language in set-theoretic terms with respect to that domain. A *permutation of the domain of a model* is a function which maps each object in that domain to a unique object from the same domain. Within a model, first-order predicates can e.g. be interpreted as sets of objects and first-order relational predicates accordingly as sets of tuples of objects. Logical expressions are also given a set-theoretic interpretation, so that first-order quantifiers can e.g. be interpreted in terms of relations between predicates, i.e. sets of tuples of sets of objects. The sets corresponding to material predicates in a model, such as e.g. the relational predicate "is larger than" in a model which is used to interpret a fragment of natural language involving the predicate, vary under at least some permutations of a model's domain. There will e.g. be a permutation which maps two objects *a* and *b* which stand in this relation to other objects from the domain which do not (e.g. simply to *b* and *a*, respectively). The idea underlying Tarski's characterization is that no such thing can happen to logical expressions; the logical expressions retain their intended meaning in a model, no matter under which permutation of the objects in the model's domain we consider them.[14]

One of the main questions about the notion of logical consequence is how the precise, model-theoretic notion relates to the intuitive, pre-theoretical notion of logical entailment with which we operate in ordinary reasoning. The idea that the former can be extracted from natural language, and in particular Glanzberg's recent critique of this idea, are discussed in Gil Sagi's contribution to the special issue.

That there is an explanatory gap to be filled here has already been pointed out by Tarski, who writes that

> the concept of following is not distinguished from other concepts of everyday language by a clearer content or more precisely delineated denotation [...] and one has to reconcile oneself in advance to the fact that every precise definition of the concept [...] will to a greater or lesser degree bear the mark of arbitrariness. (2002, 176)

An influential contribution to the debate about logical consequence which takes this question as its starting point is Etchemendy (1990). Roughly, Etchemendy argues that Tarski's model-theoretic definition of logical consequence fails to capture the intuitive notion of logical consequence, since it presupposes certain contingent, non-logical assumptions about the cardinality of the universe, putting the notion defined by Tarski at odds with the necessity of the intuitive notion.[15]

14 For a more precise explanation of the criterion, see MacFarlane (2015, sec. 5) and Bonnay (2014) for an overview of recent work on it. An influential line of objection to invariance-based characterizations of logical constants can for example be traced through Hanson (1997), McCarthy (1981), McGee (1996), Sagi (2015), and Zinke (2018b).

15 See Caret and Hjortland (2015, 5f) and Zinke (2018a, sec. 5.3) and see Zinke (2018a, sec. 5.1) for a different argument along similar lines.

doi: 10.48106/dial.v74.i2

3 Formal Languages

There are different formal methods which one can apply to evaluate the logical validity of an argument. One may for example rely on semantic methods, such as those provided by a model theoretic semantics, or on syntactical methods, such as the one provided by the natural deduction calculus.[16] In order to apply such formal methods to systematically assess the quality of an argument, the premises and conclusions of arguments have to be translated from the natural language in which they are stated into a suitable formal language. The process of translating a sentence of a natural language into a formal language is the process of formalizing in the narrow sense, as opposed to the wider sense which pertains to whole arguments.

Besides this central technical reason, there are further reasons for formalizing arguments. One important reason is that given a suitable formal language, formalizing an argument forces us to clarify, in different respects, its premises and conclusion. One respect of clarification concerns the many ambiguities present in natural language. Formal languages are often explicitly constructed to be unambiguous, so that each sentence (or formula, if one prefers) of the language is assigned a single, precise meaning. A well-worn example are ambiguous natural language sentences involving quantifier phrases such as "Every child gets a present." Translating the sentence into the formal language of first-order logic, we are forced to decide between two unambiguous readings of the sentence (that every child gets its own present(s) or that every child gets the same present(s)) by the variable-binding structure of the quantifiers of the formal language. Dutilh Novaes (2012, ch. 4 and 7), furthermore argues that there is another respect in which formalization helps us clarify the formalized parts of language, namely that formal languages serve to eliminate certain cognitive biases.

From the perspective of logic, formal languages are first and foremost mathematical objects.[17] More specifically, they are identified with sets of formulas, where a formula is a sequence of symbols which is generated from a set of symbols, the formal language's alphabet, based on a set of syntactic rules which give us a recipe for generating all well-formed formulas of the respective

16 That logic can help us decide on the validity of an argument formulated in a natural language is a standard assumption. It is however challenged by Baumgartner and Lampert (2008), who argue that the formalization of an argument should rather be understood as a means to explicate the argument by bringing out the formal structure on which the natural language argument is based.

17 But see Dutilh Novaes (2012, ch. 2) for discussion.

language. The resulting formal language is of course still devoid of meaning, as it merely gives us an alphabet of symbols and rules for constructing certain sequences of them. To interpret the language, a semantics which defines meanings for all well-formed formulas of the language is needed. The standard approach is to identify these meanings with truth-values, reflecting the idea that semantics is about true or false representation of an underlying structure which the sentences of a language reflect or fail to reflect. But there is also an inferentialist tradition which aims to characterize meaning in terms of the inferential rules which govern the expressions of the language.[18]

Formal languages and their semantic interpretations are legion, but what constrains our choice of a formal language when formalizing an argument? This section will focus on one rather important constraint, namely the expressive strength of the formal language. General philosophical constraints about the notions involved in an argument one wants to formalize or pragmatic or sociological constraints tied to certain context will hence not be discussed.

The notion of expressive strength is a semantic notion which concerns not only an uninterpreted formal language, but rather a pairing of such a language with a suitable semantics. It seems that, at least in some cases, there is a notable asymmetry in the relation between the language and the semantics when it comes to determining expressive strength: We cannot extend the expressive strength of some language beyond a certain threshold set by the expressions it contains by coupling it with a different semantics. An example is the language of propositional logic which simply lacks the syntactic expressions needed to capture the inner logical structure of atomic formulas which grounds the felicity of certain inferences which come out as valid in classical first-order logic. One could try to compensate for the lack of syntactic structure by adopting a particular translation scheme and by encoding the validity of the logically invalid inferences in the semantics. E.g. if the predicate "F" stands for "is a dog" and "G" for "is an animal," then the valid first-order inference from "$\forall x(Fx \to Gx)$" and "$\exists xFx$" to "$\exists xGx$" could be simulated in the language of propositional logic by assigning a propositional constant to the English sentences "All dogs are animals," "There is a dog," and "There is an animal" and by building it into one's semantics of the language of propositional logic that the two first entail the third. But there are obvious limits to this strategy, since it e.g. makes the semantics depend on a particular translation-schema from a natural into the formal language and since it would make it a matter

18 See e.g. Sellars (1953), Brandom (1994), Peregrin (2014).

doi: 10.48106/dial.v74.i2

of stipulation which propositional constants express logical truths or stand in relations of logical entailment.

In order to allow us to adequately formalize an argument, the formal language (together with a suitable semantic interpretation), has to be able to capture enough of the logical structure of the argument as stated in a natural language to make it an argument, i.e. a collection of sentences one of which stands in a relation of inferential support to the others. Intensional logic offers a wealth of examples which highlight expressive limitations of certain formal languages. A classical example from tense logic concerns the formalization of the sentence (see e.g. Cresswell 1990, 18):

(13) One day all persons now alive will be dead.

In the language of a simple tense logic which extends the language of first-order logic with the sentential tense-operators **P** ("It was the case that…") and **F** ("It will be the case that…"), if one uses the predicates A, D for "… is alive" and "… is dead" respectively, the closest one can get to an adequate formalization of (13) is:

(14) $\mathbf{F}\forall x(Ax \rightarrow Dx)$

Since this formula says that it will be the case at a future time that everyone alive at that time is dead at that time, this translation is clearly inadequate. There are different ways to remedy this lack of expressive strength. One is to add a sentential "now"-operator **N** and to introduce a double-indexed semantics for the language which allows one to evaluate formulas relative to not one but two time indices, one of which is specifies the time of evaluation.[19] Figuratively speaking, **N**'s semantic contribution to a formula is to force the evaluation of the formula in its scope at the time of evaluation. So in

(15) $\mathbf{F}\forall x(\mathbf{N}Ax \rightarrow Dx)$

N's job is to exempt the atomic formula Ax from being evaluated at the future time index introduced by **F** and to force its evaluation at the time index representing the time of evaluation, i.e. present time from the perspective of someone evaluating the formula. The result is an adequate formalization of (13) which could e.g. be used in the formalization an argument involving (15) as a premise.

19 See e.g. Vlach (1973), Kamp (1971).

Interestingly, (13) can also be expressed without temporal operators, if we instead allow the quantifiers of the language to range over times, relativize predications to times, so that "Axt" and "Dxt" stand for "x is alive at time t" and "x is dead at time t" respectively, and take t_0 to stand for the time of evaluation (Cresswell 1990, 19):

(16) $\exists t_1(t_0 < t_1 \wedge \forall x(A(xt_0) \rightarrow D(xt_1)))$

This formula seems to adequately capture what (13) says relative to a particular time of evaluation. Note that, as Cresswell (1990, 21) points out, it might be argued to be objectionable that (16) produces an eternal sentence for each value of t_0. At least it is, if we assume that the truth-value of (13) could change, if e.g. technological advances would allow humans to attain immortality.

The availability of (16) as a translation of (13) raises the question of whether it wouldn't be preferable to just work with the language of first-order logic rather than with the extended language of first-order tense logic which adds new operators. Considerations of parsimony certainly seem to favour this strategy. Why introduce additional operators if we can express the same things without them? Philosophical reasons may be brought to bear on this question. Arthur Prior for example argued that the tense logical formalization of (13) is preferable, considerations of parsimony notwithstanding, since he took tense, which is more naturally expressed using operators like **F**, **P**, and **N**, to be more fundamental than time.[20]

Questions about the choice of formal language are discussed in Hanoch Ben-Yami and, with a historical focus on Frege's *Begriffsschrift*, in Jongool Kim's contributions to the special issue.

4 Quality-Criteria for Formalization

4.1 *Translation Problems and a Simple Quality Constraint*

Assuming that a suitable formal language has been selected, determining the logical form of a natural language sentence is still not a straightforward matter. It seems clear that not every formula of such a language can equally well be used to translate every natural language sentence. But what then makes a

20 See Cresswell (1990, 22) and see Lewis (1968) for the development of counterpart theory, a theory expressible in the language of first-order logic which can express any sentence which can be expressed in the language of first-order modal logic.

doi: 10.48106/dial.v74.i2

formula or a set of formulas an adequate or a correct formalization? Can we formulate general criteria for the quality or admissibility for formalizations of a formal language?[21]

A minimal constraint on the correctness of formalization of sentences is that it should respect certain intuitively valid inferences involving these sentences. In this subsection, the focus will be on two well-known examples of problem cases for translations of natural language sentences into the language of first-order logic which illustrate two different attempts to ensure that this minimal constraint is met.

The first problem specifically concerns a particular type of sentence, namely that of action sentences. Consider the following sentence:

(17) Donald embraced Orman at noon.

The most-straightforward translation of this sentence into the language of first-order logic is

(18) *Edon*

where *Exyz* is the three-place predicate "*x* embraces *y* at time *z*" and *d*, *o*, *n* are individual constants designating Donald, Orman, and the relevant point in time respectively. The problem with this formalization of the sentence is that it does not respect the inferential relation between (17) and the following sentence:

(19) Donald embraced Orman.

Clearly, if Donald embraced Orman at noon, Donald embraced Orman. Yet, if we translate (19) in the same straightforward manner as (17), using a two-place predicate *Fxy* which stands for a sentence of the form "*x* embraces *y*," we get the following formula:

(20) *Fdo*

But this formula is not logically entailed, in classical first-order logic, by (18). A classic discussion of this problem is found in Davidson (1967). Building on previous work by Reichenbach and Kenny, Davidson's solution to the problem

21 This is a topic which has surprisingly not been discussed much in the literature. Adequacy criteria for formalizations in first-order logic are for example discussed in Baumgartner and Lampert (2008); Baumgartner (2010), Blau (1977), Brun (2004, 2012), Epstein (1994), and Sainsbury (2001).

is to propose an alternative formalization-pattern for sequences describing events. According to his proposal, (17) should be formalized as:

(21) $\exists x(Gxdo \wedge Hxn)$,

Here the predicate $Gxyz$ stands for "x is an embrace by y of z," the predicate Hxy for "x happened at time y," and the constants d, o, n retain their earlier referents. This new formula directly entails the formula

(22) $\exists xGxdo$

which, following Davidson's formalization pattern, is an adequate formalization of (19). The problem is hence solved.

Davidson's proposal gives us an example of a formalization pattern which is sensitive to the content of the formalized sentence. As Davidson put it: "Part of what we must learn when we learn the meaning of any predicate is how many places it has, and what sorts of entities the variables that hold these places range over. Some predicates have an event-place, some do not" (1967, 93). Given the previous discussion about the distinction between formal and material inferences, one might think that Davidson's proposal blurs the line between the two kinds of inferences, if such a line can at all be drawn. One might indeed think that both the example discussed by Davidson and the example to be discussed next illustrate that it is, even in the case of first-order logic, a genuinely open question to which extent formal logic can account for the informal notion of entailment, including ostensibly material entailments such as those from (7) to (8) and from (9) to (10).

The second example illustrates a problem case of formalization which arises even if one accepts external constraints on formalization. A classical example discussed in the literature is De Morgan's problem:[22]

(23) All horses are animals.
(24) ∴ All heads of horses are heads of animals.

There is a straightforward way to formalize (23) by simply translating "is a horse" using the predicate-letter F and "is an animal" using the predicate letter G:

(25) $\forall x(Fx \rightarrow Gx)$

22 See Brun (2004, sect. 9, 189ff). See also Brun (2012).

doi: 10.48106/dial.v74.i2

If we formalize (24) in the same manner using the predicate-letter H for "is a head of a horse" and I for "is the head of an animal," we end up with:

(26) $\forall x(Hx \rightarrow Ix)$

If we just consider (24) in isolation, this is may be a fine formalization, but (26) is inadequate in the context of a formalization of the argument from (23) to (24). The inference captured in this argument is intuitively correct, but (25) does not logically entail (26) .

There are different formalizations of (24) which solve the problem (cf. Brun 2004, 193). One solution is to formalize (24) as follows, using the binary predicate K to translate "is the head of" in addition to F and G which are still used to translate "is a horse" and "is an animal" respectively:

(27) $\forall x \forall y((Fy \wedge Kxy) \rightarrow (Gy \wedge Kxy))$

Alternatively, the following formula also does the trick:

(28) $\forall x(\exists y(Fy \wedge Kxy) \rightarrow \exists y(Gy \wedge Kxy))$

Both (27) and (28) are logical consequences of (25), so both (25) and (27), as well as (25) and (28) give us formalizations of the argument from (23) to (24) which can be said to meet the minimal requirement set out earlier in this section. Interestingly however, (27) is logically stronger than (28) in the sense that (28) is a logical consequence of (27), but (27) not of (28). The fact that we can have two different, but non-equivalent ways of formalizing the argument from (23) to (24) raises several general questions about the formalization of arguments (cf. Brun 2004, 194). We might for example ask whether the two variants can be compared concerning their quality as formalizations of the natural language argument they translate, and if so, which one of them offers us the better formalization.

The discussion of the two classical formalization problems illustrate two important general aspect of how we determine the correctness of a formalization. The first and quite obvious point is that the intuitive notion of inference we apply when reasoning using natural language gives us a corrective for correct formalization. The correctness of a formalization can never be a completely formal matter; i.e. logic alone can never tell us whether a formula is a correct formalization of a sentence.[23] Second, whether a formula of a formal

23 Which is of course not to say that we cannot use formal methods to reason about correctness, see Paseau (2019).

language is an adequate formalization of a natural language sentence cannot be determined by considering the sentence in isolation. Correctness rather is a holistic notion which has to take relevant inferential patterns in natural language into account. (Cf. Friedrich Reinmuth's contribution to this special issue.)

These two points give us constraints on adequate formalization, but they obviously fall short of giving us general criteria for the adequateness of formalizations which might, e.g. answer the mentioned questions about the comparative quality of equally admissible alternative formalizations.

4.2 *General Quality Criteria*

What shape could such a general criterion take? Brun distinguishes two kinds of quality criteria, *correctness criteria* and *adequacy criteria* (see 2004, 11). In his terminology, a formalization is *correct* if its validity-relevant features are just those of the sentence or of the argument which it formalizes. But there is a fundamental problem for formalizing arguments which shows that correctness alone is not enough to guarantee that a formalization is a good formalization. Following Blau (1977), this problem has come to be known as the problem of unscrupulous formalization.[24] To see the problem, consider the following example given in Brun (2004, 238):

(29) Every prime number is odd or equal to 2.
(30) There is no prime number which is not odd and not equal to 2.

These two sentences can arguably be recognized to say the same without thinking much about their logical form, e.g. by pondering the meanings of "every" and "there is no." Let us, for the sake of the argument, assume that we accept on an intuitive level that (29) and (30) are equivalent. Using "P" for "is a prime number" and "O" for "is an odd number," a scrupulous formalization of the two sentences would give us the two following formulas:

(31) $\forall x(Px \rightarrow (Ox \vee x = 2))$
(32) $\neg\exists x(Px \wedge (\neg Ox \wedge \neg x = 2))$

Given these translations, we could now provide a formal explanation of our informal judgement that (29) and (30) are equivalent by proving that the two formulas are equivalent in first-order logic. An unscrupulous formalization in

24 Blau's German term is "skrupellose Formalisierung" (see 1977, 18).

doi: 10.48106/dial.v74.i2

contrast would for example be one which translates both (29) and (30) as (31). The goal of our exercise in formalization is to show that we can confirm our informal judgement that (29) and (30) are equivalent and there is no easier equivalence proof than one which demonstrates that a formula, trivially, but correctly, is equivalent to itself. The point of the example is that if correctness is all that matters, then there the unscrupulous formalization is as good as the scrupulous one.

The example of unscrupulous formalizations shows that correctness alone is not a guarantee of the quality of a formalization. This is where adequacy enters the picture. Adequacy is a stricter quality-criterion than correctness, that is, each adequate formalization is a correct formalization, but not vice versa. The notion of adequacy hence allows us to rule out correct, but still problematic formalizations of the sort just discussed. Unscrupulous formalization give us a clear adequacy-constraint: Adequate formalizations do not trivialize non-trivial inferential connections between the resulting formulas, ruling out e.g. a formalization which translates both (29) and (30) as (31). Accordingly, adequacy criteria go beyond correctness criteria in the sense that they ensure that the formalization not only captures the validity-relevant features of the formalized sentences or argument, but also does so in a non-trivial way.

There are, just as in case of the notion of logical entailment, two different conceptions of correctness which are tied to two conceptions of what validity-relevant features are. First, these features can be the truth-conditions of the relevant sentences and formulas, giving us a semantic conception of correctness. The idea then is that a formalization is correct if the formalization has the same truth-conditions as the sentence it formalizes relative to a logic and a translation-schema (or correspondence schema in Brun's terms) which specifies the translations of all relevant expressions of natural language into the relevant formal language.[25]

The validity-relevant features can however also be inferential features, giving us a syntactic conception of correctness. For arguments, the formalization and the formalized argument as stated in natural language have to have the same inferential structure, whereas for the formalization of a single sentence, the formalization is correct if the formally correct inferences in which it can occur are also valid in an informal sense for the corresponding inferences made in natural language.[26]

25 See the correctness principle (WK) in Brun (2004, 210).
26 See the correctness principle (SK) in Brun (2004, 214).

The minimal constraint mentioned in the previous subsection hence concerns the second, the inferential, notion of correctness. Sainsbury discusses the following adequacy criterion for formalizations of English sentences:

> QC1. A formalization is adequate only if each of its logical constants is matched by a single English expression making the same contribution to truth conditions. (Sainsbury 2001, 352)

This proposal is motivated by Sainsbury's discussion of what he calls the "Tractarian vision," that every entailment is a logical entailment. Friends of this idea might be tempted to ensure that material entailments are really logical entailments by putting more structure into the formalizations than the surface form of the sentences requires. They might for example try to ensure that the argument from (7) ("The ball is red") to (8) ("The ball is coloured") counts as logically valid by formalizing its premise and conclusion as follows:

(33) $Rb \wedge Cb$
(34) Cb

A problem with this sort of translation and, more generally, with the Tractarian vision is that it appears to conflate the two distinct projects of analysing the meaning of a sentence and of isolating its logical form.[27] The motivation for formalizing (7) as (33) has to draw on the semantic fact that to say that an object is red is, implicitly, to say that it is coloured. To ensure that the entailment is logical, the proposed formalization hence draws on a fact about the meaning of the non-logical expressions involved in (7). So while the formalization of the argument works on the formal level, it indirectly violates the formality requirement: The formality of the logical entailment between (33) and (34) is not mirrored by the premise and conclusion of the argument as stated in English. Sainsbury's adequacy criterion QC1 systematically blocks ad hoc logicalizations of arguments of this sort.[28]

A drawback of QC1 is that it also threatens Davidson's proposed formalization schema for action sequences: There is arguably no single English

27 See Sainsbury (2001, 354). Note that such translations would also count as unscrupulous in Blau's and Brun's sense.

28 Note that this problem would not arise in the first place in a logically perfect language of the sort which Wittgenstein characterizes in the *Tractatus*. In such a language, all logically simple sentences are fully analyzed in the sense that they do not contain any hidden logical or semantic structure which could be brought out by formalizing them.

doi: 10.48106/dial.v74.i2

expression in “Donald embraced Orman at noon” which makes the same contribution to the sentences’s truth conditions as the existential quantifier in its formalization (21) does with respect to that formula of first-order logic.

Purists who eschew the content sensitivity of Davidson’s formalization pattern might see this as an advantage rather than a drawback, but Brun argues that QC1 suffers from two further problems which are less specific and more severe (see Brun 2004, 253f). First, it presupposes an explanation of what it means for a natural language expression to match or correspond to a logical constant in a formula of the formal language into which one translates. Second, putting the first problem aside, while QC1 rules out some problematic formalizations, such as (33), it likewise rules out uncontroversial formalizations, including in particular:

(35) Müller is sad, Schmidt is happy.
(36) $Sm \wedge Hs$
(37) Crocodiles are green.
(38) $\forall x(Cx \rightarrow Gx)$
(39) Hans owns a red bicycle.
(40) $\exists x(Bx \wedge Rx \wedge Ohx)$

The comma in (35) can hardly be said to make the same contribution to its truth-conditions as the conjunction in (36) and the same can be said about the quantifier and the material conditional in (38) and the existential quantifier, as well as the two conjunctions in (40). QC1 helps rooting out some inadequate formalizations, but it throws the baby out with the bathwater by classifying a range of standard formalizations as inadequate.

There are however better adequacy criteria than QC1, such as the following, (a simplified version of) Brun’s criterion of less precise formalization which gives us a necessary condition for the adequacy of a formalization:

> QC2. For a formula ϕ to be a correct formalization of a sentence A, every formula ψ which is less precise than ϕ has to be such that there is a correct formalization of A which is a notational variant of ψ.[29]

29 Cf. principle (UGK), Brun (2004, 349).

This principle needs a bit of unpacking.[30] First of all, "less precise" is here understood to be a relation which holds between two formulas ϕ and ψ relative to a formalism (i.e. a logic), which are formalizations of the same sentence and which are such that ψ can be generated from ϕ by substituting a logically more complex formula for a sub-formula of ϕ. Of two such formulas, one is less precise than the other if the former gives us a less detailed picture of the logical structure of the sentence. Consider for example the following sentence:

(41) Paul Otto Alfred is an adopted son.

Letting the constant a stand for the name "Paul Otto Alfred" and the predicate P for "is an adopted son," we can formalize (41) as:

(42) Pa

However, we could also use the two predicates Q and R, standing for "is adopted" and "is a son" to formalize (41) as:

(43) $Qa \wedge Ra$

Or we could still be more precise and formalize (41) as follows using the predicate S to translate "is male" and T to translate "is the father of":

(44) $Qa \wedge Sa \wedge \exists x(Txa)$

(42)–(44) are all formalizations of the same sentence, namely (41); furthermore, each of the three formulas can be generated by substitution from the others;[31] finally, the three formulas are increasingly precise, revealing more and more of the formalized sentence's logical structure.

QC2 also involves the notion of a notational variant. This notion can be understood in terms of substitution: A formula ϕ is a notational variant of a formula ψ if, and only if, ϕ can be transformed into ψ by a one-to-one substitution of non-logical predicates and vice versa (see Brun (2004), 301).

Now how does QC2 work? We can think of a logically complex formalization as the result of a step-by-step procedure which starts with an atomic formula and then begins capturing more of the formalized sentence's logical structure

30 Just as with the principle itself, I will in the following simplify the details of Brun's account which is explained in full detail in (2004, sec. 13.2 and 13.4).

31 E.g. we get (43) from (42) by substituting Pa by $Qa \wedge Ra$ and (44) from (43) by substituting $Sa \wedge \exists x(Txa)$ for Ra.

doi: 10.48106/dial.v74.i2

by analyzing it in terms of more complex formulas which all are correct in the semantic sense of having the right truth-conditions. What QC2 tells us is basically that to be an adequate formalization is to only contain logical complexity which can be the result of such a process of refinement. (44) for example counts as adequate in this sense, since if we condense the second conjunction into a single formula, we in any case get a formula which is a notational variant of (43), and which is a semantically correct formalization of the sentence.

With that said, let us return to De Morgan's problem and the two non-equivalent, but seemingly both admissible formalizations of (24), (27) and (28):

(27) $\forall x \forall y((Fy \wedge Kxy) \rightarrow (Gy \wedge Kxy))$
(28) $\forall x(\exists y(Fy \wedge Kxy) \rightarrow \exists y(Gy \wedge Kxy))$

Can QC2 help us decide whether one of the two is a more adequate formalization of (24), the conclusion of De Morgan's argument? Note first that neither of these two formulas is more precise than the other in the relevant sense, since the quantifiers and variables the two formulas contain prevent us from generating one from the other by substituting a logically more complex formula for a sub-formula in either of the two. However, only one of the two formulas, namely (28) stands in the "is more precise than"-relation to (26):

(26) $\forall x(Hx \rightarrow Ix)$

We can generate (28) from (26) by substituting $\exists y(Fy \wedge Kxy)$ and $\exists y(Gy \wedge Kxy)$ for Hx and Ix respectively. (27) cannot be generated in the same way, since the second universal quantifier in (27) cannot be introduced by substituting logically more complex formulas for sub-formulas of (26). The closest we can get to (26) is:

(45) $\forall x \forall y(Mxy \rightarrow Nxy)$

However, it is not clear what the predicates M and N could stand for. Since both are relational predicates, M would have to correspond to something like "is a horse head of" and N to "is an animal head of." Be that as it may, since (45) is a less precise formula than (27), QC2 tells us that (27) is an inadequate formalization of (24), unless there is a notational variant of (45) which is an adequate formalization of (24) ("All heads of horses are heads of animals"). If (45) turned out to be a notational variant of (26), then this condition would be met. However, this is not the case, since due to the presence of the second universal quantifier in (27), we cannot generate it from (26) by one-for-one substituting its non-logical predicates. So whether (27) is an adequate formalization of (24) depends on whether (45) is an adequate formalization of (24).

This opens up a way to informally argue that only (28) is an adequate formalization of (24) by arguing that (45) is not a notational variant of an adequate formalization of (24). Given QC2, the adequacy of (45) cannot be justified by pointing out that it is a less precise formula than the adequate formalization (27) since it is exactly the adequacy of (27) which is at issue, so an independent justification is needed. One might then for example argue that the additional logical complexity of (45) gives us a reason to prefer (26) instead, or one might also target the seemingly unnatural translation schema one would have to adopt to make sense of (45).[32]

5 Choice of Logic

Since our focus here is on deductive logic, the formalisms one has to choose from when formalizing an argument are different logics. The one logic which has the claim to being the default choice is classical first-order logic. It has this status in virtue of some of its formal properties—classical first-order logic is e.g. complete and sound—and its expressive strength. First-order logic can be used to formalize a range of mathematical theories, including e.g. some set theories and, as we have seen, it can be used to express the same, or at least similar claims, as intensional logics such as tense logic or modal logic (see Lewis 1968).

Still, there appear to be reasons to rely on alternative logics. One reason is that one may be compelled to reject logical principles or inference schemata which hold in e.g. classical first-order logic with respect to certain contexts, or topics, or more generally for philosophical reasons. Free logic provides an example of the latter sort. As Karel Lambert describes it, free logic is "free of existence assumptions with respect to its terms, general and singular" (1981, 123). Classical first-order logic involves the assumption that every singular term (e.g. each constant) refers to an object in the domain of quantification.[33] This, free logicians argue, is problematic. Consider for example the sentence:

32 Note that Brun uses an additional adequacy criterion to more formally argue that (28), and not (27), is an adequate formalization of (24) (see (2004, 352–356).

33 See e.g. Frege (1893, 9, note 31).

(46) Heimdallr exists.

In the language of first-order logic, this sentence can be formalized as follows, using the constant h for Heimdallr:

(47) $\exists x(h = x)$

Literally, this formula says that there exists something the same as Heimdallr. Both this logico-literal restatement and (46) itself are, at least insofar as common sense is concerned, false, since Heimdallr is an object of fiction, i.e. an object which does not exist. Given the mentioned assumption about the reference of singular terms, this formula is however a logical truth of classical first-order logic. If we accept first-order logic, we hence seem to be forced to accept an obvious falsehood as true.[34] Free logic offers a way out of this problem, since it allows for the falsity of formulas like (47). This is because unlike in classical logic, the rule of Existential Generalization:

(48) $A \vdash \exists x A(x/t)$

fails in free logic. Here, A is a formula of the language of first order logic and $A(x/t)$ is the formula which results if we replace any occurrence of the individual constant t by the variable x (if there are any). Existential Generalization allows us to e.g. infer from (the formalization in the language of first-order predicate logic of) "Heidallr owns Gjallarhorn" to the existence of something which owns Gjallarhorn. In free logic, this inference is not valid, since, briefly put, that a sentence is satisfied by a particular individual constant does not entail the existence of an object in the domain of discourse which satisfies the formula.[35] Other reasons for adopting particular (non-classical) logics which have been given in the philosophical literature include its adequacy for explaining vagueness (cf. e.g. Machina (1976) or Smith (2008)), or the need to move to a non-classical logic in order to avoid semantic paradoxes such as the liar paradox (cf. e.g. Kripke 1975).

It is a fact that there are different logics, but which one should we rely on in analyzing arguments? Carnap famously adopted a tolerant stance towards logic. He assumed that any choice of logic is permissible in principle and that

34 There are ways to evade this argument, e.g. by adopting the descriptivist theory of proper names famously proposed in Russell, B.A.W. (1905). The dominant view about the reference of proper names, according to which they are directly referential (cf. Kripke 1980), however, speaks against Russell's theory.

35 See Nolt (2020) for a general overview and further explanation.

which logic one relies on is ultimately a matter of its usefulness for a particular purpose.[36] However, Carnap's tolerant attitude is not shared by everyone and we may ask whether, despite the fact that there are different logics, there is one logic which is correct in the sense that it gives us the one correct notion of logical consequence. This question is asked in the recent discussion about logical pluralism, the view that there is more than one correct logic and therefore also more than one correct notion of logical consequence.[37] A recently proposed methodology for choosing between logics based on reflective equilibrium is criticized in Bogdan Dicher's contribution to the special issue. A question about the independence of formalization and choice of logic is raised in Roy Cook's contribution.

6 Genesis of the Special Issue and Acknowledgements

The initial idea for this special issue came about during the workshop "Making it (too) precise" which I organized together with Dominik Aeschbacher and Maria Scarpati in July 2017 at the University of Geneva as part of the SNSF-funded research project "Indeterminacy and Formal Concepts" (project nr. 156554) led by Prof. Kevin Mulligan. After the editorial committee of *Dialectica* approved the proposal for the special issue, an open call for papers was published online. 18 papers in total were submitted, including some of those presented at the workshop in Geneva. All of these paper were subject to the same review process which mirrored that passed by regular submissions to *dialectica*, with the sole differences being that the guest editor was both responsible for the organization of the review process and for the initial internal review. The 13 papers which passed this initial step were double-anonymously reviewed by two expert reviewers. In a third and final step, the papers which were selected by the guest editor based on the recommendations of the reviewers were presented to the editorial committee and the editors who approved the guest editor's decision.

First and foremost, I would like to thank the authors for contributing their papers and allowing them to be published in this special issue. My second greatest debt is to all the reviewers whose work made it possible for an interested bystander like myself to take editorial decisions. I would also like to thank the editorial committee of *Dialectica*, especially Matthias Egg for his

36 See in particular Carnap's principle of tolerance, as set out e.g. in Carnap (1947, sec. 17).

37 See Beall and Restall (2006) and Shapiro (2014) for developments of the position, Field (2009), Priest (2006), Read (2006) for opposing views, and Russell, G.K. (2023) for an overview.

doi: 10.48106/dial.v74.i2

helpful comments and its managing editor Philipp Blum, for giving me the opportunity to edit and for approving the special issue and the Swiss National Science Foundation for financial support at the outset ("Indeterminacy and Formal Concepts," University of Geneva 2014–17, project number 156554, PI: Kevin Mulligan). Finally, I would like to thank Philipp Blum and all the people involved for the work they put into turning *Dialectica* into an open access journal. It is a very happy coincidence, one which only materialized after the reviewing process had been well under way, that this special issue would be one of the first issues of the journal to be freely and openly accessible to anyone over the internet.

7 Overview of the Papers of the Special Issue

In his paper "The Quantified Argument Calculus and Natural Logic," Hanoch Ben-Yami relates his Quantified Argument Calculus (acronym: *Quarc*) to Larry Moss's Natural Logic. The main selling point of both of these logical systems is that they give us logics which are able to account for the validity of certain intuitively correct argument types, such as for example the argument from (7) to (8), which are invalid in classical first-order logic. Ben-Yami shows that Quarc is able to account for the same extended range of arguments which Moss's Natural Logic is designed to capture and furthermore argues that Quarc has the advantage that it does not require to posit negative nouns to do so.

In "Reflective Equilibrium on the Fringe: The Tragic Threefold Story of a Failed Methodology for Logical Theorising," Bogdan Dicher critizises the idea due to Peregrin and Svoboda (2017) that reflective equilibrium can serve as a method for choosing a logic. The core idea of this approach is that the fact that the rules of inference of a logic and the inferences in natural language which it is supposed to formalize can be brought into a (virtuously circular) agreement with each other provides us with a criterion for that logic's adequacy. Dicher's argument against this idea is based on three case studies, one focusing on the impact on harmony of moving from single- to multiple-conclusion, another focusing on the question of how we may distinguish between logics which deliver the same valid logical entailments, focusing on classical first-order logic and strict-tolerant logic (Cobreros et al. 2012), and a third focusing on an application of the logic of first-degree entailment (Anderson and Belnap 1975) by Beall.

Jongool Kim's paper "The Primacy of the Universal Quantifier in Frege's Concept-Script" focuses on the question of why Frege adopted the universal, rather than the existential quantifier as a primitive of the formal system developed in his Frege (1879). This question is not only of historical interest, given that Frege's book is one of the most important contributions to the development of contemporary logic, but also raises a general systematic question about factors motivating the choice of a particular formal language. While Frege never explicitly answered this question, Kim extracts, develops, and discusses three arguments which support this choice from Frege's works and singles out one of them, a philosophical argument based on the idea that choosing the existential quantifier as a primitive instead would have undermined Frege's logicist project of putting arithmetic on a purely logical foundation, as the strongest.

Friedrich Reinmuth's paper "Holistic Inferential Criteria of Adequate Formalization" focuses on adequacy criteria for logical formalization. Following e.g. Brun (2004), Peregrin and Svoboda (2017) and others, Reinmuth assumes that such criteria have to be holistic in the sense that they have to take into account the consequences of the choice one makes in formalizing a particular natural language sentence not only for the target argument, but also for all other arguments involving the same sentence as a premise or conclusion. He points out shortcomings in existing proposals and motivates and develops criteria which extend from arguments to more complex sequences of logical reasoning and which e.g. allow one to distinguish between equivalent formalizations of arguments which nonetheless lead to differences when embedded in such sequences.

Gil Sagi's paper "Considerations on Logical Consequence and Natural Language" focuses on the relation between the notion of logical consequence and ordinary language. Sagi in particular targets three recent arguments due to Glanzberg (2015) to the conclusion that the relation of logical consequence cannot be simply read off natural language. Her paper rebuts these arguments and argues that one of the two positive proposals made by Glanzberg for how one might go beyond natural language in order to get at logical consequence is in fact compatible with the view that this relation exists in natural language.

In " 'Unless' is 'Or,' Unless '$\neg A$ Unless A' is Invalid," Roy T. Cook discusses the formalization of arguments involving the expression "unless," focussing in particular on the differences between formalizations which rely on the same formal language, that of propositional logic, but differ in that they assume classical or intuitionistic logic as the background logic. One of Cook's main points is that his discussion questions the assumption that translations from informal into formal language are logic neutral, in the sense that we can settle for a logical formalization independently of first adopting a particular logic.

Vladan Djordjevic's paper "Assumptions, Hypotheses, and Antecedents" focuses on an important distinction between three ways in which deductive arguments can be cast both in formal languages and in natural language. Djordjevic distinguishes "arguments from assumptions," which are arguments in which each premise is assumed to be logically true and the logical truth of the conclusion is to be established, from "arguments from hypotheses," in which the validity of an inference from the premises to the conclusion is at issue, and from assertions of conditionals which contain the premises of an argument in their antecedent and its conclusion in its consequent. The three categories are often conflated and Djordjevic argues that certain philosophical puzzles, including a standard argument for fatalism and McGee's counterexample to Modus Ponens can be resolved based on these distinctions.

Robert Michels
0000-0003-4982-7239
LanCog, University of Lisbon
robert.michels@edu.ulisboa.pt

References

Anderson, Alan Ross and Belnap, Nuel D., Jr. 1975. *Entailment: The Logic of Relevance and Necessity. Volume 1*. Princeton, New Jersey: Princeton University Press.

Baumgartner, Michael. 2010. "Informal Reasoning & Logical Formalization." in *P.F. Strawson – Ding und Begriff / Object and Concept*, edited by Sarah-Jane Conrad and Silvan Imhof, pp. 11–34. Logos n. 18. Heusenstamm b. Frankfurt: Ontos Verlag, doi:10.1515/9783110323702.11.

Baumgartner, Michael and Lampert, Timm. 2008. "Adequate Formalization." *Synthese* 164(1): 93–115, doi:10.1007/s11229-007-9218-1.

Beall, J. C. and Restall, Greg. 2006. *Logical Pluralism*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199288403.001.0001.

Beall, J. C., Restall, Greg and Sagi, Gil. 2019. "Logical Consequence." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2019/entries/logical-consequence/.

Betz, Gregor. 2010. *Theorie dialektischer Strukturen*. Philosophische Abhandlungen n. 101. Frankfurt a.M.: Vittorio Klostermann, doi:10.5771/9783465136293.

—. 2013. *Debate Dynamics: How Controversy Improves Our Beliefs*. Synthese Library n. 357. Dordrecht: Springer, doi:10.1007/978-94-007-4599-5.

Blau, Ulrich. 1977. *Die dreiwertige Logik der Sprache: Ihre Syntax, Semantik und Anwendung in der Sprachanalyse*. Berlin: de Gruyter, doi:10.1515/9783110838794.

Bonnay, Denis. 2008. "Logicality and Invariance." *The Bulletin of Symbolic Logic* 14(1): 29–68, doi:10.2178/bsl/1208358843.

—. 2014. "Logical Constants, or How to Use Invariance in Order to Complete the Explication of Logical Consequence." *Philosophy Compass* 9(1): 54–65, doi:10.1111/phc3.12095.

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

Brun, Georg. 2003. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. Logos n. 2. Heusenstamm b. Frankfurt: Ontos Verlag. Second edition: Brun (2004), doi:10.1515/9783110323528.

—. 2004. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. 2nd ed. Heusenstamm b. Frankfurt: Ontos Verlag. First edition: Brun (2003).

—. 2012. "Adequate Formalization and de Morgan's Argument." *Grazer Philosophische Studien* 85: 325–335, doi:10.1163/9789401208338_017.

Caret, Colin R. and Hjortland, Ole Thomassen. 2015. "Logical Consequence: Its Nature, Structure, and Application." in *Foundations of Logical Consequence*, edited by Colin R. Caret and Ole Thomassen Hjortland, pp. 3–31. Mind Association Occasional Series. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198715696.003.0001.

Carnap, Rudolf. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 1st ed. Chicago, Illinois: University of Chicago Press. Second edition: Carnap (1956).

—. 1956. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed. Chicago, Illinois: University of Chicago Press. Enlarged edition of Carnap (1947).

Cobreros, Pablo, Egré, Paul, Ripley, David and van Rooij, Robert. 2012. "Tolerant, Classical, Strict." *Journal of Philosophical Logic* 41(2): 347–385, doi:10.1007/s10992-010-9165-z.

Cresswell, Maxwell J. 1990. *Entities and Indices*. Studies in Linguistics and Philosophy n. 41. Dordrecht: Kluwer Academic Publishers, doi:10.1007/978-94-009-2139-9.

Davidson, Donald. 1967. "The Logical Form of Action Sentences." in *The Logic of Decision and Action*, edited by Nicholas Rescher, pp. 81–95. Pittsburgh, Pennsylvania: University of Pittsburgh Press. Reprinted in Davidson (1980, 105–148).

—. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press. Second, enl. edition: Davidson (2001).

—. 2001. *Essays on Actions and Events: Philosophical Essays, Volume 1*. 2nd ed. Oxford: Oxford University Press. Enlarged, doi:10.1093/0199246270.001.0001.

Douven, Igor. 2021. "Abduction." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/sum2021/entries/abduction/.

Dummett, Michael A. E. 1976. "What is a Theory of Meaning (II)." in *Truth and Meaning: Essays in Semantics*, edited by Gareth Evans and John Henry McDowell, pp. 67–137. Oxford: Oxford University Press. Reprinted in Dummett (1993, 34–93), doi:10.1093/oso/9780198250074.003.0004.

—. 1991. *The Logical Basis of Metaphysics*. London: Gerald Duckworth & Co.

—. 1993. *The Seas of Language*. Oxford: Oxford University Press, doi:10.1093/0198236212.001.0001.

Dutilh Novaes, Catarina. 2011. "The Different Ways in which Logic is (said to be) Formal." *History and Philosophy of Logic* 32(4): 303–332, doi:10.1080/01445340.2011.555505.

—. 2012. *Formal Languages in Logic: A Philosophical and Cognitive Analysis*. Cambridge: Cambridge University Press, doi:10.1017/cbo9781139108010.

Epstein, Richard L. 1994. *The Semantic Foundations of Logic. [Volume 2:] Predicate Logic*. Oxford: Oxford University Press.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Cambridge, Massachusetts: Harvard University Press.

Feigl, Herbert and Sellars, Wilfrid, eds. 1949. *Readings in Philosophical Analysis*. New York: Appleton Century Crofts.

Field, Hartry. 2009. "Pluralism in Logic." *The Review of Symbolic Logic* 2(2): 342–359, doi:10.1017/s1755020309090182.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a.S.: Louis Nebert.

—. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, 1. Band*. Jena: Hermann Pohle. Reissued as Frege (1966).

—. 1966. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, 1. Band*. Hildesheim: Georg Olms. Reprografischer Nachdruck der Ausgabe Frege (1893).

Gentzen, Gerhard. 1935. "Untersuchungen über das logische Schließen." *Mathematische Zeitschrift* 39: 176–210, 405–431. Republished as Gentzen (1969), doi:10.1007/BF01201353.

—. 1969. *Untersuchungen über das logische Schliessen*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Glanzberg, Michael. 2015. "Logical Consequence and Natural Language." in *Foundations of Logical Consequence*, edited by Colin R. Caret and Ole Thomassen Hjortland, pp. 71–120. Mind Association Occasional Series. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198715696.003.0003.

Goodman, Nelson. 1955. *Fact, Fiction and Forecast*. Cambridge, Massachusetts: Harvard University Press.

Groarke, Leo. 2021. “Informal Logic.” in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/fall2021/entries/logic-informal/.

Haack, Susan. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511812866.

Hale, Bob and Wright, Crispin, eds. 1997. *A Companion to the Philosophy of Language*. Blackwell Companions to Philosophy. Oxford: Blackwell Publishers. Second edition: Hale, Wright and Miller (2017).

Hale, Bob, Wright, Crispin and Miller, Alexander, eds. 2017. *A Companion to the Philosophy of Language*. 2nd ed. Blackwell Companions to Philosophy. Chichester: Wiley-Blackwell. First edition: Hale and Wright (1997), doi:10.1002/9781118972090.

Hanson, William H. 1997. “Review of Etchemendy (1990).” *The Philosophical Review* 106(3): 365–409, doi:10.2307/2998398.

Institut International de Collaboration Philosophique, ed. 1936. *Actes du Congrès International de Philosophie Scientifique, Paris, 1935. Volume 7: Logique*. Actualités scientifiques et industrielles n. 394. Paris: Hermann & cie.

Kamp, Hans. 1971. “Formal Properties of ‘Now’.” *Theoria* 37(3): 227–273. Reprinted in Kamp (2013, 11–52), doi:10.1111/j.1755-2567.1971.tb00071.x.

—. 2013. *Meaning and the Dynamics of Interpretation: Selected Papers of Hans Kamp*. Leiden: E.J. Brill. Edited by Klaus von Heusinger and Alice ter Meulen, doi:10.1163/9789004252882.

Kripke, Saul A. 1975. “Outline of a Theory of Truth.” *The Journal of Philosophy* 72(19): 690–716. Reprinted in Kripke (2011, 75–98), doi:10.2307/2024634.

—. 1980. *Naming and Necessity*. Oxford: Basil Blackwell Publishers.

—. 2011. *Philosophical Troubles: Collected Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199730155.001.0001.

Lambert, Karel. 1981. “On the Philosophical Foundations of Free Logic.” *Inquiry* 24(2): 147–203. Reprinted and “much revised” in Lambert (2002, 122–175), doi:10.1080/00201748108601931.

—. 2002. *Free Logic: Selected Essays*. Cambridge: Cambridge University Press, doi:10.1017/cbo9781139165068.

Lewis, David. 1968. “Counterpart Theory and Quantified Modal Logic.” *The Journal of Philosophy* 65(5): 113–126. Reprinted, with a postscript (Lewis 1983b), in Lewis (1983a, 26–39), doi:10.2307/2024555.

—. 1983a. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

—. 1983b. “Postscript to Lewis (1968).” in *Philosophical Papers, Volume 1*, pp. 39–46. Oxford: Oxford University Press, doi:10.1093/0195032047.003.0003.

MacFarlane, John. 2000. "What does it Mean to Say that Logic is Formal?" PhD dissertation, Pittsburgh, Pennsylvania: University of Pittsburgh, https://web.archive.org/web/20030510033523id_/http://socrates.berkeley.edu:80/~jmacf/Diss.pdf.

—. 2015. "Logical Constants." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/sum2015/entries/logical-constants/.

Machina, Kenton F. 1976. "Truth, Belief, and Vagueness." *Journal of Philosophical Logic* 5(1): 47–78, doi:10.1007/bf00263657.

Mares, Edwin D. 2020. "Relevance Logic." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/win2020/entries/logic-relevance/.

McCarthy, Timothy G. 1981. "The Idea of a Logical Constant." *The Journal of Philosophy* 78(9): 499–523, doi:10.2307/2026088.

McGee, Vann. 1996. "Logical Operations." *Journal of Philosophical Logic* 25(6): 567–580, doi:10.1007/bf00265253.

Nolt, John E. 2020. "Free Logic." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/win2020/entries/logic-free/.

Paseau, Alexander C. 2019. "A Measure of Inferential-Role Preservation." *Synthese* 196(7): 2621–2642, doi:10.1007/s11229-015-0705-5.

Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. London: Palgrave Macmillan, doi:10.1057/9781137452962.

Peregrin, Jaroslav and Svoboda, Vladimír. 2017. *Reflective Equilibrium and the Principles of Logical Analysis: Understanding the Laws of Logic*. London: Routledge, doi:10.4324/9781315453934.

Plato, Jan von. 2014. "The Development of Proof Theory." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/win2014/entries/proof-theory-development/.

Prawitz, Dag. 1974. "On the Idea of a General Proof Theory." *Synthese* 27(1/2): 63–77, doi:10.1007/bf00660889.

—. 2005. "Logical Consequence From a Constructivist Point of View." in *The Oxford Handbook of Philosophy of Mathematics and Logic*, edited by Stewart Shapiro, pp. 671–695. Oxford Handbooks. Oxford: Oxford University Press, doi:10.1093/oxfordhb/9780195325928.003.0022.

Priest, Graham. 2006. *Doubt Truth to Be a Liar*. Oxford: Oxford University Press, doi:10.1093/0199263280.001.0001.

Prior, Arthur Norman. 1960. "The Runabout Inference-Ticket." *Analysis* 21(2): 38–39. Reprinted in Prior (1976, 85–87), doi:10.1093/analys/21.2.38.

—. 1976. *Papers in Logic and Ethics*. London: Gerald Duckworth & Co. Edited by Peter Geach and Anthony Kenny.

Read, Stephen. 2006. "Monism: The One True Logic." in *A Logical Approach to Philosophy: Essays in Honour of Graham Solomon*, edited by David DeVidi and Timothy Kenyon, pp. 193–209. The University of Western Ontario Series in Philosophy of Science n. 69. Dordrecht: Springer, doi:10.1007/1-4020-4054-7_10.

Rumfitt, Ian. 2017. "Against Harmony." in *A Companion to the Philosophy of Language*, edited by Bob Hale, Crispin Wright, and Alexander Miller, 2nd ed., pp. 225–249. Blackwell Companions to Philosophy. Chichester: Wiley-Blackwell. First edition: Hale and Wright (1997), doi:10.1002/9781118972090.ch10.

Russell, Bertrand Arthur William. 1905. "On Denoting." *Mind* 14(56): 479–493. Reprinted in Feigl and Sellars (1949, 103–115), in Russell, B.A.W. (1956), Russell, B.A.W. (1973, 103–119) and in Russell, B.A.W. (1994, 414–428), doi:10.1093/mind/fzi873.

—. 1956. *Logic and Knowledge: Essays, 1901–1950*. London: George Allen & Unwin. Edited by Robert Charles Marsh.

—. 1973. *Essays in Analysis*. London: George Allen & Unwin. Edited by Douglas Lackey.

—. 1994. *Foundations of Logic, 1903–1905*. The Collected Papers of Bertrand Russell, The McMaster University Edition n. 4. London: Routledge. Edited by Alasdair Urquhart, with the assistance of Albert C. Lewis.

Russell, Gillian K. 2023. "Logical Pluralism." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/fall2023/entries/logical-pluralism/.

Ryle, Gilbert. 1954. *Dilemmas*. Cambridge: Cambridge University Press. The Tarner Lectures 1953, doi:10.1017/cbo9781316286586.

Sagi, Gil. 2014. "Formality in Logic: From Logical Terms to Semantic Constraints." *Logique et Analyse* 57(227): 259–276, doi:10.2143/LEA.227.0.3053506.

—. 2015. "The Modal and Epistemic Arguments Against the Invariance Criterion for Logical Terms." *The Journal of Philosophy* 112(3): 159–167, doi:10.5840/jphil2015112 39.

Sainsbury, Richard Mark. 1991. *Logical Forms: An Introduction to Philosophical Logic*. Oxford: Blackwell Publishers. Second edition: Sainsbury (2001).

—. 2001. *Logical Forms: An Introduction to Philosophical Logic*. 2nd ed. Oxford: Blackwell Publishers. First edition: Sainsbury (1991).

Salmon, Wesley C. 1963. *Logic*. Foundations of Philosophy Series. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Second edition: Salmon (1973).

—. 1973. *Logic*. 2nd ed. Foundations of Philosophy Series. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. First edition: Salmon (1963).

doi: 10.48106/dial.v74.i2

Schröder-Heister, Peter. 2018. "Proof-Theoretic Semantics." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2018/entries/proof-theoretic-semantics/.

Sellars, Wilfrid. 1953. "Inference and Meaning." *Mind* 62(247): 313–338. Reprinted in Sellars (1980) and in Sellars (2007, 3–27), doi:10.1093/mind/lxii.247.313.

—. 1980. *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*. Atascadero, California: Ridgeview Publishing Co. Edited by Jeffrey F. Sicha.

—. 2007. *In the Space of Reasons: Selected Essays*. Cambridge, Massachusetts: Harvard University Press. Edited by Kevin Scharp and Robert B. Brandom.

Shapiro, Stewart. 2014. *Varieties of Logic*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199696529.001.0001.

Sher, Gila Y. 1991. *The Bounds of Logic: A Generalized Viewpoint*. Cambridge, Massachusetts: The MIT Press.

Shoesmith, David J. and Smiley, Timothy J. 1978. *Multiple Conclusion Logic*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511565687.

Smith, Nicholas J. J. 2008. *Vagueness and Degrees of Truth*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199233007.001.0001.

Steinberger, Florian. 2011. "What Harmony Could and Could Not Be." *Australasian Journal of Philosophy* 89(4): 617–639, doi:10.1080/00048402.2010.528781.

Tarski, Alfred. 1936a. "Über den Begriff der logischen Folgerung." in *Actes du Congrès International de Philosophie Scientifique, Paris, 1935. Volume 7: Logique*, edited by Institut International de Collaboration Philosophique, pp. 1–11. Actualités scientifiques et industrielles n. 394. Paris: Hermann & cie. German translation of Tarski (1936b).

—. 1936b. "O pojęciu wynikania logicznego." *Przegląd Filozoficzny* 39: 58–68. Translated into German (Tarski 1936a) and English (Tarski 1956b).

—. 1956a. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. 1st ed. Oxford: Oxford University Press. Trans. J.H. Woodger, 2nd edition: Tarski (1983).

—. 1956b. "On the Concept of Logical Consequence." in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, 1st ed., pp. 409–420. Oxford: Oxford University Press. Trans. J.H. Woodger, 2nd edition: Tarski (1983).

—. 1983. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. 2nd ed. Indianapolis, Indiana: Hackett Publishing Co. Trans. J.H. Woodger, edited and introduced by John Corcoran, 1st edition: Tarski (1956a).

—. 1986. "What are Logical Notions?" *History and Philosophy of Logic* 7(2): 143–154, doi:10.1080/01445348608837096.

—. 2002. "On the Concept of Following Logically." *History and Philosophy of Logic* 23(3): 176–196. Translation of Tarski (1936a) by Magda Stroińska and David Hitchcock, doi:10.1080/0144534021000036683.

Tennant, Neil W. 1987. *Anti-Realism and Logic: Truth as Eternal*. Oxford: Oxford University Press.

Vlach, Frank. 1973. "'Now' and 'Then': A Formal Study in the Logic of Tense Anaphora." PhD dissertation, Los Angeles, California: University of California, https://philpapers.org/archive/VLANAT.pdf.

Williamson, Timothy. 1985. "Converse Relations." *The Philosophical Review* 94(2): 249–262, doi:10.2307/2185430.

Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. International Library of Psychology, Philosophy and Scientific Method. London: Kegan Paul, Trench, Trübner & Co.

Zinke, Alexandra. 2018a. *The Metaphysics of Logical Consequence*. Studies in Theoretical Philosophy n. 6. Frankfurt a.M.: Vittorio Klostermann, doi:10.5771/9783465143451.

—. 2018b. "A Bullet for Invariance: Another Argument Against the Invariance Criterion for Logical Terms." *The Journal of Philosophy* 115(7): 382–388, doi:10.5840/jphil2018115723.

doi: 10.48106/dial.v74.i2

The Quantified Argument Calculus and Natural Logic

Hanoch Ben-Yami

The formalisation of natural language arguments in a formal language close to it in syntax has been a central aim of Moss's Natural Logic. I examine how the Quantified Argument Calculus (Quarc) can handle the inferences Moss has considered. I show that they can be incorporated in existing versions of Quarc or in straightforward extensions of it, all within sound and complete systems. Moreover, Quarc is closer in some respects to natural language than are Moss's systems—for instance, it does not use negative nouns. The process also sheds light on formal properties and presuppositions of some inferences it formalises. Directions for future work are outlined.

Despite the successes of the Predicate Calculus, based on Frege's *Begriffsschrift* (1879), there have been recurrent attempts to develop different logic systems, closer in various respects to natural language. Strawson's (1950, 1952) and Sommers' (1982) are two such familiar earlier ones.

More recently, Lawrence Moss has published a series of works, some co-authored with Pratt-Hartmann, which engage in the similar project of *Natural Logic* (Pratt-Hartmann and Moss 2009; Moss 2010b, 2010c, 2010a, 2011, 2015). Natural Logic has several aims. One main aim is to "construct a system [whose] syntax is closer to that of a natural language than is first-order logic" and give "logical systems in which one can carry out as much simple reasoning in language as possible" (Moss 2010b, 538–539). Moss's works "attempt to make a comprehensive study of the entailment relation in fragments of language," "to go beyond truth conditions and examples, important as they are, and to aim for more global characterizations" (2010b, 561). "The subject of natural logic," Moss writes, "might be defined as 'logic for natural language, logic in natural language.' By this, we aim," he clarifies, "to find logical systems that deal with inference in natural language, or something close to it" (2015, 563). Moss has tried to faithfully represent in his systems standard quantifiers, passive-active voice relations, comparative adjectives, and more.

A different system with similar aspirations which has also been recently developed is the Quantified Argument Calculus, or Quarc.[1] Quarc is a powerful formal logic system, first introduced in Ben-Yami's "The Quantified Argument Calculus" (2014), based on work published by Ben-Yami in the preceding decade (primarily 2004) and closely related to the calculus introduced in Lanzet and Ben-Yami (2004). It is closer in its syntax than is the Predicate Calculus to natural language, sheds light on the logical role of some of the latter's features which it incorporates (such as copular structure, converse relation terms and anaphora), and it is also closer to natural language in the logical relations it validates. Ben-Yami (2014) contains a Lemmon-style natural deduction system for Quarc and a truth-valuational, substitutional semantics; this system has been shown to be sound and complete (Ben-Yami 2014; Ben-Yami and Pavlović 2022). Quarc has since been extended into a sound and complete three-valued system with defining clauses, using model-theoretic semantics (Lanzet 2017). In this latter version it was shown to contain a semantically isomorphic image of the Predicate Calculus. Thus, Quarc has been shown to be at least as strong as the first-order Predicate Calculus, and moreover, the proofs in these papers shed light on the nature of quantification in the Predicate Calculus (see there for details). In other works (Pavlović 2017; Pavlović and Gratz 2019), a sequent calculus has been developed for several versions of Quarc and various properties of the system, such as cut-elimination, subformula property and consistency were proved. Quarc has also been used to investigate Aristotelian logic, both assertoric and modal, in works mentioned above as well as in Raab (2018). Raab concludes that the Quarc-reconstruction he provides of Aristotle's logic is "much closer to Aristotle's original text than other such reconstructions brought forward up to now" (abstract).

It would be interesting to compare what Natural Logic has achieved with what has or can be achieved by Quarc. The present paper embarks on this inquiry. Only *embarks*, for limitations of space and time force us to leave out a comparative study of some central questions of the Natural Logic project. An important issue for Moss is that of *decidability*. He would like to determine whether the logic systems he constructs to incorporate reasoning in natural language, systems which are more limited in their expressive power than the first-order Predicate Calculus, are decidable. Moss and Pratt-Hartmann write:

1 A related approach is developed in Francez, N. (2014).

> From a computational point of view [...] expressive power is a double-edged sword: roughly speaking, the more expressive a language is, the harder it is to compute with. In the last decade, this trade-off has led to renewed interest in *inexpressive* logics, in which the problem of determining entailments is algorithmically decidable with (in ideal cases) low complexity. The logical fragments subjected to this sort of complexity-theoretic analysis have naturally enough tended to be those which owe their salience to the syntax of first-order logic, for example: the two-variable fragment, the guarded fragment, and various quantifier-prefix fragments. But of course it is equally reasonable to consider instead logics defined in terms of the syntax of *natural languages*. (2009, 647–648)

Moss also thinks that decidable systems with less expressive power might represent more faithfully actual human reasoning (2015, 563). Interesting and important as decidability questions are, they will not be addressed in this paper but be left for future work.

The primary concern of this paper is Quarc's capacity to incorporate the natural language inferences studied by Natural Logic. Natural Logic's starting point is a variety of inferences in natural language, all apparently formally valid. Formal systems are then built to incorporate some of these inferences. I shall examine whether Quarc can incorporate these inferences or how it should be extended to accomplish this. I shall also discuss the soundness and completeness of the systems I consider.

Quarc is introduced in the next section; I develop it there only to the extent needed for its application later in the paper. In the section following it, I first present several arguments which Moss considers, and then address each of them in a separate subsection. Along the way I also consider whether, with Moss, we should allow nouns to be negated. I end with a short conclusion, which also includes directions for future work.

1 Introduction to Quarc

By now, Quarc has been presented in several works and in several versions (Ben-Yami 2014; Lanzet 2017; Pavlović and Gratz 2019; Ben-Yami and Pavlović 2022) and there is therefore no need for an additional detailed exposition. Moreover, for our purposes below we do not need to employ the full version of

doi: 10.48106/dial.v74.i2

Quarc that was introduced in Ben-Yami (2014). Accordingly, although I shall first informally introduce the full Quarc language of that paper, the following formal introduction will be of a reduced version (but with the addition of identity), one which we shall then continue to use.

1.1 *Informal Introduction of the System*

Consider a simple subject–predicate or argument–predicate sentence:

(1) Alice is polite.

Its grammatical form can be represented by

(2) (Alice) is polite

with the argument in parenthesis, followed by the copula and then the predicate. In the Predicate Calculus, we formalise this sentence by

(3) $P(a)$

Quarc does not depart from this formalisation, apart from a typographical change: the arguments, in Quarc, are written to the *left* of the predicate:

(4) $(a)P$

Similarly,

(5) Alice loves Bob.

is formalised, in Quarc, as

(6) $(a, b)L$

Consider next the quantified sentence,

(7) Every student is polite.

Its grammatical form can be represented by

(8) (every student) is polite

Here, grammatically, the argument is the noun phrase "every student." In it, the quantifier "every" attaches to the one-place predicate "student," and together they form *a quantified argument*. This is the way quantification is incorporated in Quarc:

(9) $(\forall S)P$

Namely, quantifiers are *not* sentential operators. Rather, they attach to one-place predicates to form quantified arguments. Some other examples:

(10) Some students are polite.
(11) Every girl loves Bob.
(12) Every girl loves some boy.

are formalised (respectively; likewise below) by,

(13) $(\exists S)P$
(14) $(\forall G, b)L$
(15) $(\forall G, \exists B)L$

This basic departure in the treatment of quantification requires a few additional ones. One is the need to reintroduce the copular structure and, with it, modes of predication, as in Aristotelian logic. In natural language, we can negate sentence (1), "Alice is polite," in two ways:

(16) It's not the case that Alice is polite.
(17) Alice isn't polite.

The Predicate Calculus allows only the first mode of negation—the one rarer and somewhat artificial in natural language—namely, sentential negation. Quarc, however, also allows the negation symbol to be written between the argument or arguments and the predicate, signifying negative predication, by contrast to affirmative one. These two sentences are thus formalised, respectively, by

(18) $\neg((a)P)$
(19) $(a)\neg P$

Parentheses can be omitted without ambiguity in these formulas, and they can be written as $\neg aP$ and $a\neg P$. Since the argument is singular, these two formulas are equivalent, and they shall be defined as such both in the proof system and in the semantics below. However, the equivalence does not hold when the argument is quantified:

doi: 10.48106/dial.v74.i2

(20) It's not the case that some students are polite.
(21) Some students aren't polite.

formalised by:

(22) $\neg(\exists SP)$
(23) $(\exists S)\neg P$

These formulas will not be equivalent either in the proof system or in the semantics.

Some adjectives have a corresponding negative form: *polite* and *impolite*, for instance. Yet even if "Alice isn't polite" means the same as "Alice is impolite," this is not the case with all such pairs of adjectives. Often, the negative form designates not the contradictory but the contrary of the positive one: while "reverent" means, feeling or showing deep and solemn respect, "irreverent" means, showing a lack of respect for people or things that are generally taken seriously (Oxford definitions); one's attitude towards, say, religion can be neither reverent nor irreverent. Moreover, many adjectives have no negative form: *tall*, *asleep*, *red*; and relation words usually don't—e.g. "loves" or "teacher of." For these and other reasons (see below on negative nouns), the work done by negative predication cannot generally be accomplished by negative predicates.

All natural languages have the means of reordering the noun-phrases in relational sentences without changing, if the arguments are all singular, what is said by the sentences. Different languages achieve this by different means. English often accomplishes it by changing from active- to passive-voice:

(24) Alice loves Bob.
(25) Bob is loved by Alice.

In the singular case, the two are logically equivalent. But again, this is not generally the case when the arguments are quantified:

(26) Every girl loves some boy.
(27) Some boy is loved by every girl.

Quarc incorporates this reordering by having an n-place predicate written with a permutation of the $1, 2, \ldots, n$ sequence as superscripts to its right. Sentences (24) to (27) are then formalised by,

(28) $(a, b)L$

(29) $(b, a)L^{2,1}$
(30) $(\forall G, \exists B)L$
(31) $(\exists B, \forall G)L^{2,1}$

As with negation, the formulas with singular arguments alone are defined as equivalent in both proof system and semantics, while this equivalence will not generally hold for sentences with quantified arguments.

The last additional feature of Quarc is its use of anaphora. Consider the two sentences,

(32) John loves John.
(33) John loves himself.

The former is rarely used, although one of its uses is to explain the use of the reflexive pronoun "himself" in the latter. The reflexive pronoun "himself" in (33) is anaphoric on the earlier occurrence of "John," its *source*, in the sense that it can be replaced by its source and the sentence will have the same meaning. This eliminable anaphor is what Geach called pronoun of laziness (1962, sec. 76). Quarc incorporates it by using a Greek letter for the anaphor, also written as a subscript to the right of its source. Accordingly, it formalises (32) and (33) by:

(34) $(j, j)L$
(35) $(j_\alpha, \alpha)L$

The formalisation of quantified sentences in which quantified arguments have anaphors is similar:

(36) Every man loves himself.
(37) $(\forall M_\alpha, \alpha)L$

As with negation and reordering, if all arguments are singular, then a Quarc formula with an anaphor and the formula with that anaphor replaced by its source are defined as equivalent in both proof system and semantics. However, the anaphor is no longer generally replaceable by its source when the latter is quantified, neither in natural language nor in Quarc.

With this I conclude the informal introduction of Quarc and turn to the more rigorous introduction of the formal system. However, for the purposes of the discussion below, we don't need to use formulas with anaphora. I therefore introduce a *reduced* version of Quarc, in this respect, which will make it easier to follow and focus on the main argument of this paper. The interested reader is referred to the works mentioned above to see how anaphora is incorporated in the full version of Quarc.

1.2 *Vocabulary of Quarc*

The language of Quarc contains the following symbols:

(38) (**Vocabulary**)

- Predicates: $P, Q, R, \dots$, denumerably many and each with a fixed number of places, including the two-place predicate $=$.
- Singular arguments (SAs): $a, b, c, \dots$, denumerably many.
- Sentential operators: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$.
- Quantifiers: $\forall, \exists$.
- Numerals used as indices, comma, parentheses.

If P is a one-place predicate, then $\forall P$ and $\exists P$ will be called *quantified arguments* (QAs). An *argument* is a singular argument or a quantified one. For every n-place predicate R, $n > 1$, apart from $=$, R^{π}, where π is any permutation of $1, \dots, n$ (including the identity permutation), is called a *reordered* form of R; R^{π} is also an n-place predicate.

1.3 *Formulas of Quarc*

The following rules specify all the ways in which formulas can be generated.

(39) (**Formulas**)

1. (**Basic formula**) If P is a non-reordered n-place predicate and $c_1, \dots, c_n$ singular arguments (SAs), then $(c_1, \dots, c_n)P$ is a formula, called a *basic* formula.
2. (**Reorder**) If P is a reordered n-place predicate, $n > 1$, and $c_1, \dots, c_n$ SAs, then $(c_1, \dots, c_n)P$ is a formula.
3. (**Negative predication**) If P is an n-place predicate and $c_1, \dots, c_n$ SAs, then $(c_1, \dots, c_n)\neg P$ is a formula.
4. (**Identity**) If c_1 and c_2 are SAs then $c_1 = c_2$ is a formula. $c_1 = c_2$ is an alternative way of writing $(c_1, c_2) =$.

5. (**Sentential operators**) If ϕ and ψ are formulas, so are $\neg(\phi)$, $(\phi) \wedge (\psi)$, $(\phi) \vee (\psi)$, $(\phi) \rightarrow (\psi)$ and $(\phi) \leftrightarrow (\psi)$. The parentheses surrounding formulas are called *sentential parentheses*.
6. (**Quantification**) If ϕ is a formula containing an occurrence of an SA c, and substituting the quantified argument qP (i.e. $\forall P$ or $\exists P$) for c will result in qP *governing* ϕ (see definition below), then $\phi[qP/c]$ is a formula. ($\phi[qP/c]$ is the formula in which qP replaced the occurrence of c.)

Formulas of the form, $(c_1, \ldots, c_n)P$, in which P is a *reordered* predicate are not considered basic formulas, as this simplifies the semantic definitions below.

The notion of governance, which is related to that of scope in the Predicate Calculus, is defined as follows:

(40) (**Governance**) An occurrence qP of a QA *governs* a string of symbols A just in case qP is the leftmost QA in A and A does not contain any other string of symbols (B), in which the displayed parentheses are a pair of sentential parentheses, such that B contains qP.

Once anaphors are introduced, the notion of governance becomes non-trivial and its definition needs elaboration. Since they are not introduced in this formal part, determining whether a quantified argument governs a formula is straightforward. For instance, $\exists S$ governs the formulas $(\exists S)P$, $(\exists S)\neg P$, $(a, \exists S)L$, $(\exists S, \forall P)L$ and $(\exists S, \forall P)L^{1,2}$ – the last two because it is to the left of $\forall P$. By contrast, $\exists S$ does not govern $\neg((\exists S)P)$, since it is contained in $((\exists S)P)$; nor $((\exists S)P) \wedge (aQ)$, as it is contained in $((\exists S)P)$; nor $(\forall Q, \exists S)L$, since $\forall Q$ is to its left. For the reduced Quarc language of this paper, a somewhat simpler definition of governance could be provided, practically listing the schemas of formulas governed by a QA; I prefer to use this definition in order to facilitate the transition to fuller Quarc languages. We shall often omit parentheses where no ambiguity arises.

1.4 *Truth-Valuational, Substitutional Semantics*

As in Ben-Yami (2014), I use here a truth-valuational, substitutional semantics for Quarc. Justification of the approach and answers to some common or possible objections, neither specific to Quarc but as a general semantic approach, can be found in Ben-Yami (2014) and Ben-Yami and Pavlović (2022). The results below do not depend on the use of this semantics: a model-theoretic semantics for Quarc can and has been developed. A precursor of Quarc with model-theoretic semantics is found in Lanzet and Ben-Yami (2004) and a three-valued version of Quarc with model-theoretic semantics is found in Lanzet (2017).

(41) (**Truth-Value Assignments**) The following holds for any truth-value assignment, or valuation:

1. (**Basic formula**) Every basic formula is assigned the truth-value of *true* or *false*, but not both.
2. (**Reorder**) Let P be a non-reordered n-place predicate, $n > 1$, and $\pi = \pi 1, \ldots, \pi n$ a permutation of $1, 2, \ldots, n$. Then, the truth-value assigned to $(c_{\pi 1}, \ldots, c_{\pi n})P^{\pi}$ is that assigned to $(c_1, \ldots, c_n)P$.
3. (**Law of Identity**) Every formula of the form $c = c$ is *true*.
4. (**Indiscernibility of Identicals**) If $t = c$ is *true* and the formula $\phi[t_1, \ldots, t_n]$ is a basic formula containing the instances $t_1, \ldots, t_n$ of an SA t, then $\phi[c/t_1, \ldots, c/t_n]$ is *true* if $\phi[t_1, \ldots, t_n]$ is *true*.
5. (**Instantiation**) For every one-place predicate P there is some SA c such that $(c)P$ is *true*.
6. (**Sentential operators**) Let ϕ and ψ be formulas. Then, $\neg(\phi)$ is *true* just in case ϕ is *false*, etc.
7. (**Negative predication**) Let P be an n-place predicate and $c_1, \ldots, c_n$ SAs. The truth-value of $(c_1, \ldots, c_n)\neg P$ is that of $\neg(c_1, \ldots, c_n)P$.
8. (**Quantification**) Let $\phi[\forall P]$ ($\phi[\exists P]$) be a formula governed by an occurrence of $\forall P$ ($\exists P$). If for every (some) SA c for which $(c)P$ is *true*, $\phi[c/\forall P]$ ($\phi[c/\exists P]$) is *true*, then $\phi[\forall P]$ ($\phi[\exists P]$) is *true*. If for some (every) c for which $(c)P$ is *true* $\phi[c/\forall P]$ ($\phi[c/\exists P]$) is *false*, then $\phi[\forall P]$ ($\phi[\exists P]$) is *false*.

(42) (**Validity**) An argument whose premises are all and only the formulas in the set of formulas $\mathfrak{S}$ and whose conclusion is the formula ϕ is *valid*, written $\mathfrak{S} \vDash \phi$, just in case every valuation that makes all the formulas in $\mathfrak{S}$ *true* also makes ϕ *true*, even if we add or eliminate singular arguments from our language (of course, only singular arguments not occurring in $\mathfrak{S}$ or ϕ can be eliminated). We also say that $\mathfrak{S}$ *entails* ϕ.

For a discussion of these definitions, see Ben-Yami (2014).

5 *Proof System*

The proof system used here is based on that found in Ben-Yami (2014) and Ben-Yami and Pavlović (2022), with the omission of the rules for anaphora. I use a Lemmon-style natural deduction system, based on the one introduced in Jaśkowski (1934) and further developed and streamlined in Fitch (1952), Lemmon (1965) and elsewhere. Proofs are written as follows:

(43) (**Proof**) A *proof* is a sequence of lines of the form $\langle L, (i), \phi, R\rangle$, where L is a possibly empty list of line numbers; (i) the *line number* in parenthesis; ϕ a formula; and R the *justification*, a name of a *derivation rule* possibly followed by line numbers, written according to one of the derivation rules specified below. ϕ is said to *depend* on the formulas listed in L. The line numbers in L are written without repetitions and in ascending order. The formula in the last line of the proof is its *conclusion*. If there is a proof with the formula ϕ as conclusion, depending only of formulas from the set $\mathfrak{S}$, then ϕ is *provable* from $\mathfrak{S}$, or $\mathfrak{S} \vdash \phi$.

I next list the derivation rules of the system.

(44) (**Derivation rules**)

1. (**Premise**) As any line of a proof, any formula can be written, depending on itself, its justification being Premise:

$i\ (i)$	ϕ	Premise

2. (**Propositional Calculus Rules, PCR**) We allow the usual derivation rules of the Propositional Calculus.
3. (**Identity Introduction, =I**) As any line of the proof a formula of the form $c = c$ can be written, depending on no premises, with its justification being =I.

(i)	$c = c$	=I

4. (**Identity Elimination, =E**) (This and the following rules specify how to add a line to a proof which contains preceding lines of the

specified forms.) Let ϕ be a basic formula containing occurrences $t_1, \dots, t_n$ of the singular argument t (ϕ may also contain additional occurrences of t).

L_1	(i)	ϕ	
L_2	(j)	$t = c$	
L_1, L_2	(k)	$\phi[c/t_1, \dots, c/t_n]$	$=$E i, j

Where L_1, L_2 is the list of numbers occurring either in L_1 or in L_2.

5. (**Sentence negation to Predication negation, SP**) Let P be an n-place predicate and $c_1, \dots, c_n$ singular arguments.

L	(i)	$\neg(c_1, \dots, c_n)P$	
L	(j)	$(c_1, \dots, c_n)\neg P$	SP i

6. (**Predication negation to Sentence negation, PS**) Let P be an n-place predicate and $c_1, \dots, c_n$ singular arguments.

L	(i)	$(c_1, \dots, c_n)\neg P$	
L	(j)	$\neg(c_1, \dots, c_n)P$	PS i

7. (**Reorder, R**) Let P be an n-place predicate, $n > 1$, and $\pi = \pi 1, \dots \pi n$ and $\rho = \rho 1, \dots \rho n$ two permutations of $1, 2, \dots, n$ (the identity permutation included).

L	(i)	$(c_{\pi 1}, \dots, c_{\pi n})P^{\pi}$	
L	(j)	$(c_{\rho 1}, \dots, c_{\rho n})P^{\rho}$	R i

8. (**Universal Introduction, $\forall$I**) Let $\phi[\forall P]$ be a formula governed by $\forall P$. Assume that neither $\phi[\forall P]$ nor the formulas in lines L apart from $(c)P$ in line i contain any occurrence of the singular argument c.

i	(i)	$(c)P$	Premise
L	(j)	$\phi[c/\forall P]$	
$L - i$	(k)	$\phi[\forall P]$	$\forall$I i, j

Where $L - i$ is the possibly empty list of numbers occurring in L apart from i.

9. (**Universal Elimination, $\forall$E**) Let $\phi[\forall P]$ be a formula governed by $\forall$P.

L_1	(i)	$\phi[\forall P]$	
L_2	(j)	$(c)P$	
L_1, L_2	(k)	$\phi[c/\forall P]$	$\forall$E i, j

10. (**Particular[2] Introduction, $\exists$I**) Let $\phi[\exists P]$ be a formula governed by $\exists P$.

L_1	(i)	$\phi[c/\exists P]$	
L_2	(j)	$(c)P$	
L_1, L_2	(k)	$\phi[\exists P]$	$\exists$I i, j

11. (**Instantial Import, Imp**)[3] Let q stand for either $\exists$ or $\forall$, and $\phi[qP]$ be governed by qP. Assume c does not occur in $\phi[qP]$, ψ or any of the formulas L_1, and in no formula L_2 apart from j and k.

L_1	(i)	$\phi[qP]$	
j	(j)	$(c)P$	Premise

2 Why the quantifier is called, in Quarc, *particular* and not *existential* is explained in Ben-Yami (2004, sec. 6.5; 2014, 123).

3 In Ben-Yami (2014, 133) this rule was called *Instantiation*. "Instantial Import," however, is preferable for several reasons. First, in this way the ambiguity of "Instantiation" is avoided, as it is used only for the truth-value assignment rule in Definition 41.5. Secondly, unlike "Instantiation," the phrase "Instantial Import" does not imply that this derivation rule presupposes that any one-place predicate *has* instances. What it does presuppose is that for a formula as in (i) to be true, *P should have* instances; and this is the case even if we allow some one-place predicates to be empty and adopt a three-valued system as in Lanzet (2017). Lastly, "Instantial Import" hints at a relation of this rule to the Predicate Calculus' existential import.

doi: 10.48106/dial.v74.i2

k	(k)	$\phi[c/qP]$	Premise
L_2	(l)	ψ	
$L_1, L_2 - j - k$	(m)	ψ	Imp i, j, k, l

As examples, I provide three proofs, which between them demonstrate all the derivation rules apart from the rules for identity, which are not special to Quarc, and Reorder, which is used later. First, $(\forall S)P \vdash (\exists S)P$:

1	(1)	$(\forall S)P$	Premise
2	(2)	aS	Premise
3	(3)	aP	Premise
2, 3	(4)	$(\exists S)P$	∃I 2, 3
1	(5)	$(\exists S)P$	Imp 1, 2, 3, 4

This inference, being part of the Aristotelian Square of Opposition, is invalid on the standard translation of these sentences to the Predicate Calculus. Quarc is closer in this respect to Aristotelian Logic; for discussion, see Ben-Yami (2004, 2014), Lanzet (2017), Raab (2018).

Secondly, the Aristotelian Barbara, i.e. $(\forall S)M, (\forall M)P \vdash (\forall S)P$:

1	(1)	$(\forall S)M$	Premise
2	(2)	$(\forall M)P$	Premise
3	(3)	aS	Premise
1, 3	(4)	aM	∀E 1, 3
1, 2, 3	(5)	aP	∀E 2, 4
1, 2	(6)	$(\forall S)P$	∀I 3, 5

And lastly, an Aristotelian conversion: "No P is S" follows from "No S is P." Instead of introducing into Quarc a negative quantifier translating "no"—something that *can* be done—these sentences are translated here as synonymous with "Every/any S is not P" or $(\forall S)\neg P$, and $(\forall P)\neg S$, and we show that $(\forall S)\neg P \vdash (\forall P)\neg S$:

1	(1)	$(\forall S)\neg P$	Premise
2	(2)	aP	Premise

3	(3)	aS	Premise
1, 3	(4)	$a\neg P$	∀E 1, 3
1, 3	(5)	$\neg aP$	PS 4
1, 2	(6)	$\neg aS$	PCR (¬I) 3, 2, 5
1, 2	(7)	$a\neg S$	SP 6
1	(8)	$(\forall P)\neg S$	∀I 2, 7

For additional examples, see Ben-Yami (2014) and Ben-Yami and Pavlović (2022).

2 Incorporation in Quarc of the Inferences Motivating the Natural Logic Project

2.1 *The Inferences to be Considered*

In different works, Moss provides different examples of the kinds of inference he discusses in the context of his Natural Logic project. I shall use here, as our point of departure, the inferences he lists in his “Natural Logic” (Moss 2015, 561–562). This list is more detailed and more recent than those found elsewhere in his writings.[4]

1. Passive voice

   Some dog sees some cat.
   ⍰ Some cat is seen by some dog.

2. Conjunctive predicates

   Bao is seen and heard by every student.
   Amina is a student.
   ⍰ Amina sees Bao.

3. Comparative adjectives

4 A reviewer drew my attention to two other relevant works by Moss (2016) and Moss and Topal (2020) (the latter published, online only, shortly before this paper was submitted), in which additional inferences involving comparative quantifiers are involved. I comment on them when discussing comparative quantifiers below.

doi: 10.48106/dial.v74.i2

Every giraffe is taller than every gnu.
Some gnu is taller than every lion.
Some lion is taller than some zebra.
⍰ Every giraffe is taller than some zebra.

4. Defining clauses

All skunks are mammals.
⍰ All who fear all who respect all skunks fear all who respect all mammals.

5. Comparative quantifiers

More students than professors run.
More professors than deans run.
⍰ More students than deans run.

I shall examine the incorporation of inferences of these kinds in Quarc, each in a separate subsection. But before turning to them, I address a different feature which some of Moss's systems contain: negative nouns.

2.2 *Negative Nouns*

Some of Moss's formal systems contain devices intended to represent "negated nouns such as 'non-man' or 'non-animal' " (Pratt-Hartmann and Moss 2009, 648). Moss thinks that "this is rather unnatural in standard speech but it would be exemplified in sentences like *Every non-dog runs*" (2015, 567–568). Other examples Moss provides there are *All non-apples on the table are blue* and *Bernadette knew all non-students at the party* (Pratt-Hartmann and Moss 2009, 564).

But when such sentences *are* used, which I suspect is rarely, they are surely used as elliptical for sentences like, "All *fruits* on the table which aren't apples are blue" or "Bernadette knew all non-student *guests* at the party." There were also breadcrumbs on the table, but we didn't mean to say that *they* were blue; and there were also drinks and finger food at the party.

This ellipsis understanding is also shared by Moss. In his (2010b, 539–540), we find an introductory dialogue between A, Moss's mouthpiece, and a Questioning Q. Q requests "an example of some non-trivial inference carried out in natural language," to which A responds by mentioning an inference containing the premise, *Every non-pineapple is bigger than every unripe fruit*. Q immediately remonstrates: " 'non-pineapple'?! I thought this was supposed to be natural language"; and A excuses himself with, "Take it as a shorthand for 'piece of fruit which is not a pineapple'." Regrettably, Q acquiesces: "Ok, I get it."

Yet if, instead of Q, A would have encountered Critical C, she might have retorted, "So why not stay with 'fruits which aren't pineapples'? Should Logic turn a shorthand into a formal syntactic feature?! And you anyway intend to incorporate defining clauses in your system, for instance when formalising 'all *who respect all skunks*,' so you *shall* have the resources for 'fruits which aren't pineapples.' If your goal is, as you stated, 'logic for natural language, logic in natural language,' then try avoiding non-men, non-dogs and other non-natural creatures."

C's point is supported by an observation due to Aristotle. In his *Categories* (~BC330), when discussing primary, individual substances—an individual man or horse, for instance—and secondary substances, like "man" and "animal" as species and genera, he notes: "Another mark of substance is that it has no contrary. What could be the contrary of any primary substance, such as the individual man or animal? It has none. Nor can the species or the genus have a contrary" (*Cat.* 5, 3b24). Since there is no contrary to man or animal, "non-man" and "non-animal" cannot function, on their own, as noun phrases.

The actual natural language sentences which Moss formalises by means of formal negative nouns, designated by a bar ($\overline{q}$ for non-q's), are sentences like, "Some p aren't q" and "Some p don't r any q," formalised by $\exists(p, \overline{q})$ and $\exists(p, \forall(q, \overline{r}))$ (2015, 573). (We don't need to go into the details of Moss's syntax, since for our purposes the idea is sufficiently clear from these examples.) These two sentences are formalised in Quarc by $(\exists P)\neg Q$ and $(\exists P, \forall Q)\neg R$. Accordingly, Quarc can formalise these sentences without recourse to negative nouns but by using negation as a mode of predication, as it is indeed used in natural language.

I think that finding the idea of negative nouns acceptable is influenced by the semantic idea of a *domain of discourse*. If, when quantifying, the plurality over which we quantify is that of a domain of discourse, then we can single out a part of it either as containing all items to which a predicate p applies, or all those to which *it does not apply*. Indeed, when Moss develops a semantics for languages that include negative nouns, his model or structure $\mathfrak{U}$ contains a non-empty set A which functions as the domain, and if $p^{\mathfrak{U}} \subseteq A$,

doi: 10.48106/dial.v74.i2

then $\overline{p}^{\mathfrak{U}} = A\backslash p^{\mathfrak{U}}$ (Pratt-Hartmann and Moss 2009, 651). However, a domain of discourse, in the technical sense in which the idea is employed in semantics, is an artefact of Fregean Logic, whose quantified sentences contain no expression specifying the plurality over which they quantify. For this reason, the semantics must introduce an otherwise implicit domain. Natural language sentences, by contrast, do specify the plurality over which they quantify: when I say, "All *your students* came to class," I specify your students as the relevant plurality. Quarc follows natural language in this respect, and needs no domain of discourse or of quantification (Ben-Yami 2004, 59–60; Lanzet 2017). Once the domain is eliminated, "non-man" and "non-animal" have nothing to designate and should be eliminated as well.

For these reasons, I think that negative nouns are not needed and should not be included in a logic which aspires to be a logic for natural language. As argued above, the rare sentences which apparently use them are better seen as elliptical: as such they can be formalised in Quarc, which therefore does not need to contain negative nouns.

2.3 *Passive Voice*

(45) Some dog sees some cat.
⍰ Some cat is seen by some dog.

Quarc was developed to incorporate reordering devices such as the active–passive voice distinction. If "*a* sees *b*" is formalised, "$(a, b)S$," then "*b* is seen by *a*" is formalised, "$(b, a)S^{2,1}$." We show that,

(46) $(\exists D, \exists C)S \vdash (\exists C, \exists D)S^{2,1}$

Proof.

1	(1)	$(\exists D, \exists C)S$	Premise
2	(2)	aD	Premise
3	(3)	$(a, \exists C)S$	Premise
4	(4)	bC	Premise
5	(5)	$(a, b)S$	Premise
5	(6)	$(b, a)S^{2,1}$	R 5
2, 5	(7)	$(b, \exists D)S^{2,1}$	∃I 2, 6
2, 4, 5	(8)	$(\exists C, \exists D)S^{2,1}$	∃I 4, 7

2, 3	(9)	$(\exists C, \exists D)S^{2,1}$	Imp 3, 4, 5, 8
1	(10)	$(\exists C, \exists D)S^{2,1}$	Imp 1, 2, 3, 9

□

Quarc with truth-valuational semantics has been shown to be sound and complete in Ben-Yami (2014) and Ben-Yami and Pavlović (2022); a model-theoretic version of this result is found, for an earlier version of the system and for a three-valued version of it, in Lanzet and Ben-Yami (2004) and Lanzet (2017). Accordingly, Quarc is a sound and complete formal system, with a syntax modelled on natural language's, which incorporates inferences like (45).

2.4 *Conjunctive Predicates*

(47) Bao is seen and heard by every student.
Amina is a student.
⍰ Amina sees Bao.

The new element in this inference is the conjunctive verb, or more generally conjunctive predicate, "see and hear." We shall extend Quarc to incorporate it.

We take our cue for the incorporation of conjunctive predicates in Quarc from the way negative predication, reordering and anaphora were incorporated in it. Namely, we shall define valuation- and derivation rules for the case in which all arguments are singular terms, and show that these together with the other rules which have already been defined provide us with desirable results for the more complex cases as well.

2.4.1 Vocabulary

We do not extend the basic vocabulary of Quarc but define,

(48) (**Conjunctive predicates**) If P and Q are n-place predicates, so is $(P) \wedge (Q)$, which is called a *conjunctive predicate*.

Conjunction of predicates can be iterated. Assuming P, Q and R are n-place predicates, so are $((P)\wedge(Q))\wedge(R)$, $(P)\wedge((Q)\wedge(R))$, $((P)\wedge(Q))\wedge((R)\wedge(P))$, and so on. However, as can be proved, formulas with the same predicates ordered and grouped in whichever way, with or without repetition, are equivalent both semantically and proof-theoretically. This allows us to omit parentheses for some conjunctive predicates: both $((P) \wedge (Q)) \wedge (R)$ and $(P) \wedge ((Q) \wedge (R))$ can be written as $P \wedge Q \wedge R$.

Notice that many-place conjunctive predicates can be reordered like any other many-place predicate.

2.4.2 Formulas

No new rules. If P and Q are one-place predicates, then $(a)(P) \wedge (Q)$ is a formula. Similarly for any n-place predicates and any arguments.

2.4.3 Semantics

(49) (**Conjunctive Predication**). Let P and Q be n-place predicates, and $c_1, \ldots, c_n$ singular arguments. The truth-value assigned to $(c_1, \ldots, c_n)(P) \wedge (Q)$ on a valuation is that assigned to $(c_1, \ldots, c_n)P \wedge (c_1, \ldots, c_n)Q$.

Examples. If, on a given valuation, aP, aQ and aR are *true*, then so are, according to our definition, $a(P) \wedge (Q)$, $a(Q) \wedge (R)$ and $a(R) \wedge (P)$. Accordingly, so are $a((P) \wedge (Q)) \wedge (R)$, $a(P) \wedge ((Q) \wedge (R))$ and $a((P) \wedge (Q)) \wedge ((R) \wedge (P))$. If aP is *false*, then so are $a(P) \wedge (Q)$, $a(P) \wedge ((Q) \wedge (R))$ and $a(R) \wedge (P)$; and so on.

This rule yields the desirable results for the two different sentences,

(50) Every linguist knows and admires some philosopher.

formalised as,

(51) $(\forall L, \exists P)(K) \wedge (A)$

and

(52) Every linguist knows some philosopher and every linguist admires some philosopher.

Formalised as,

(53) $(\forall L, \exists P)K \wedge (\forall L, \exists P)A$

According to *Universal Quantification*, (51) is *true* on a valuation just in case so are all formulas of the form, $(l, \exists P)(K) \wedge (A)$, where for l the formula lL is *true*. The formula $(l, \exists P)(K) \wedge (A)$ is *true*, according to *Particular Quantification*, just in case so is some formula of the form, $(l, p)(K) \wedge (A)$, where for p the formula pP is *true*. Next, according to *Conjunctive Predication*, $(l, p)(K) \wedge (A)$ is *true* just in case so is $(l, p)K \wedge (l, p)A$. Namely, (51) is *true* iff every linguist knows some philosopher and admires *the same* philosopher. By contrast, since (53) is *true* just in case so is each of its conjuncts, we shall not get that every linguist need admire a philosopher he knows.

2.4.4 Proofs

We add an introduction and an elimination rules for conjunctive predicates:

(54) (**Conjunctive Predication Introduction, CP-I**) Let P and Q be n-place predicates, $c_1, \ldots, c_n$ singular arguments.

L	(i)	$(c_1, \ldots, c_n)P \wedge (c_1, \ldots, c_n)Q$	
L	(j)	$(c_1, \ldots, c_n)(P) \wedge (Q)$	CP-I i

(55) (**Conjunctive Predication Elimination, CP-E**) Let P and Q be n-place predicates, $c_1, \ldots, c_n$ singular arguments.

L	(i)	$(c_1, \ldots, c_n)(P) \wedge (Q)$	
L	(j)	$(c_1, \ldots, c_n)P \wedge (c_1, \ldots, c_n)Q$	CP-E i

It is straightforward to see that soundness is preserved.

The completeness of Quarc on the truth-valuational approach is proved in Ben-Yami and Pavlović (2022) by adapting Henkin's proof (1949). We won't provide here the complete proof but only specify its features that are relevant for proving that the completeness of the system is preserved with the additional structures introduced in this paper. As part of the proof, a "Henkin Theory" is specified, consisting of all formulas falling under certain schemas. It is then shown that any valuation that respects the truth-value assignment rules for the connectives of the propositional calculus while making all the formulas of the Henkin Theory *true*, respects all the truth-value assignment rules of Quarc as well. Later, some of the formulas of the Henkin Theory are shown to be theorems of Quarc.

To prove that completeness is preserved, we should add to the Henkin theory the axiom schema,

(56) $(c_1, \ldots, c_n)(P) \wedge (Q) \leftrightarrow ((c_1, \ldots, c_n)P \wedge (c_1, \ldots, c_n)Q)$

doi: 10.48106/dial.v74.i2

Any valuation that respects the truth-value assignment rule for the connective $\leftrightarrow$ while making all the formulas of this form *true*, clearly respects Conjunctive Predication (49) as well. And, given CP-I and CP-E, this is a schema of theorems of Quarc. See Henkin (1949) and Ben-Yami and Pavlović (2022) for further details.

We can now turn to a proof of the argument opening this subsection. We formalise it as follows:

Bao is seen and heard by every student: $(b, \forall S)(C \wedge H)^{2,1}$
Amina is a student: aS
$\therefore$ Amina sees Bao: $(a, b)C$

We show that,

(57) $(b, \forall S)(C \wedge H)^{2,1}, aS \vdash (a, b)C$

Proof.

1	(1)	$(b, \forall S)(C \wedge H)^{2,1}$	Premise
2	(2)	aS	Premise
1, 2	(3)	$(b, a)(C \wedge H)^{2,1}$	$\forall$E 1, 2
1, 2	(4)	$(a, b)C \wedge H$	R 3
1, 2	(5)	$(a, b)C \wedge (a, b)H$	CP-E 4
1, 2	(6)	$(a, b)C$	PCR ($\wedge$ E) 5

□

2.5 *Comparative Adjectives*

(58) Every giraffe is taller than every gnu.
Some gnu is taller than every lion.
Some lion is taller than some zebra.
$\therefore$ Every giraffe is taller than some zebra.

Most comparative adjectives are *transitive*: if Alice is *younger* than Bob, and Bob younger than Charlie, then Alice is younger than Charlie. It might thus seem that this transitivity is built into language as a formal rule, for any comparative adjective of the form, ϕ-*er*. There are, however, exceptions, as we learn from Rock–Paper–Scissors: in this game, paper is strong*er* or bett*er* than rock, rock is stronger than scissors, yet scissors is stronger than paper.

Such exceptions notwithstanding, we shall treat in this subsection comparative adjectives of the form ϕ-*er* as transitive. I do not think that the transitivity of adjectives of the ϕ-*er* structure is merely a frequent albeit contingent fact. Rather, we have here a rule of grammar which allows exceptions. That the past tense of "go" is "went" does not show it not to be a rule that the past tense of verbs is formed by adding "ed." With comparative adjectives we have a different kind of rule and exception, concerning not syntax but meaning; yet this does not affect the fact that transitivity is a rule for the use of comparative adjectives, to be overridden only if the exception is explicitly introduced.

2.5.1 Vocabulary and formulas

We add to the language denumerably many two-place *comparative predicates*, $P_{\text{er}}, Q_{\text{er}}, R_{\text{er}} \ldots$ No new formula rules.

2.5.2 Semantics

(59) (**Comparative Adjective Transitivity**). Let P_{er} be a comparative predicate, and c_1, c_2 and c_3 singular arguments. If the truth-value assigned to $(c_1, c_2)P_{\text{er}}$ and $(c_2, c_3)P_{\text{er}}$ on a valuation is *true*, then that assigned to $(c_1, c_3)P_{\text{er}}$ is also *true*.

2.5.3 Proofs

(60) (**Comparative Adjective Transitivity, CAT**) Let P_{er} be a comparative predicate, c_1, c_2 and c_3 singular arguments.

$L1$	(i)	$(c_1, c_2)P_{\text{er}}$	
$L2$	(j)	$(c_2, c_3)P_{\text{er}}$	
$L1, L2$	(k)	$(c_1, c_3)P_{\text{er}}$	CAT i, j

Soundness is again immediate. Completeness is proved by adding to the Henkin theory all the formulas which fall under the schema,

(61) $(c_1, c_2)P_{\text{er}} \wedge (c_2, c_3)P_{\text{er}} \rightarrow (c_1, c_3)P_{\text{er}}$

Any valuation that respects the truth-value assignment rules for the connectives $\wedge$ and $\rightarrow$ while making all the formulas of this form *true*, respects (59)

doi: 10.48106/dial.v74.i2

as well. All formulas of this form are theorems of Quarc, provable from CAT. See again Ben-Yami and Pavlović (2022) for further details.

The proof of (58) is quite tedious and adds no interesting element to what we learn from proofs of simpler inferences. I shall therefore formalise and prove instead the following:

(62) Every giraffe is taller than every wildebeest: $(\forall G, \forall W)T_{\text{er}}$
Some wildebeest is taller than every lion: $(\exists W, \forall L)T_{\text{er}}$
⍰ Every giraffe is taller than every lion: $(\forall G, \forall L)T_{\text{er}}$

We show that:

(63) $(\forall G, \forall W)T_{\text{er}}, (\exists W, \forall L)T_{\text{er}} \vdash (\forall G, \forall L)T_{\text{er}}$

Proof.

1	(1)	$(\forall G, \forall W)T_{\text{er}}$	Premise
2	(2)	$(\exists W, \forall L)T_{\text{er}}$	Premise
3	(3)	gG	Premise
1, 3	(4)	$(g, \forall W)T_{\text{er}}$	∀E 1, 3
5	(5)	wW	Premise
1, 3, 5	(6)	$(g, w)T_{\text{er}}$	∀E 4, 5
7	(7)	$(w, \forall L)T_{\text{er}}$	Premise
8	(8)	lL	Premise
7, 8	(9)	$(w, l)T_{\text{er}}$	∀E 7, 8
1, 3, 5, 7, 8	(10)	$(g, l)T_{\text{er}}$	CAT 6, 9
1, 3, 5, 7	(11)	$(g, \forall L)T_{\text{er}}$	∀I 8, 10
1, 5, 7	(12)	$(\forall G, \forall L)T_{\text{er}}$	∀I 3, 11
1, 2	(13)	$(\forall G, \forall L)T_{\text{er}}$	Imp 2, 5, 7, 12

□

2.5.4 Asymmetry

Another property of comparative adjectives is asymmetry. If Alice is younger than Bob, then Bob isn't younger than Alice. Unlike transitivity, asymmetry seems to have no exception for comparative adjectives.

This property can also be straightforwardly incorporated in Quarc. Nothing needs to be added to either vocabulary or formula rules. In the semantics, the rule should be that if $(c_1, c_2)P_{\text{er}}$ is *true* on a valuation, then $(c_2, c_1)P_{\text{er}}$ is *false* on it. And the rule of inference should allow the inference $(c_1, c_2)P_{\text{er}} \vdash \neg(c_2, c_1)P_{\text{er}}$. We shall not develop this further here.

2.6 *Defining Clauses*

(64) All skunks are mammals.
⍰ All who fear all who respect all skunks fear all who respect all mammals.

Those who respect the skunks and mammals, as well as those who fear the former, are presumably not respectful triangles or fearful ideas, say. Which respectful and fearful "things" are referred to would depend on context, but something more specific does seem to be meant. We shall assume here that the conclusion is about *creatures* generally, and consider it as elliptical for,

(65) All *creatures* who fear all *creatures* who respect all skunks fear all *creatures* who respect all mammals.

This will enable us to treat inference (64) by means of the extended, three-valued Quarc system developed in Lanzet (2017), which has the syntactic and semantic resources to represent defining clauses and can straightforwardly translate sentences such as (65).

One might object and claim that the conclusion of (64) is about *absolutely everything*. Triangles and ideas, so might one continue, also fall within its purview, only they happen not to fear or respect anything, ipso facto skunks and mammals. I find this approach unconvincing when applied to natural language, whose logic both Natural Logic and Quarc aim to represent. However, the issue need not be decided for the purpose of formalising inference (64) in Quarc: the means for representing absolute generality are provided in both Lanzet and Ben-Yami (2004) and Lanzet (2017), in each somewhat differently, by the introduction of a special predicate, *Thing* or T. Very roughly, the idea is that everything is a Thing: for every constant c, cT is *true*. (This special predicate also helps explore the relations between Quarc and the Predicate Calculus.) We shall not develop this idea further here, though, but continue with the assumption that a predicate with narrower application *is* assumed, and use *creature* as in (65).

The three-valued Quarc system of Lanzet (2017) is too complex to be fully presented in this paper. I shall therefore introduce only some of its features,

doi: 10.48106/dial.v74.i2

which will enable us to get an idea of how sentence (65) and consequently inference (64) are handled by it. The reader is referred to Lanzet (2017) for a full exposition. Since we are not inquiring into decidability in this paper but leaving it as a subject for future work, neither shall we inquire whether a restricted, simpler yet complete and decidable version of that system suffices for the formalisation of the relevant arguments.

2.6.1 Compound Predicates

Consider the sentence,

(66) Alice is a woman who knows Bob.

It is logically equivalent to,

(67) Alice is a woman and Alice knows Bob.

While (67) is formalised in Quarc as,

(68) $aW \wedge (a, b)K$

we shall formalise (66) by:

(69) $aW_x : (x, b)K$

The chain of symbols, $W_x : (x, b)K$, is considered *a compound predicate*.

More generally, if $\phi[a]$ is a formula and P a one-place predicate, then $P_x : \phi[x]$ is a *compound predicate*, which is also a one-place predicate. $\phi[a]$ contains no occurrence of x (to avoid ambiguity), and x replaced some or all occurrences of a in $\phi[a]$. $P_x : \phi[x]$ can be read, *P which is* ϕ. $(b)P_x : \phi[x]$ is *true* on a valuation just in case bP and $\phi[b/x]$ are *true* on that valuation.

With this in place, we can formalise the following compound predicates:

creatures who respect Mumbo	$C_x : (x, m)R$
creatures who respect all mammals	$C_x : (x, \forall M)R$
creatures who fear all creatures	$C_x : (x, \forall C)F$
creatures who fear all creatures who respect Mumbo	$C_x : (x, \forall C_y : (y, m)R)F$
creatures who fear all creatures who respect all mammals	$C_x : (x, \forall C_y : (y, \forall M)R)F$

And we can now formalise sentence (65) as well, "All creatures who fear all creatures who respect all skunks fear all creatures who respect all mammals":

(70) $(\forall C_x : (x, \forall C_y : (y, \forall S)R)F, \forall C_y : (y, \forall M)R)F$

2.6.2 Proofs

Lanzet (2017) develops a three-valued system, allowing for some formulas to lack a truth value. “All my children work in the coal mines” is neither true nor false when uttered by a childless person. Similarly, $\exists SP$ and $\forall SP$ will lack a truth value when S has no instances. If our conception of validity in a three-valued system is that truth entails truth, and this *is* Lanzet’s conception, then this three-valued framework complicates the proof system. The classical Negation Introduction rule, for instance, cannot be employed. In addition, some of the rules for quantifiers should be modified, because in some cases we should guarantee that the predicate occurring in the argument position, say P, has instances. This can be done in several ways, one of them by having $(\exists P)P$ among our premises: this formula is *true* if and only if P has instances. For these two reasons, the ∀-Introduction rule is replaced by two rules. Lanzet uses a proof system which operates on sequents, although resembling a natural deduction system in its inference rules. Adapting his rules to the system used in this paper, his ∀I1 rule will be:

i	(i)	cP	Premise
L_1	(j)	$\phi[c/\forall P]$	
L_2	(k)	$\exists PP$	
$L_1 - i, L_2$	(l)	$\phi[\forall P]$	∀I1 i, j, k

Where $\forall P$ governs $\phi[\forall P]$ and c does not occur in L_1 apart from i, in L_2 or in $\phi[\forall P]$.

Returning to the inference with which we opened this subsection, on the conception of validity as truth entails truth, sentence (65), “All creatures who fear all creatures who respect all skunks fear all creatures who respect all mammals,” follows from “All skunks are mammals” only if we assume that the compound predicates in the conclusion’s argument positions, “creatures who fear all creatures who respect all skunks,” and “creatures who respect all mammals” have instances. Otherwise, if no one respected mammals, say, there would be no one to fear in the conclusion, and a true premise would have a conclusion which is neither true nor false.—We *can* develop a different conception of validity for three-valued systems, in which, instead of truth leading to truth, an argument is valid just in case, if its premises are not false, then its conclusion isn't false either (Halldén 1949). Another option is to define validity for a three-valued system as, if the premises are true then the conclusion isn't false (strict-to-tolerant validity, Cobreros et al. 2013). On either conception, a valid argument with true premises may have a conclusion which has no truth-value, and no additional premise should be added to (64). Both options are worth exploring, but here we shall limit ourselves to the option Lanzet adopts and take validity to mean, truth entails truth.

We should, therefore, add to (64) the two premises,

(71) $(\exists C_x : (x, \forall C_y : (y, \forall S)R)F)C_x : (x, \forall C_y : (y, \forall S)R)F$

(72) $(\exists C_y : (y, \forall M)R)C_y : (y, \forall M)R$

and show the following:

(73) $\forall SM,$
$(\exists C_x : (x, \forall C_y : (y, \forall S)R)F)C_x : (x, \forall C_y : (y, \forall S)R)F,$
$(\exists C_y : (y, \forall M)R)C_y : (y, \forall M)R \ \vdash$
$(\forall C_x : (x, \forall C_y : (y, \forall S)R)F, \forall C_y : (y, \forall M)R)F$

The proof is long and requires familiarity with the rules of Lanzet (2017), so instead of providing it we shall show that the inference is valid. Since the system of that paper was proved there to be complete, it follows that the inference can be proved.

Proof. Proof. We should show that, if on a valuation $\mathfrak{B}$ the three premises of (73) are *true*, then for every instance a of $C_x : (x, \forall C_y : (y, \forall S)R)F$ and every instance b of $C_y : (y, \forall M)R$, the following is also *true*, $(a, b)F$. From premises (71) and (72), we know that each of these compound predicates has instances. So suppose $(a)C_x : (x, \forall C_y : (y, \forall S)R)F$ is *true* on $\mathfrak{B}$ with a specific set of SAs (remember that on the truth-valuational semantics, we may add or eliminate singular arguments from our language). Then so are aC and $(a, \forall C_y : (y, \forall S)R)F$. But this means that $C_y : (y, \forall S)R$ has instances on $\mathfrak{B}$, and that for any of its instances c, $(a, c)F$ is *true* on $\mathfrak{B}$. For any such c, since $cC_y : (y, \forall S)R$ is *true* on $\mathfrak{B}$, cC and $(c, \forall S)R$ are *true* on $\mathfrak{B}$. And again, for any instance d of S on $\mathfrak{B}$, $(c, d)R$ is *true* on $\mathfrak{B}$.

On $\mathfrak{B}$, if b is an instance of $C_y : (y, \forall M)R$, then both bC and $(b, \forall M)R$ are *true* on $\mathfrak{B}$. So for any instance e of M on $\mathfrak{B}$, $(b, e)R$ is *true* on $\mathfrak{B}$. Now, if d is an instance of S on $\mathfrak{B}$, from the first premise of (73), $\forall SM, dM$ is also *true* on $\mathfrak{B}$, and therefore $(b, d)R$ is *true* on $\mathfrak{B}$. So $(b, \forall S)R$ is also *true* on $\mathfrak{B}$. Since bC

is also *true*, $bC_y:(y,\forall S)R$ is *true* on $\mathfrak{B}$. But we saw that $(a,\forall C_y:(y,\forall S)R)F$ is *true* on $\mathfrak{B}$. So $(a,b)F$ is *true* on $\mathfrak{B}$.

□

We see that inference (64) can be incorporated in an existing powerful version of Quarc. Moreover, in the process, Quarc has brought to light two features of Moss's original formulation which needed to be addressed: completion of an ellipsis and making two presuppositions explicit. We therefore proved here a revised inference, (73).

2.7 *Comparative Quantifiers*

(74) More students than professors run.
More professors than deans run.
? More students than deans run.

The four kinds of inference we discussed above did not pose serious issues for their incorporation in Quarc, syntactically, semantically, or proof-theoretically. The active–passive-voice distinction and defining clauses were already incorporated in Quarc, the latter in a three-valued version of it; and conjunctive predicates and comparative adjectives required rather straightforward extensions for their incorporation. Comparative quantifiers, however, pose several challenges, only some of which will be met in this paper.

The quantifiers of Quarc, ∃ and ∀, translate natural language's "some," "a," "all," "any" and "every" in various of their uses. All these quantifiers are *unary determiners*: they attach to one general noun to form a noun phrase. "Some boys," "a girl," "all men," "any woman" and "every person" are a few examples. This is also true of some other natural language quantifiers, for instance *three*, *at least seven*, *infinitely many*, *most* and *many*. Translating these quantifiers in Quarc will require additional vocabulary but not additional syntactic roles.

By contrast, comparative quantifiers, in their use exemplified in (74), are *binary* determiners: they attach to *two* general nouns to form a noun-phrase. As, for instance, in "more *students* than *professors*" and "more *professors* than *deans*" (Ben-Yami 2009). Translating them into Quarc will therefore necessitate an additional syntactic role: a quantifier which attaches to an ordered pair of one-place predicates to form a quantified argument.

doi: 10.48106/dial.v74.i2

2.7.1 Vocabulary and Formulas

We add a new *binary quantifier*, Π, read "more." If P and Q are one-place predicates, then $\Pi(P, Q)$ is a *binary quantified argument*.

2.7.2 Semantics

To capture the truth-conditions of "more" within a truth-valuational substitutional semantics, as well as those of many other, unary quantifiers—e.g. "three," "at least seven," "many" and "most"—we should overcome a difficulty related to the fact that several names might name the same thing (Lewis 1985). Suppose we defined "Two men married Olivia Langdon" as true if there are two different substitution instances of names for "two men," each verifying "x is a man," which yield a true sentence of the form, "x married Olivia Langdon." We would then get that the sentence is true, since both "Mark Twain is a man" and "Samuel Clemens is a man" are true, as are "Mark Twain married Olivia Langdon" and "Samuel Clemens married Olivia Langdon." Yet Mark Twain *is* Samuel Clemens, and only this single man married Olivia Langdon.

To overcome this difficulty, we first define for each one-place predicate P on each valuation $\mathfrak{V}$ a *maximal substitution set* $\mathfrak{S}$. This is a set for which,

- only names a for which aP is *true* on $\mathfrak{V}$ are in $\mathfrak{S}$.
- for any different a and b in $\mathfrak{S}$, $a = b$ is *false* on $\mathfrak{V}$
- for any c for which cP is *true* on $\mathfrak{V}$, $a = c$ is *true* on $\mathfrak{V}$ for some a in $\mathfrak{S}$, possibly c itself.

In this way we make sure that every P is counted exactly once, so to say, by the names in P's maximal substitution set. It is easy to show that on each valuation, all maximal substitution sets of a given predicate have the same number of members, or cardinality.

We can now define the truth value of a formula $\phi[\Pi(P, Q)]$, governed by $\Pi(P, Q)$, on a valuation $\mathfrak{V}$. We consider two maximal substitution sets $\mathfrak{S}_P$ and $\mathfrak{S}_Q$. $\phi[\Pi(P, Q)]$ is *true* on $\mathfrak{V}$ just in case more substitution cases of the form, $\phi[a/\Pi(P, Q)]$ with $a \in \mathfrak{S}_P$ are *true* on $\mathfrak{V}$ than such substitution instances with $a \in \mathfrak{S}_Q$.

Turning to inference (74), we can formalise it and show the validity of the formalisation in Quarc. Its formalisation will be,

(75) $(\Pi(S, P))R, (\Pi(P, D))R \vDash (\Pi(S, D))R$

We have to show that if both premises are *true* on a valuation $\mathfrak{V}$, then so is the conclusion. We choose three maximal substitution sets on $\mathfrak{V}$, $\mathfrak{S}_S$, $\mathfrak{S}_P$ and $\mathfrak{S}_D$. If $(\Pi(S,P))R$ is *true* on $\mathfrak{V}$, then there are more members a in $\mathfrak{S}_S$ for which aR is *true* on $\mathfrak{V}$ than members b in $\mathfrak{S}_P$ for which bR is *true* on $\mathfrak{V}$; and similarly, there are more such members b than members c of $\mathfrak{S}_D$ for which cR is *true* on $\mathfrak{V}$. So there are more members a in $\mathfrak{S}_S$ for which aR is *true* on $\mathfrak{V}$ than members c in $\mathfrak{S}_D$ for which cR is *true* on $\mathfrak{V}$. Accordingly, $(\Pi(S,D))R$ is *true* on $\mathfrak{V}$.

2.7.3 Proofs

This is the part of the challenge comparative quantifiers pose which will not be met in this work. How is it possible to reflect the logic of the quantifier Π in a proof system, is a question we shall here leave unanswered. In fact, even the more basic question, *whether* it is possible to capture content by form for Π in argument–predicate sentences, will not be addressed here either.

To the best of my knowledge, Moss does not try to incorporate inference (74) or the quantifier "more," as used in *argument–predicate* sentences, in his Natural Logic systems (but see below on the use of this quantifier in *'existential'* sentences). In Moss (2015), he mentions inference (74) in order to show the apparent inadequacy of first-order logic as a means of representing the logic of natural language:

> [In] the first-order language with one-place relations student(x), professor(x), and run(x), there is no first-order sentence ϕ with the property that for all (finite) models M, ϕ is true in M if and only if "More students than professors run" is true in M in the obvious sense. This failure already suggests that first-order logic might not be the best "host logical system" for natural language inference. (2015, 563)

I agree with Moss on what he takes this inability to suggest. (See also Ben-Yami 2009 for a discussion of generalised quantifiers and comparative quantifiers.) What we managed to show in this paper is that Quarc does not have this shortcoming as a system for representing the logic of natural language. Quarc can incorporate natural language's comparative quantifiers as binary quantifiers, imitating their natural language syntax, and it does that by providing the correct truth conditions for these sentences. We saw this being done for "more" with a truth-valuational substitutional semantics; the way to generalise this

doi: 10.48106/dial.v74.i2

approach to other comparative quantifiers (e.g. “at least as many”) or construct a model-theoretic semantics for them is straightforward. Accordingly, we have managed to show an advantage of Quarc over the Predicate Calculus in this respect.

2.7.4 Comparative Quantifiers in “Existential” Sentences

In more recent work, Moss and Topal extended Natural Logic and applied it to sentences of the form, “There are at least as many *p* as *q*” and “There are more *p* than *q*” (2016; 2020) (see fn. 4). They have developed sound and complete proof systems for cardinality comparisons, for both finite (Moss 2016) and infinite sets (Moss and Topal 2020). This is impressive work, and it would be interesting to inquire whether Quarc can deliver anything comparable. This, however, will not be attempted in this paper, for several reasons.

There are obvious space considerations. For instance, the proof system of Moss (2016) contains 24 rules, of which 16 involve his formalisations of “at least as many” and “more”; the corresponding numbers for the proof system of Moss and Topal (2020) are 21 and 12. Accordingly, a Quarc system formalising these inferences might involve significantly more additions than the extended systems considered above. Similarly, a completeness proof for this extended system would not be established by minor additions to the one provided in Ben-Yami and Pavlović (2022). This is a topic for a separate paper.

Moreover, a Quarc treatment of sentences of the form, “There are at least as many *p* as *q*” and “There are more *p* than *q*,” will depart from Moss’s in some important fundamental respects. Moss formalises these sentences by sentences similar in form to those formalising “All/some *p* are/aren’t *q*.” For instance, “Some *p* are *q*” is formalised by $\exists(p, q)$, and “There are more *p* than *q*” by $\exists^{>}(p, q)$. Namely, apart from the different quantifier, no syntactic distinction is drawn between the argument–predicate sentence, “Some *p* are *q*,” in which the argument is “some *p*,” and the so-called *existential sentence*, “There are *x*,” in which *x* is a noun phrase formed by a comparative quantifier, “more *p* than *q*.” However, the existential sentence, “There are more *p* than *q*” is no argument–predicate one. A sentence similar to it in form using the quantifier “some” will be, “There are some *p*,” and not, “Some *p* are *q*.” An argument–predicate sentence with the quantifier “more” would have the form of the sentence considered above, “More students than professors run.” As mentioned earlier, Moss hasn’t developed a proof system for *these* sentences.

The distinction between existential sentences and argument–predicate sentences seems to be a linguistic universal. Moreover, existential sentences show important differences from quantified argument–predicate ones (Ben-Yami 2004, sec. 6.5; Francez, I. 2009; McNally 2011). Accordingly, a system that aims to be a logic for natural language informed by the latter's syntax should formalise existential sentences differently than it does argument–predicate ones. It should distinguish the two constructions and explore the logical relations between them. As part of such a general treatment of existential sentences, those with a noun-phrase of the form "more *p* than *q*" as their pivot (see Francez, I. 2009; McNally 2011 for the terminology) can also be introduced and discussed, as well as those with other comparative quantifiers. A general inquiry into the logic and formalisation of existential sentences has not been attempted by Moss and shall not be attempted here either.

3 Conclusions and Future Work

This paper tried to assess the ability of Quarc, in its current or extended versions, to represent the kinds of inference which have served as the basis of Moss's constructions of Natural Logic systems. We have shown how Quarc can incorporate, sometimes with some extensions, passive–active voice distinctions, conjunctive predicates (*see and hear*), comparative adjectives (*taller*), and defining clauses (*who respect all mammals*). All these were incorporated within sound and complete systems. We have also shown how Quarc can be syntactically extended to incorporate comparative quantifiers (*more … than …*) and provided a semantics but not a proof system for this extension.

All this was done by using a language with a syntax close to that of natural language. In this respect we followed Moss's dictum for his Natural Logic project, "logic for natural language, logic in natural language" (2015, 563). I believe that in some respects we improved on Natural Logic, for instance by not using negative nouns.

The process also helped shed light on some of the inferences we discussed. The constraints of the formal system brought us to recognise an ellipsis and presuppositions involved in the conclusion of inference (64), "All who fear all who respect all skunks fear all who respect all mammals."

A main aim of the Natural Logic project which we did not address here was the question of decidability. Apart from the theoretical interest, this is relevant to questions of the applicability of computer programmes for determining validity. I hope this question will be addressed in future work, by myself or others.

doi: 10.48106/dial.v74.i2

Another topic which was not addressed in this paper but which has engaged Natural Logic is that of *monotonicity* (Moss 2015, sec. 4). Moss's work is based on van Benthem's (1986, 1991), which generated additional inquiries as well (see van Benthem 2008 for a historical survey). Whether and how can Quarc analyse the phenomena of monotonicity is again left for future work.

The last topic mentioned as subject for future work is the formalisation of the so-called existential sentences—"There are *x*"—in Quarc. Once this is done, existential sentences with comparative quantifiers—"There are more *p* than *q*" and "There are at least as many *p* as *q*"—can also be formalised, and Moss's work on these last sentences can be comparatively studied.

So, there is still work to be done. Yet hopefully, we have shown that in addition to the earlier successes in its application to the analysis of the logic of natural language, Quarc can also represent the inferences that motivated Moss's Natural Logic.*

Hanoch Ben-Yami
0000-0002-4903-854X
Central European University
benyamih@ceu.edu

References

VAN BENTHEM, Johan. 1986. *Essays in Logical Semantics*. Studies in Linguistics and Philosophy n. 29. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-009-4540-1.

—. 1991. *Language in Action: Categories, Lambdas, and Dynamic Logic*. Studies in Logic and the Foundations of Mathematics n. 130. Amsterdam: North-Holland Publishing Co.

—. 2008. "Natural Logic: A View from the 1980s." in *Logic, Navya-Nyāya & Applications. Hommage to Bimal Krishna Matilal*, edited by Mihir Kumar CHAKRABORTY, Benedikt LÖWE, Madhabendra NATH MITRA, and Sundar SURAKKAI, pp. 21–42. Studies in Logic n. 15. London: College Publications.

BEN-YAMI, Hanoch. 2004. *Logic & Natural Language: On Plural Reference and Its Semantic and Logical Significance*. Farnham, Surrey: Ashgate, doi:10.4324/9781315250267.

* The research leading to this work was supported by the funding programme, Research Stays for University Academics, 2018 (57381327), of the German Academic Exchange Service (DAAD), and by the Senior Fellowship programme of the Edelstein Center, The Hebrew University of Jerusalem.

—. 2009. “Generalized Quantifiers, and Beyond.” *Logique et Analyse* 52(208): 309–326, https://logiqueetanalyse.be/archive/issues87-220/LA208/01ben-yami.pdf.

—. 2014. “The Quantified Argument Calculus.” *The Review of Symbolic Logic* 7(1): 120–146, doi:10.1017/s1755020313000373.

Ben-Yami, Hanoch and Pavlović, Edi. 2022. “Completeness of the Quantified Argument Calculus on the Truth-Valuational Approach.” in *Human Rationality: Festschrift for Nenad Smokrović*, edited by Boran Berčić, Aleksandra Golubović, and Majda Trobok, pp. 53–78. Rijeka: Faculty of Humanities and Social Sciences, University of Rijeka, urn:nbn:hr:186:273617.

Cobreros, Pablo, Egré, Paul, Ripley, David and van Rooij, Robert. 2013. “Reaching Transparent Truth.” *Mind* 122(488): 841–866, doi:10.1093/mind/fzt110.

Fitch, Frederic B. 1952. *Symbolic Logic: An Introduction*. New York: Ronald Press Co.

Francez, Itamar. 2009. “Existentials, Predication, and Modification.” *Linguistics and Philosophy* 32(1): 1–50, doi:10.1007/s10988-009-9055-4.

Francez, Nissim. 2014. “A Logic Inspired by Natural Language: Quantifiers As Subnectors.” *Journal of Philosophical Logic* 43(6): 1153–1172, doi:10.1007/s10992-014-9312-z.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle a.S.: Louis Nebert.

Geach, Peter Thomas. 1962. *Reference and Generality, an Examination of Some Medieval and Modern Theories*. Ithaca, New York: Cornell University Press. Third edition: Geach (1980).

—. 1980. *Reference and Generality, an Examination of Some Medieval and Modern Theories*. 3rd ed. Ithaca, New York: Cornell University Press. First edition: Geach (1962).

Halldén, Sören. 1949. *The Logic of Nonsense*. Uppsala: Uppsala Universitet.

Henkin, Leon. 1949. “The Completeness of the First-Order Functional Calculus.” *The Journal of Symbolic Logic* 14(3): 159–166, doi:10.2307/2267044.

Jaśkowski, Stanisław. 1934. “On the Rules of Suppositions in Formal Logic.” *Studia Logica (wydawnictwo poświęcone logice i jej historji)* 1: 5–32. Reprinted in McCall (1967, 232–258), with considerable change in notation, https://www.logik.ch/daten/jaskowski.pdf.

Lanzet, Ran. 2017. “A Three-Valued Quantified Argument Calculus: Domain-Free Model-Theory, Completeness, and Embedding of FOL.” *The Review of Symbolic Logic* 10(3): 549–582, doi:10.1017/s1755020317000053.

Lanzet, Ran and Ben-Yami, Hanoch. 2004. “Logical Inquiries into a New Formal System with Plural Reference.” in *First-Order Logic Revisited*, edited by Vincent F. Hendricks, Fabian Neuhaus, Stig Andur Pedersen, Uwe Scheffler, and Heinrich Theodor Wansing, pp. 173–224. Logische Philosophie n. 12. Berlin: Logos Verlag.

Lappin, Shalom, ed. 1996. *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell Publishers. Second edition: Lappin and Fox (2015).

Lappin, Shalom and Fox, Chris J., eds. 2015. *The Handbook of Contemporary Semantic Theory*. 2nd ed. Hoboken, New Jersey: John Wiley and Sons, Inc. First edition: Lappin (1996), doi:10.1002/9781118882139.

Lemmon, Edward John. 1965. *Beginning Logic*. London: Thomas Nelson and Sons Ltd.

Lewis, Harry A. 1985. "Substitutional Quantification and Nonstandard Quantifiers." *Noûs* 19(3): 447–451, doi:10.2307/2214953.

McCall, Storrs, ed. 1967. *Polish Logic 1920–1939*. Oxford: Oxford University Press.

McNally, Louise. 2011. "Existential Sentences." in *Semantics: An International Handbook of Natural Language Meaning. Volume 2*, edited by Klaus von Heusinger, Claudia Maienborn, and Paul H. Portner, pp. 1829–1848. Handbooks of Linguistics and Communication Science n. 33.2. Berlin: de Gruyter Mouton, doi:10.1515/9783110255072.1829.

Moss, Lawrence S. 2010a. "Syllogistic Logics with Verbs." *Journal of Logic and Computation* 20(4): 947–967, doi:10.1093/logcom/exn086.

—. 2010b. "Logics for Two Fragments beyond the Syllogistic Boundary." in *Fields of Logic and Computation. Essays Dedicated to Yuri Gurevich on the Occasion of His 70th Birthday*, edited by Andreas Blass, Nachum Dershowitz, and Wolfgang Reisig, pp. 538–564. Berlin: Springer, doi:10.1007/978-3-642-15025-8_27.

—. 2010c. "Natural Logic and Semantics." in *Logic, Language and Meaning. 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009. Revised Selected Papers*, edited by Maria Aloni, Harald Andreas Bastiaanse, Tikitu de Jager, and Katrin Schulz, pp. 84–93. Berlin: Springer, doi:10.1007/978-3-642-14287-1_9.

—. 2011. "Syllogistic Logics with Comparative Adjectives." *Journal of Logic, Language, and Information* 20(3): 397–417, doi:10.1007/s10849-011-9137-x.

—. 2015. "Natural Logic." in *The Handbook of Contemporary Semantic Theory*, edited by Shalom Lappin and Chris J. Fox, 2nd ed., pp. 561–592. Hoboken, New Jersey: John Wiley and Sons, Inc. First edition: Lappin (1996), doi:10.1002/9781118882139.ch18.

—. 2016. "Syllogistic Logic with Cardinality Comparisons." in *J. Michael Dunn on Information Based Logics*, edited by Katalin Bimbó, pp. 391–416. Outstanding Contributions to Logic n. 8. Cham: Springer, doi:10.1007/978-3-319-29300-4_18.

Moss, Lawrence S. and Topal, Selçuk. 2020. "Syllogistic Logic with Cardinality Comparisons, on Infinite Sets." *The Review of Symbolic Logic* 13(1): 1–22, doi:10.1017/S1755020318000126.

Pavlović, Edi. 2017. "The Quantified Argument Calculus: An Inquiry into Its Logical Properties and Applications." PhD dissertation, Budapest: Central European University, https://www.etd.ceu.edu/2017/pavlovic_edi.pdf.

PAVLOVIĆ, Edi and GRATZ, Norbert. 2019. "Proof-Theoretic Analysis of the Quantified Argument Calculus." *The Review of Symbolic Logic* 12(4): 607–636, doi:10.1017/s1755020318000114.

PRATT-HARTMANN, Ian and MOSS, Lawrence S. 2009. "Logics for the Relational Syllogistic." *The Review of Symbolic Logic* 2(4): 647–683, doi:10.1017/s1755020309990086.

RAAB, Jonas. 2018. "Aristotle, Logic, and QUARC." *History and Philosophy of Logic* 39(4): 305–340, doi:10.1080/01445340.2018.1467198.

SOMMERS, Fred. 1982. *The Logic of Natural Language*. New York: Oxford University Press.

STRAWSON, Peter Frederick. 1950. "On Referring." *Mind* 59(235): 320–344. Reprinted, with some added footnotes, in Strawson (1971, 1–27) and in Strawson (2004, 1–20), doi:10.1093/mind/LIX.235.320.

—. 1952. *Introduction to Logical Theory*. London: Methuen & Co.

—. 1971. *Logico-Linguistic Papers*. London: Methuen & Co. Reprinted as Strawson (2004), doi:10.4324/9781351153607.

—. 2004. *Logico-Linguistic Papers*. 2nd ed. London: Routledge, doi:10.4324/9781315250250.

Reflective Equilibrium on the Fringe

The Tragic Threefold Story of a Failed Methodology for Logical Theorising

Bogdan Dicher

Reflective equilibrium, as a methodology for the "formation of logics," fails on the *fringe*, where intricate details can make or break a logical theory. On the fringe, the process of theorification cannot be methodologically governed by anything like reflective equilibrium. When logical theorising gets tricky, there is nothing on the pre-theoretical side on which our theoretical claims can reflect of—at least not in any meaningful way. Indeed, the fringe is exclusively the domain of theoretical negotiations and the methodological power of reflective equilibrium is merely nominal.

Reflective equilibrium has been proposed as a methodology for logical theorising and, indeed, as a procedure for justifying our logical knowledge at least since Goodman's "new riddle of induction."[1]

In recent years, interest in it resurged, particularly in the wake of the advances of the *anti-exceptionalist* programme in logic. The general background for this paper will be given by a modest form of anti-exceptionalism, compatible with logical immanentism—the view that logic is immanent in language (see e.g. Brandom 2000)—which claims that the epistemology of logics is fallibilist (see e.g. Peregrin and Svoboda 2013, 2016, 2017; Read 2000).[2]

In this paper, I will argue against the thesis that reflective equilibrium is a viable methodology for logical theorising. This negative thesis does not deny that the *phenomenology* of logical inquiry could be described, at least in part, in accordance to the pattern provided by reflective equilibrium (hereafter often abbreviated as "*RE*"). This I gladly grant and duly deplore, for I believe

1 In Goodman (1955). The name, of course, is of a later date, being first used in Rawls (1971).

2 Full-blooded anti-exceptionalism is, roughly, the view that logic is not special, but rather contiguous with the empirical sciences (Hjortland 2017; Priest 2014; Russell 2014; Williamson 2007).

that, ultimately, it is the plausibility of this way of describing logical inquiry that is at the core of the misguided tenet that *RE* is a meaningful methodology for logic. Instead, my claim is that the processes normally associated with logical investigations are too complex, too abstract, and too "theoretical" to be in any *substantive* sense guided by *RE*. I will present my arguments against reflective equilibrium via three case studies of currently debated issues among logicians. These vignettes will, I hope, drive home the following three points:

- The first is that logical theorising is systematically biased in favour of theoretical considerations and so *RE* is, *qua* methodology, too weak.
- The second is that *RE* underdetermines both the identification of the specific problems one encounters in "the formation of logics," i.e. problematisation, and the problem-solving process itself.
- The third and final point I wish to make is that *RE* systematically favours weaker logics.

1 Reflective Equilibrium

So what is reflective equilibrium? In its most exalted sense, it is the ultimate justification procedure open to some of our beliefs, including our logical beliefs. In a more modest sense, it is a methodology in processes like formalisation, theorification, modelling, etc. These two senses of *RE* are connected and it takes but a small (up and ahead) step from the latter to the former. Both are evident in a celebrated remark of Goodman's, worth reproducing here *in extenso*:

> Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields inacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences. This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are*

> *unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* [...] [I]n the agreement achieved lies the only justification needed for either. (1955, 63–64)

Much of what I have to say will target *RE qua* methodology. This is because I take it that whatever problems beset it in this quality, also affect its status as a state that justifies a body of beliefs: *RE* is supposed to generate an eponymous doxastic state in which one's logical beliefs are justified. But if the process does not warrant the cogency of its outcomes, then what value can there be to either? A state of *RE* may be seen as one where no further developments of one's theories is possible because there are no more apparent problems to resolve.[3] Yet the same situation could ensue as an effect of lack of curiosity, of having a deficit of imagination, or low epistemic standards. This kind of epistemic "tranquillity" is a non-specific symptom. Insofar at is has any value, this is due to the inherent virtues of the process that lead to it.

So what is this methodology? Goodman's original description refers only to inferences, principles of inference and the relation between them. But we may well suppose that articulating this relation involves a few more ingredients. So, expanding a bit on the original schematic proposal, we can easily get a *prima facie* plausible story that goes along the following lines: One starts with a body of inchoate, perhaps practical or intuitive, knowledge of a certain domain—for instance, that associated with the dispositions to infer manifested in the daily ratiocinative practice, or even that obtained by a modicum of reflection on the practice. That is, one starts with the knowledge expressed in pre- or quasi-theoretical claims like "this argument is valid," "that doesn't follow," or perhaps even "valid arguments are truth-preserving," etc. Call this "*1-knowledge*."[4]

This body of pre-theoretical knowledge is apt for further regimentation, precisification and expansion—by fine-tuning the conceptual apparatus behind it, by discovering novel, perhaps more abstract or more general, relations between its objects, by forming new hypotheses, proving general statements,

3 This is a somewhat implausible contention, as it is not clear how, for instance, the effort to achieve a *simpler* theory could be massaged into the simple picture of *RE*. But let us grant it for the sake of the argument.

4 I do not wish to attach any precise philosophical sense to the word "knowledge." Instead, it is to be taken in the intuitive sense. To the extent that it is explicit knowledge, it consists of both statements (factive, prescriptive, normative, etc.) and the conceptual apparatus (predicates, relations, etc.) underlying them. However, I am not assuming that this knowledge must be explicit; it can well be, at least partly, knowledge-how.

doi: 10.48106/dial.v74.i2

etc. Thus, one moves from the knowledge that a particular item is an argument to a general account of what arguments are, from the belief that valid arguments preserve truth to beliefs like "valid deductive arguments preserve designated value on Tarskian models," etc. Call (all) this "*2-knowledge*."

The development and refinement of *2-knowledge*—or, in one word, *theorification*—proceeds and is kept in check by balancing it against *1-knowledge*. Theoretical pronouncements are measured against the pre-theoretical knowledge that inspired them in the first place. For instance, a rather bad putative definition of *argument* as "speech in which, out of two given things, a third follows" is suitably modified upon realising that many (things that are usually called) arguments have more or less than two premises (given things) and may well derive a conclusion (third thing) that is, in fact, identical to (one of) the premise(s).

At the same time, *1-knowledge* is, at least potentially, modifiable in light of *2-knowledge*. For instance, it may be that *1-knowledge* does not provide for a distinction between inductive and deductive arguments (though maybe it could), whereas *2-knowledge* does. This theoretical distinction may inform *1-knowledge* and we may see hosts of savvy informal reasoners resorting to it in everyday contexts. Or it may be that pre-theoretically we are disposed to infer in accordance with a certain form of argument but, in virtue of general principles of validity developed as part of *2-knowledge*, we come to see that this is not the case (cf. *infra*, the discussion of the ω-rule for an illustration of this case.)

Our logical theories and, with them, logical knowledge, are obtained and justified as a result of this trade-off between pre-theoretical and theoretical beliefs.[5]

2 Formalisation and the Formation of Logics

Goodmanian reflective equilibrium seems to presuppose a non-conventionalist view of logic. At any rate, it is easier to grasp the problems of *RE* if we assume, without loss of generality, such a view. Recall Carnap's famous *principle of tolerance*:

> In logic there are no morals. Everyone is at liberty to build his own logic, i.e. his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must

5 For a more detailed discussion of the method see the opinionated survey in Cath (2016).

> state his methods clearly, and give syntactical rules instead of philosophical arguments. (1937, sec. 17)

For Carnap, the standard for the success of logics is not the extent to which they "correspond" to natural language, the medium of human reasoning, but rather their usefulness relative to the purposes for which they were designed.

Not so for the view that will provide the background for the present discussion. On it, the relation between natural language and the logical formalism must go beyond the latter's usefulness in analysing the former. For specificity's sake, let our underlying view of logic be that it is obtained via a process of *formalisation*, understood as "a kind of extraction [...] of logical form" out of natural language (Peregrin and Svoboda 2016, 4)—see also Peregrin and Svoboda (2013, 2017).[6]

The image suggested by *RE* is readily seen to fit some scenarios of "formalisation" which are marked by but two parameters:

1. An informal argument like (arg): "Socrates is mortal because all men are mortal."
2. A target logical system (e.g. first-order logic) or perhaps merely a target logical syntax (e.g. Fregean syntax, by which I mean the sort of syntax that explicitly features sentential operators and construes atomic declarative sentences as having function-argument from, as opposed to, say, subject-predicate form).[7]

Suppose now that we go about formalising (arg) in the Fregean syntax—our target (tar). We already know its syncategoremata: expressions like "all," "some," the (grammatical) conjunctions "and," "or," "if … then," etc. We also know, by and large, how to deal with them in (tar). All in all, we could arrive at the following schematic rendering of (arg):

$$\frac{\forall x Mx}{Ms}$$

of which we make sense via a key that says that "*M*" stands for *mortal*, "*x*" is a variable ranging over the extension of "man," and "*s*" an individual constant, standing for *Socrates*.

6 For an alternative account of formalisation, see Brun (2014). For a monographic analysis of the many problems raised by this deceptively simple concept, see Brun (2004).

7 This is not inconsistent with the Peregrin-Svoboda view of formalisation, as the "target" need not be thought of as being antecedently available. It can be just as well be "extracted" in the process of formalisation.

doi: 10.48106/dial.v74.i2

It's no achievement to see that this is a suboptimal—indeed, plainly wrong—formalisation of (arg). For one thing, "All men are mortal" was rendered formally rather dumbly. For instance, *man* and *mortal* were placed in distinct grammatical categories. Not only is this unpleasantly non-uniform, but it also obscures the predicate status of *man*. We would do better to render this premise as "$\forall x(Wx \to Mx)$," with "W" standing for *man* and x ranging over a (generic) class of objects. (Note that this is already a good step away from the"surface" grammar of English.) So we get an improved rendering of (arg), namely:

$$\frac{\forall x(Wx \to Mx)}{Ms}$$

the validity of which we check in (tar).[8] Obviously, it is not.

Does this mean that the conclusion of (arg) does not follow logically from the premise? Well, yes, it does mean that; still, we wouldn't want to say that "Socrates is mortal" may be false when "All men are mortal" is true. In this sense, we would not want to revise our commitment to (arg). We figure out that we need another premise, "Socrates is a man," in order to validate both (arg) and its formalisation.

And so on and so forth: I am not particularly bent on boring the reader with logical trivia. The salient point is that all this happens within the confines of a more or less precise target formalism. At this level, of *formalisation*, it is quite plausible to see our endeavours as governed by *RE*.

The *formation of logics*, to appropriate a term used by Peregrin and Svoboda (2016, 2017), is, as it were, the next level of formalisation-qua-extraction. One obtains a logic by making explicit (cf. Brandom 1994) *and* bringing together into a coherent ensemble the principles governing informal reasoning. No matter how generous our notion of formalisation is, this is no *mere* formalisation, as a few examples will show.

Consider first the case of a working mathematician who believes, in the first instance, that the ω-rule:

$$\frac{P(0) \qquad P(1) \qquad \dots \qquad P(n) \qquad \dots}{\forall x(x \in \mathbb{N} \to Px)}$$

8 Actually, since (tar) is rather imprecise, the validity check would have to be performed in a logic based on the Fregean syntax or, at the very least, in a fragment of such a logic that contains enough information about →, ∀, and the horizontal "inference" line that ended up rendering "because."

is *logically* valid.

Subsequently, and in light of various *2-knowledge* beliefs—inference rules are finitary, logic is topic-neutral, “natural number” does not express a logical property, logicism fails because of Russell’s paradox, etc.—she changes her mind and decides not only that the ω-rule is not part of logic, but also that its syntactic structure, and in particular its infinite number of premises, make it not an inference rule at all.[9]

Take now Peano’s axiom of induction. Its natural formulation involves quantification over properties:

$$\forall P(P(0) \wedge \forall n(P(n) \rightarrow P(n+1)) \rightarrow \forall n P(n))$$

For various (theoretical) reasons, this kind of formalisation was thought best to be avoided and first-order logic, in which the quantifiers range only over individuals, became the norm (for more on this, see Eklund 1996). The demise of second order formalisms has little to do with what goes on in natural language, where (apparent) quantification over properties is certainly present. It was and, to the extent that the controversy is alive, it still is a matter of deploying heady theoretical considerations.[10] Languages may carry logics inside them, but it is still up to the logicians to decide what to bring to the surface and how.

A third example will also illustrate the fact that, in many cases, the practice is not at all coherent and it cannot light our way in a simple fashion. Take the following rules governing a truth predicate T:

$$\frac{A}{T\langle A\rangle}\ T\text{-I} \qquad\qquad \frac{T\langle A\rangle}{A}\ T\text{-E}$$

They seem innocuous enough. But add some equally innocuous reasoning principles and pick the sentence named by $\langle A\rangle$ so that it is “This sentence is false” and all hell breaks loose, i.e. any sentence follows from any sentence.[11] Deciding how to handle these issues significantly exceeds what can be reasonably characterised as a process of formalisation.

Thus, *in practice* the formation of logics is a rough-going process of theorification responsible to the pre-formal practice, informed by it and, allegedly

9 This example may also serve to illustrate the modification of *1-knowledge* in virtue of *2-knowledge* discussed at the end of the previous section.

10 Famously, Quine rejected second-order logic as set theory “in sheep’s clothes” (1970, 66). But the same logic was forcefully defended by Shapiro, S. (1991).

11 For more on this, see below, section 4.

doi: 10.48106/dial.v74.i2

at least, placed under its control to a certain extent. The process goes beyond simple formalisation and is not at all unproblematic.

RE is meant to guide us on the righteous path of smoothing out these asperities and forming a justified logic, by debunking whatever tensions may arise between *1-* and *2-knowledge*. Can it really do this? I think not and in the next three sections, I will explore three cases of current logical debates, consideration of which will explain why I am sceptical about the promises of *RE*.

3 Case Study no.1: Multiple Conclusions

Orthodox logical theorising (Dummett 1991; Steinberger 2011) teaches that an argument has one or more premises and only one conclusion. In this it is faithful to the practice, insofar as it appears that natural language arguments have but one conclusion. At the same time, inferences of the form:

$$\frac{\neg\neg A}{A}\ \text{DNE}$$

are generally accepted in the daily ratiocinative practice. That is, one tends to accept inferences by *double negation elimination* (DNE).

As it turns out, these pre-theoretical commitments stand in an uneasy tension, albeit one that needs a rather sophisticated background theory to surface fully. This background theory is a version of logical inferentialism, better known as *proof-theoretic semantics* (Prawitz 1965, 1974; Schröder-Heister 2018; Francez 2015), whose roots can be traced back to Gentzen (1935). Proof-theoretic semantics theorists hold that the meaning of the logical operators is determined by the primitive rules of inference that govern how sentences in which they feature as principal operators are, respectively, introduced and eliminated from proofs. These two kinds of rules for an operator must match; to put it in jargon: they must be in *harmony* (Dummett 1991). If harmony does not obtain, then the operator is illegitimate and so is the inferential behaviour it sanctions. Moreover, the test for the "match" between the introduction and elimination rules is syntactic in nature. There must be a syntactically assessable property the obtaining of which witnesses the harmonious character of the pairing.[12]

12 This is why proof-theoretic semantics is salient for spotting the aforementioned tension: It requires meaning explanations to proceed in terms of syntactical properties against the background of the

DNE is obviously an elimination rule for negation. The corresponding introduction rule is the (intuitionistic) *reductio ad absurdum*:

$$\dfrac{\begin{array}{c}[A]_j \\ \vdots \\ \neg A\end{array}}{\neg A}\ \text{iRAA}, j$$

It turns out that these two rules cannot be harmonised *if* arguments (and the formal proofs representing them) are single-conclusion. A familiar, if bitterly contested, account of harmony has it that a set of introductions and eliminations for a logical constant is harmonious only if its addition to a proof system is conservative (Dummett 1991).[13] That is, to the extent that the addition generates new valid arguments, then these must involve the novel vocabulary. Famously, Peirce's law

$$((A \to B) \to A) \to A$$

despite containing only one logical operator, the conditional, is not provable in intuitionistic logic. *A fortiori*, it is not provable using only the rules for the conditional. However, once one adds DNE to intuitionistic logic—thus ensuring that negation behaves classically—there is a proof of it. (I leave the construction of the proof as an exercise for the reader.) It follows from this that classical negation is not harmonious. The strongest correct rules for negation are those of intuitionistic logic.

But this holds water only if arguments and the formal proofs representing them are single-conclusion. Only in this case does classical negation yield a nonconservative extension of intuitionistic logic. If multiple conclusions are allowed, classical negation is conservative and hence harmonious. In such systems there are proofs of Peirce's law in the implicational fragment alone:

rules used and the structure of the proofs. On truth-conditional approaches to the meaning of the logical terms, the syntax of the proof system matters not at all. The behaviour of the logical operators is determined by their truth conditions and it is plain that, at least if one assumes a bivalent notion of truth, there is no way of making A false when $\neg\neg A$ is true. That's the end of the story: whether this behaviour is best tracked by a single- or a multiple-conclusion proof system is irrelevant for the validity of DNE.

13 Not much hinges on this contested account of harmony. It features here because it is the best known. For a defence of it, see Dicher (2016); for criticism, see Read (2000). For a more recent proposal see Gratzl and Orlandelli (2017).

doi: 10.48106/dial.v74.i2

$$
\cfrac{\cfrac{\cfrac{\cfrac{[A]_1}{A, B}\,\text{Weakening}}{A, A \to B}\,{\to}\text{I}, 1 \qquad [(A \to B) \to A]_2}{\cfrac{A, A}{A}\,\text{C}}\,{\to}\text{E}}{((A \to B) \to A) \to A}\,{\to}\text{I}, 2
$$

Now let us find our way out of this, guided by *RE*. Assume that our background theory, i.e. the commitment to inferentialism and the account of harmony as conservativeness, is sacrosanct.[14]

The first thing to notice is that the tension we ought to resolve is not between the pre-formal practice and our theoretical commitments. Rather, it is a tension within the practice—albeit one that comes to the fore only against the background of a commitment to a proof-theoretic account of the meaning of the logical vocabulary.[15] It seems that in order to even be able to "reflect equilibristically" on the matter, one must antecedently form some reasonably justified theoretical beliefs about validity, the structure of proofs, etc. In other words, one needs (some theory in order) to *generate* a tension between *1-knowledge* and *2-knowledge*.[16]

On the flip side, this picture suggests that revisions that put in accord the practice with the theory—against the background of its more abstract pronouncements—are somehow inescapable. Alas, it seems to me that it also leads to the demise of *RE* as a *significant* methodological constraint in logical theorising: If we agree that any theory will mutilate in some way some aspects of the practice to which we would otherwise wish to remain faithful, then it follows that any and all resolutions of conflicts must, ultimately, do violence to the practice or, which amounts to the same thing, to *1-knowledge*. Note that the assumption made is not at all surprising, given that theorification

14 To be sure, this is a contentious assumption. I will say a bit more by way of motivating it in footnote 16.

15 For characterisations of *RE* involving the appeal to a background theory, see Brun (2004, 2013, 2014) and the references therein. Notice that Brun's "background theories" may be more encompassing than those described here.

16 But why would anyone do that? Why not outrightly modify the background theory so that there is no conflict? Presumably, that background theory, including its tension generating aspects, is not embraced idiosyncratically. One clings to it because it explains better other aspects of the practice one is theorising about. It is, in other words, the best theory one has thus far about the target practice. Besides, it is not a stretch to expect that modifications to the background theory will generate other tensions, pertaining perhaps to other parts of the practice. Indeed, it would be foolishly optimistic to expect otherwise.

presupposes a great deal of systematisation. In the particular scenario at hand and, consequently, in all scenarios relevantly analogous to it, it is indeed unavoidable, since the practice itself is less than coherent.

The moral of the story is that logical *facts*, as discernible in the vernacular ratiocinative practice, are fragile.[17] They are bound to succumb to the pressures exerted by needs peculiar to theorification or to its perceived benefits. Resolving conflicts is not so much a matter of finding some equilibrium between the practice and the theory, as it is a matter of finding a convenient excuse to obliterate the inconvenient aspects of the practice.

This may appear to blatantly contradict another problem raised with respect to *RE* by Woods (2019). Woods, following Wright (1986), accuses the procedure of suffering irremediably of the problem of "too many degrees of freedom." That is, it leaves open too many areas for revision, mainly with respect to what I have termed here the "background theory." In particular, even the beliefs that brought about the conflict may be subject to revisions. I believe that the contradiction is merely apparent. I've blocked that possibility and kept the background theory unchangeable precisely in order to avoid the degrees of freedom problem *because* I believe that Woods' diagnosis is correct in the absence of that assumption. Now we see that even with it *RE* fares less than stellarly.

One may argue that this does not go against *RE*, which does not require that the resolution of the conflicts be balanced, or "just," etc. All that *RE* requires is that we resolve the tensions between the practice and the theory, even if, as I have claimed, this will systematically ensue in the theory gaining the upper hand. But then it seems that *RE*, as a methodological requirement, amounts to little more than the injunction to pay *some* attention to the domain one is theorising about. This, of course, is a piece of eminently reasonable advice. It is also about as useful in guiding our investigations of that domain as the prophecies of the oracle of Delphi would be in planning one's future.

This, then, is the first complaint that I have against the thesis that *RE* is a meaningful guide to the formation of logics: that "real" equilibrium matters little for it, and that the process of achieving what we may call "internal" equilibrium, is heavily rigged in favour of theoretical considerations.

17 This is abundantly illustrated by the actual solutions to the problem of multiple conclusions; see Dicher (2020).

doi: 10.48106/dial.v74.i2

4 Case Study no.2: Which Logic is This?

I have already mentioned classical logic. Despite its many merits, few logicians expect classical logic to perform well in the presence of of paradox-generating vocabulary like vague predicates or transparent truth. But are they right in thinking this?

Contrary to these common beliefs, an impressive case has been put forward by Cobreros et al. (2012, 2013) on behalf of classical logic being able to handle the aforementioned troublesome vocabulary without degenerating into a trivial consequence relation (see also Ripley 2012, 2013). To be sure, this is classical logic in a particular and rather special guise—special enough to give it a name of its own: "*ST*," pronounced "strict-tolerant." Let us see us how classical logic and *ST* handle the paradoxes and in what sense the latter is classical.

Our starting point is Gentzen's sequent calculus for classical logic, *LK* (1935). Recall that this contains the Cut rule:

$$\frac{X : Y, A \qquad A, X : Y}{X : Y}$$

Now if one were to add e.g. the *T*-rules from above to *LK*, then the system would become trivial: any conclusion would follow from any premisses. To see this, let λ be a sentence such that $\lambda \equiv_{df} \neg T\langle\lambda\rangle$. Thus λ is the (strengthened) *Liar*: "This sentence is not true."[18]

Then we can derive the empty sequent:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{T\langle\lambda\rangle : T\langle\lambda\rangle}\,\text{Id}}{\neg T\langle\lambda\rangle : \neg T\langle\lambda\rangle}\,\neg\text{-L}, \neg\text{-R}}{\lambda : \lambda}\,\text{df}}{T\langle\lambda\rangle : \lambda}\,T\text{-L}}{: \neg T\langle\lambda\rangle, \lambda}\,\neg\text{-L}}{: \lambda}\,\text{df, Contraction} \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{T\langle\lambda\rangle : T\langle\lambda\rangle}\,\text{Id}}{\neg T\langle\lambda\rangle : \neg T\langle\lambda\rangle}\,\neg\text{-L}, \neg\text{-R}}{\lambda : \lambda}\,\text{df}}{\lambda : T\langle\lambda\rangle}\,T\text{-R}}{\neg T\langle\lambda\rangle, \lambda :}\,\neg\text{-R}}{\lambda :}\,\text{df, Contraction}}{:}\,\text{Cut}$$

from which in turn $A : B$ follows for any A, B via Weakening.

18 The truth predicate is essential for expressing λ, though it is not the only required ingredient. The name forming operator $\langle \ldots \rangle$ is equally important. For more technical details about this setup, including the matter of how to render λ expressible, see Ripley (2012).

Gentzen (1935) proved that Cut is eliminable from *LK* in the sense that any derivable *LK*-sequent is derivable without using Cut; hence *LK* and its cut-less variant, LK^-, are equivalent in that they derive the same sequents. Since in the above proof Cut is essential for deriving the troublesome empty sequent, we have two proof systems that, although equivalent in the absence of the truth predicate, behave differently when extended with the rules governing it.

LK^- can be used to formalise *ST*,[19] which has the same valid sequents as classical logic but allows for non-trivial and conservative extensions with the sort of vocabulary that generates troubles classically. Semantically, its consequence relation can be characterised by the strong Kleene valuations (Kleene 1952), given below for conjunction, disjunction and negation, when *A* follows from some premises (bundled in the set) *X* iff, whenever each of the statements in *X* has the value 1, the conclusion *A* has a value in {1,½}:[20]

$\wedge$	1	½	0
1	1	½	0
½	½	½	0
0	0	0	0

$\vee$	1	½	0
1	1	1	1
½	1	½	½
0	1	½	0

$\neg$	
1	0
½	½
0	1

This brings about a wealth of questions of paramount importance for logical theorising:

- Is *ST* truly the same logic as classical logic or are they different logics? And, if the latter, in what may their difference consist of?
- Is transitivity, as encapsulated by Cut, an essential property of a logic or is it something that we can dispense with?
- And, for that matter, just what (kind of) properties are Cut and similar, sequent-to-sequent, structures?

One thing that seems plain in light of the above discussion is that, if in deciding what logic we are dealing with we keep track only of provable sequents (over the usual language of classical logic), then there is no way to spot the difference between *ST* and classical logic. Is there any (good) reason to so identify logics?

Indeed there is. Sequents are usually construed as *inferences* or claims that the formula(e) on the right-hand side of the symbol ":" *follow from* the

19 Or rather LK^- together with the inverses of the operational rules, see Dicher and Paoli (2021).

20 This interpretation of *LK* goes back to Girard (1976). Note also that, usually, the consequence relation of *ST* is taken to be multiple-conclusion: a set of conclusions follows from a set of premises whenever all the premises are 1 and at least one of the conclusions has a value in {1,½}.

formula(e) on the left-hand side of that same symbol. Thus *ST* and classical logic have the same logically valid inferences.

But is this enough when it comes to unequivocally determining the identity of the logic expressed by a formal proof system?[21] The case of *ST* seems to suggest otherwise. One place where the difference between classical logic and *ST* comes to the fore is in the sequent-to-sequent rules they validate. *ST* loses Cut and many other classically valid *sequent-to-sequent inferences* or *metainferences* as they have become known in the literature (Barrio, Rosenblatt and Tajer 2015; Barrio, Pailos and Szmuc 2021). Indeed, it has been proved (Barrio, Rosenblatt and Tajer 2015; Dicher and Paoli 2019) that while the valid sequents of *ST* determine classical logic, its valid metainferences determine the logic of paradox, *LP* (cf. Priest 1979).

The *ST*-theorists are well aware and unperturbed by this fact. For them, these metainferences, or rather the rules they generate, are mere "closure principles" which a consequence relation may or may not obey (cf. Cobreros et al. 2013). Alas, whether or not this is the correct way to look at Cut and other metainferences is a disputed matter. It certainly isn't the only one. For instance, Dicher and Paoli (2021) have argued that a logic is actually an equivalence class determined in a suitable way by those metainferences that are valid in the following sense: any valuation that satisfies the premise sequents also satisfies the conclusion sequents.[22] From this perspective, *ST* is not classical logic, but rather *LP*.

So much for *ST* and its properties; now let us return to *RE*. Suppose that at the end of a careful process of formalising various natural language arguments we end up with the class of classically valid sequents as a codification of the class of valid inferences. Have we thereby also settled the matter of whether we have formalised classical or strict-tolerant logic? I believe that we have not and that we have *formed* our logic while somehow failing to form an accurate idea of which logic it is. For that, we need to answer a few more questions: What are we to make of the loss of Cut and other metainferences in *ST*? Or of the fact that *ST*, unlike classical logic, appears to be somehow ambiguous between two different consequence relations, the classical one and that of

21 This question can be asked with respect to similar, if simpler situations, see e.g. Hjortland (2013), where it is shown how one proof-system can express two different logics. See also Dicher (2020).

22 This is "local" metainferential validity. In contrast, one speaks of global metainferential validity when the universal quantifier is wide scope: for any valuation, if it satisfies the premises, then it satisfies the conclusion.

LP? These are central, albeit very abstract, problems in logical theorising and certainly salient issues in the *formation* of logics.

Is there any hope that *RE* can meaningfully guide us when we set about settling them? At first blush, one may expect that it ought to: after all, the debate is ultimately a debate over the role and status of Cut. The scenario, boiling down to deciding whether a particular (and rather special) metainference rule is valid seems to fit quite well in the Goodmanian framework. But this deceptively simple question quickly spirals out of control, becoming an arcane matter about obscure properties of logical systems and even about how these systems codify consequence relations. It is not just a case of revising, say, our concept of consequence such as to allow non-transitive relations to count as such.

The sort of questions raised by *ST* and its designation as "classical" cannot be answered by following the imperative of reaching an equilibrium between (intuitively acceptable) inferences one is not willing to give up and one's views about which rules of inference ought to be accepted. Even the framing of the problem exceeds the resources available within the *RE* model.

As with problematisation, so with problem-solving.[23] Reaching a *RE* underdetermines the issues at hand. To see this, assume for the sake of the argument that the problem can be meaningfully framed as a typical Goodmanian problem (and also bracket the many details at play in the debate around *ST*).

What is apparent is that something has to go, either the principle of inference codified by Cut or the vocabulary that makes it possible to express Liars, together with its associated inferential resources.[24] Whatever "firm" anchor point the pre-formal practice might provide us, such as, for instance, the almost universal acceptance of transitivity as a property of consequence relations, rather quickly loses its appeal. This inference principle generates inferences we are unwilling to accept, *if* we let it interact with other, equally intuitive, principles such as the *T*-rules. Plainly, *RE* cannot tell us which way to proceed and what to sacrifice—at least because all the inference principles at play have a good pre-theoretical hold on us.

---

23 This is where the "too many degrees of freedom" problem, already hinted at above creeps upon us.

24 Indeed, other options are possible, but I stick to the limits of the scenario above. Notice also that it is not just liars that are problematic. Vagueness, for instance, can lead to the same problems and be treated in like manner.

doi: 10.48106/dial.v74.i2

This is not incompatible with it being possible to defend one or another solution. But those solutions and their defences must, of necessity, rely on something more than doing justice to the pre-formal intuitions. Moreover, their virtue simply cannot be that they have balanced our pre-theoretical commitments with our pre-theoretical practice, for this virtue could be boasted by many rival solutions.

5 Case Study no. 3: Paraconsistent Christology and *FDE*

Very recently, JC Beall (2019) took to investigating the so-called *fundamental problem of christology* (cf. Pawl 2016) in light of his favourite logic, *FDE* or *first-degree entailment*. Briefly, the problem is that Patristic theology consecrates the dual nature, divine and human, of Christ. Being divine, Christ is immutable; being human, he is mutable. As a god, Christ is omnipotent; as a human, his powers are limited, etc. Christ, in other words, is possessed of inconsistent attributes. Of him, it is true both that "Christ is *P*" and that "Christ is not *P*," for a good number of essential predicates *P*. Because contradictions are bad in that they do not further the objective of achieving rational knowledge of the object that "embodies" them, this is a problem for christology.

Beall argues that the best solution to this problem is also the simplest: bite the bullet and accept that Christ is a contradictory object. That, however, is not really a bad thing. In particular, he argues, it does not entail that rational theological inquiry about Christ is impossible. Contradictions may be true of Christ, but they are not as *bad* as traditional (Aristotelian, classical, etc.) logicians took them to be. They can be handled by appropriate logics. Thus Beall argues that the proper logic for analytic Christology is the paraconsistent *FDE* (Anderson and Belnap 1975; Belnap 1977).

In its most common guise, *FDE* is a four-valued, truth-functional, and structural logic that recognises, as Beall puts it, a space of logical possibilities that allows a statement to be *true* (= 1), *false* (= 0), *both true and false* (= *b*, a "glut"), and *neither true nor false* (= *n*, a "gap"). The following matrices show how these mappings can be extended to valuations:

$\wedge$	1	b	n	0
1	1	b	n	0
b	b	b	0	0
n	n	0	n	0
0	0	0	0	0

$\vee$	1	b	n	0
1	1	1	1	1
b	1	b	1	b
n	1	1	n	n
0	1	b	n	0

$\neg$	
1	0
b	b
n	n
0	1

Both 1 and *b* are designated values and a conclusion *A* follows from some premises *X* if and only if, whenever the premises are at least true, the conclusion too is at least true.[25]

Theological and para-theological considerations aside, I agree with Beall, at least in the following sense: One's best hope of achieving a state of *RE* between the orthodox patristic determinations of Christ and one's logical beliefs is to endorse a paraconsistent logic. *Ceteris paribus*, *FDE* will do just marvellously.

But now suppose that one would wish to reject *FDE* on account of being too weak: it does not recognise as valid a great deal many inferences that we have a "natural" propensity to accept.[26] By the lights of *RE*-theorists, this should count against it. But could such criticism be levelled against *FDE* on the basis of *RE* considerations? Alas, it is difficult to see how this could be done. The *FDE* theorist has a very quick way out of this difficulty. All she needs point out is that the incriminated inference is not *logically* valid (after all, it is not *FDE*-valid), although it may be valid within some restricted domain of inquiry, maybe because the predicates of that domain have some special properties. By *FDE* lights, those inferences need not be rejected *simpliciter* though they are rejectable as a matter of logic. While indeed *FDE* is very weak, it can peacefully co-exist with various strictly speaking non-logical strengthenings of it.

So far, this has nothing to do with Christology, paraconsistent or otherwise. But suppose that a *FDE* theorist's *main* reasons to uphold this logic have to do with it cohering with her theological beliefs, in particular with her belief that Christ is an inconsistent object.[27] One trying to dislodge *FDE* as an (all-purpose) logic would be in quite a pickle. It seems clear that one could not move the *FDE* theorist to change her view. Indeed, why would she do so? Not only would this require that she give up a state of *RE*, but it would require her to do so despite having a very handy way of retaining it, i.e. denying the logicality of the *FDE*-invalid inferences while admitting that they are domain-limited valid (or perhaps analytical, etc.). At the limit, such a logician may even claim that *FDE* is too weak for *every* other domain but Christology.

25 *Mutatis mutandis*, the same definition applies to multiple-conclusion formalisations of *FDE*. For sequent calculi for *FDE*, see Beall (2013), Shapiro, L. (2017).

26 This task fits well with the main burden that the proponents of sub-classical logics have had to grapple historically: that of giving up as little as possible of the power of classical logic.

27 "Main" as used here is simply meant to signal the importance that our paraconsistent logician ascribes to coherence between their logical theological beliefs.

doi: 10.48106/dial.v74.i2

This is by no means an irrational claim, despite the seeming exoticism of the preoccupation with the divine nature in this age.[28] And it would certainly help her continue being in the state towards which our theorising must strive, that of *RE*.

There is nothing wrong with this in either the present or in any particular case whatsoever. The problem is that this is a pervasive trend: Setting a state of *RE* as the ultimate justification for our logical beliefs will tend to render weak logics immune to criticism. Quite simply, it seems very unlikely that an *FDE*-opponent of the kind described will ever be in as good a state of (reflective) equilibrium as an *FDE*-champion. The *FDE* theorist can be in equilibrium with respect to their mathematical, logical, theological and in particular Chalcedonian, and whatnot beliefs. And, presumably, a trivialist who believes that there are *no* logically valid arguments, can do even better.

This is a pathological condition to the extent that it means that weaker logics will systematically have a better chance of being justified by *RE*, simply because *RE* is easier to obtain for such a logic. Worse, given the role and purpose of *RE*, there is little incentive to aim for stronger logics.

One may reply that this is not so: A weaker logic means sacrificing—as far as logic is concerned—some inferences which we are generally willing to accept. But both the practice and other logical considerations may press exactly for their acceptance *qua* logically valid. That is true. But to the extent that these considerations are forced upon us by the practice, then, as we have already seen, they are easily brushed aside. The tendency to accept a given inference says nothing as to whether the inference is logically valid, restrictedly logically valid, analytically valid and so on. It is something that needs to be integrated and explained within a bigger theoretical picture. (So we reach again to our old conclusion that (seemingly) logical facts are fragile.) If, on the other hand, the aforementioned considerations are of a theoretical nature, then the justification process itself does not appear to be one whose stake is the successful or coherent integration of pre-theoretical beliefs with theoretical ones. Rather, it appears to be a game of making the best case for one's theoretical conviction. There can be no doubt that doing justice to the "facts" will be part of this process; it is just implausible that it will be the dominant part.

28 By contrast, a logician that would aspire towards coherence between her logical beliefs and the reasoning mistakes she most commonly commits would presumably be acting irrationally.

6 Epilogue

These, then, are the main problems with *RE* as a guide to logical theorising: First, theoretical considerations appear to always be able to undercut whatever tendencies may exist in the pre-formal practice. This means that understood as a methodology, *RE* is too weak because one of the "reflecting" surfaces itself is too weak. Second, I have argued that this methodology underdetermines both the identification of the specific problems one may encounter in "the formation of logics," i.e. problematisation, and the problem-solving process itself. Finally, *RE* systematically favours weaker logics. The weaker a logic is, the easier it will be to bring its prescriptions into harmony with other beliefs we may hold.

Part of the drama of reflective equilibrium is that it appears to fit parts of the (empirical) process of theorification, in particular, formalisation. There is little reason to doubt that the process of theorification starts by working on some raw materials—real inferences, made by real people in the real world. It also seems to me that it is correct to say that the processing of these data is both kept in check by the data and informs them in its turn. This much is inescapable insofar as we take logic to be an applied theory, i.e. our theory of *correct* reasoning (Priest 2006, ch.8).

That, however, does not make *RE* a plausible methodological constraint on, and even less so an appropriate account of the justification of, theorification—not when the chips are down. So, while the Goodmanian image with which we have started is tempting enough, turning it into a successful recipe for logical theorising turns out to be a hopeless job.[29]

At the fringe, reflective equilibrium becomes what the Senate and the consulate were in imperial Rome. One pays lip service to them. One uses them for ritual purposes. Every now and then one looks to them for (very) rough guidance to avoid too extravagant errors. And that's about it. The real power lies with the pretorians: the highly disciplined, highly skilled, and utterly unscrupulous theoretical considerations.

29 I am not alone in reaching this conclusion. See e.g. the previously quoted paper by Woods (2019) and also Wright (1986), Shapiro, S. (2000). For recent critical discussions of *RE* in non-logical contexts, see McPherson (2015), Kelly and McGrath (2010). An impressive array of objections to *RE* is surveyed and critically discussed in Cath (2016).

doi: 10.48106/dial.v74.i2

7 Postscript

Despite having reached the end of the story, the paper must go, because an anonymous referee asked the most important question to which I did *not* wish to answer here: "What are the viable alternatives?".

I stand by my decision not to answer this question here, because I cannot do it justice within the space of this paper. Still, a few words, gesturing towards my favoured answer, may be useful.

Let this be my starting point: I have framed reflective equilibrium as a method embodying a fallibilist epistemology of logic. My criticism of *RE* did not concern the suggestion that logical inquiry is fallible, that we can be wrong in our identification of the "laws of logic," etc. Nor did I challenge the claim that (parts) of the processes of logical theorisation and theorification can be described as proceeding according to a successive series of revisions of the "theory" in light of the "data" and conversely. What I have challenged is the claim that this can be turned into a substantive methodological requirement that would ensue in a justified logical theory.[30] To that extent, I do not wish to endorse fully an apriorist epistemology of logic.

These are the standard (or at least traditional) options in the epistemology of logic. I incline towards a different viewpoint. Thus the answer to the question "What is the best methodology for logical inquiry?" requires a preliminary answer to a deeper question, about how we should think about logic. As for the answer to this last question, Allo (2017, 546) puts it best:

> [I]t makes sense to think of logic as a kind of cognitive technology: a tool or set of tools used to reason more efficiently. The proposal to see logic as conceptual technology extends the scope of this picture, and emphasises that all the core notions that logical systems give a formal account of (like validity, consistency, possibility, and perhaps even meaning) should be understood as artefacts

30 It seems to me that this is not completely false even of a priori methodologies for logic. It is one thing to argue, however (im)plausibly, that the validity of *modus ponens* is known *a priori* by dint of knowing the meaning of *if … then*. (The disjunction between plausible and implausible, suggested by e.g. McGee's (1985) alleged counterexample to *modus ponens* should by itself give us pause.) It is a rather different thing to argue that the same is true of, e.g. vacuous discharges of assumptions, which are essential for ensuring a monotonic behaviour of the conditional. Likewise, it is one thing to argue that transitivity is an analytic note of the concept of "logical consequence" and quite another to decide whether this is to be captured at the inferential or metainferential level.

that shape deductive reasoning practices rather than as neutral descriptions or codifications of pre-existing inferential practices.

So the referee's question "What are the viable alternatives?" has a simple but hardly informative answer: Whatever methodology best serves the imperative of developing the best cognitive technology that logic can be. What that actually means is a matter for further thinking.*

Bogdan Dicher
0000-0002-2587-0649
University of Lisbon
bdicher@letras.ulisboa.pt

References

Allo, Patrick. 2017. "A Constructionist Philosophy of Logic." *Minds and Machines* 27(3): 545–564, doi:10.1007/s11023-017-9430-9.

Anderson, Alan Ross and Belnap, Nuel D., Jr. 1975. *Entailment: The Logic of Relevance and Necessity. Volume 1*. Princeton, New Jersey: Princeton University Press.

Barrio, Eduardo Alejandro, Pailos, Federico Matías and Szmuc, Damian. 2021. "Substructural Logics, Pluralism and Collapse." *Synthese* 198(suppl. 20): 4991–5007, doi:10.1007/s11229-018-01963-3.

Barrio, Eduardo Alejandro, Rosenblatt, Lucas and Tajer, Diego. 2015. "The Logics of Strict-Tolerant Logic." *Journal of Philosophical Logic* 44(5): 551–571, doi:10.1007/s10992-014-9342-6.

Beall, J. C. 2013. "LP+, K3+, FDE+, and Their 'Classical Collapse'." *The Review of Symbolic Logic* 6(4): 742–754, doi:10.1017/s1755020313000142.

—. 2019. "Christ – A Contradiction: A Defense of Contradictory Christology." *The Journal of Analytic Theology* 7: 400–433, doi:10.12978/jat.2019-7.090202010411.

Belnap, Nuel D., Jr. 1977. "A Useful Four-Valued Logic." in *Modern Uses of Multiple-Valued Logic: Invited Papers From the Fifth International Symposium on Multiple-Valued Logic, held at Indiana University, Bloomington, Indiana, May 13–16, 1975*, edited by Michael J. Dunn and George Epstein, pp. 8–40. Episteme n. 2. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-010-1161-7_2.

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

* For comments, corrections and discussion, thanks to Amanda Bryant, Bruno Jacinto, Francesco Paoli, Greg Restall, Dave Ripley, Diogo Santos and Ricardo Santos, as well as to two anonymous referees for *Dialectica*. This work was supported by the *Fundaçaõ para a Ciência e a Tecnologia (FCT)*, Portugal, through the grant SFRH/BPD/116125/2016.

doi: 10.48106/dial.v74.i2

—. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Massachusetts: Harvard University Press, doi:10.2307/j.ctvjghvz0.

Brun, Georg. 2003. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. Logos n. 2. Heusenstamm b. Frankfurt: Ontos Verlag. Second edition: Brun (2004), doi:10.1515/9783110323528.

—. 2004. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. 2nd ed. Heusenstamm b. Frankfurt: Ontos Verlag. First edition: Brun (2003).

—. 2013. "Rival Logics, Disagreement and Reflective Equilibrium." in *Epistemology: Contexts, Values, Disagreement. Proceedings of the 34th International Ludwig Wittgenstein Symposium in Kirchberg, 2011*, edited by Christoph Jäger and Winfried Löffler, pp. 355–369. Publications of the Austrian Ludwig Wittgenstein Society (new series) n. 19. Berlin: de Gruyter, doi:10.1515/9783110329018.355.

—. 2014. "Reconstructing Arguments: Formalization and Reflective Equilibrium." in *Theory and Practice of Logical Reconstruction: Anselm as a Model Case*, edited by Friedrich Reinmuth, Geo Siegwart, and Christian Tapp, pp. 94–129. Logical Analysis and History of Philosophy n. 17. Münster: Mentis Verlag, doi:10.30965/9783957439468_008.

Carnap, Rudolf. 1934. *Logische Syntax der Sprache*. Schriften zur wissenschaftlichen Weltauffassung n. 8. Wien: Julius Springer.

—. 1937. *The Logical Syntax of Language*. International Library of Psychology, Philosophy and Scientific Method. London: Kegan Paul, Trench, Trübner & Co. Translation of Carnap (1934) by Amethe Smeaton, Countess von Zeppelin.

Cath, Yuri. 2016. "Reflective Equilibrium." in *The Oxford Handbook of Philosophical Methodology*, edited by Herman Cappelen, Tamar Szabó Gendler, and John Hawthorne, pp. 213–230. Oxford Handbooks. Oxford: Oxford University Press, doi:10.1093/oxfordhb/9780199668779.013.32.

Cobreros, Pablo, Egré, Paul, Ripley, David and van Rooij, Robert. 2012. "Tolerant, Classical, Strict." *Journal of Philosophical Logic* 41(2): 347–385, doi:10.1007/s10992-010-9165-z.

—. 2013. "Reaching Transparent Truth." *Mind* 122(488): 841–866, doi:10.1093/mind/fzt110.

Dicher, Bogdan. 2016. "Weak Disharmony: Some Lessons for Proof-Theoretic Semantics." *The Review of Symbolic Logic* 9(3): 583–602, doi:10.1017/s1755020316000162.

—. 2020. "Hopeful Monsters: A Note on Multiple Conclusions." *Erkenntnis* 85(1): 77–98, doi:10.1007/s10670-018-0019-3.

Dicher, Bogdan and Paoli, Francesco. 2019. "ST, LP and Tolerant Metainferences." in *Graham Priest on Dialetheism and Paraconsistency*, edited by Can Başkent and Thomas Macaulay Ferguson, pp. 383–408. Cham: Springer Nature, doi:10.1007/978-3-030-25365-3_18.

—. 2021. "The Original Sin of Proof-Theoretic Semantics." *Synthese* 198(1): 615–640, doi:10.1007/s11229-018-02048-x.

Dummett, Michael A. E. 1991. *The Logical Basis of Metaphysics*. London: Gerald Duckworth & Co.

Eklund, Matti. 1996. "How Logic Became First-Order." *Nordic Journal of Philosophy* 1(2): 147–167, https://www.hf.uio.no/ifikk/english/research/publications/journals/njpl/files/vol1no2/howlogic.pdf.

Francez, Nissim. 2015. *Proof-Theoretic Semantics*. Studies in Logic n. 57. London: College Publications.

Gentzen, Gerhard. 1935. "Untersuchungen über das logische Schließen." *Mathematische Zeitschrift* 39: 176–210, 405–431. Republished as Gentzen (1969), doi:10.1007/BF01201353.

—. 1969. *Untersuchungen über das logische Schliessen*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Girard, Jean-Yves. 1976. *Three-Valued Logic and Cut-Elimination: The Actual Meaning of Takeuti's Conjecture*. Warszawa: PańStwow Wydawnictwo Naukowe (PWN).

Goodman, Nelson. 1955. *Fact, Fiction and Forecast*. Cambridge, Massachusetts: Harvard University Press.

Gratzl, Norbert and Orlandelli, Eugenio. 2017. "Double-Line Harmony in a Sequent Setting." in *The Logica Yearbook 2016*, edited by Pavel Arazim and Tomáš Lávička, pp. 157–171. London: College Publications.

Hjortland, Ole Thomassen. 2013. "Logical Pluralism, Meaning-Variance, and Verbal Disputes." *Australasian Journal of Philosophy* 91(2): 355–373, doi:10.1080/00048402.2011.648945.

—. 2017. "Anti-Exceptionalism About Logic." *Philosophical Studies* 174(3): 631–658, doi:10.1007/s11098-016-0701-8.

Kelly, Thomas and McGrath, Sarah. 2010. "Is Reflective Equilibrium Enough?" in *Philosophical Perspectives 24: Epistemology*, edited by John Hawthorne, pp. 325–359. Hoboken, New Jersey: John Wiley and Sons, Inc., doi:10.1111/j.1520-8583.2010.00195.x.

Kleene, Stephen Cole. 1952. *Introduction to Metamathematics*. New York: Van Nostrand Reinhold.

McGee, Vann. 1985. "A Counterexample to Modus Ponens." *The Journal of Philosophy* 82(9): 462–471, doi:10.2307/2026276.

McPherson, Tristram. 2015. "The Methodological Irrelevance of Reflective Equilibrium." in *The Palgrave Handbook of Philosophical Methods*, edited by Christopher John Daly, pp. 652–674. London: Palgrave Macmillan, doi:10.1057/9781137344557_27.

Pawl, Timothy. 2016. *In Defense of Conciliar Christology: A Philosophical Essay*. Oxford Studies in Analytic Theology. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198765929.001.0001.

Peregrin, Jaroslav and Svoboda, Vladimír. 2013. "Criteria for Logical Formalization." *Synthese* 190(14): 2897–2924, doi:10.1007/s11229-012-0104-0.

—. 2016. “Logical Formalization and the Formation of Logic(s).” *Logique et Analyse* 59(233): 55–80, doi:10.2143/LEA.233.0.3149531.

—. 2017. *Reflective Equilibrium and the Principles of Logical Analysis: Understanding the Laws of Logic*. London: Routledge, doi:10.4324/9781315453934.

Prawitz, Dag. 1965. *Natural Deduction: A Proof Theoretical Study*. Stockholm Studies in Philosophy n. 3. Stockholm: Almqvist & Wiksell.

—. 1974. “On the Idea of a General Proof Theory.” *Synthese* 27(1/2): 63–77, doi:10.1007/bf00660889.

Priest, Graham. 1979. “The Logic of Paradox.” *Journal of Philosophical Logic* 8(2): 219–241, doi:10.1007/bf00258428.

—. 2006. *Doubt Truth to Be a Liar*. Oxford: Oxford University Press, doi:10.1093/0199263280.001.0001.

—. 2014. “Revising Logic.” in *The Metaphysics of Logic*, edited by Penelope Rush, pp. 211–223. Cambridge: Cambridge University Press, doi:10.1017/cbo9781139626279.016.

Quine, Willard van Orman. 1970. *Philosophy of Logic*. Foundations of Philosophy Series. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Second edition: Quine (1986).

—. 1986. *Philosophy of Logic*. 2nd ed. Cambridge, Massachusetts: Harvard University Press. First edition: Quine (1970).

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press. Revised edition: Rawls (1999), doi:10.2307/j.ctvjf9z6v.

—. 1999. *A Theory of Justice*. Cambridge, Massachusetts: Harvard University Press, doi:10.2307/j.ctvkjb25m.

Read, Stephen. 2000. “Harmony and Autonomy in Classical Logic.” *Journal of Philosophical Logic* 29(2): 123–154, doi:10.1023/a:1004787622057.

Ripley, David. 2012. “Conservatively Extending Classical Logic with Transparent Truth.” *The Review of Symbolic Logic* 5(2): 354–378, doi:10.1017/S1755020312000056.

—. 2013. “Paradoxes and Failures of Cut.” *Australasian Journal of Philosophy* 91(1): 139–164, doi:10.1080/00048402.2011.630010.

Russell, Gillian K. 2014. “Metaphysical Analyticity and the Epistemology of Logic.” *Philosophical Studies* 171(1): 161–175, doi:10.1007/s11098-013-0255-y.

Schröder-Heister, Peter. 2018. “Proof-Theoretic Semantics.” in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2018/entries/proof-theoretic-semantics/.

Shapiro, Lionel. 2017. “LP, K3, and FDE as Substructural Logics.” in *The Logica Yearbook 2016*, edited by Pavel Arazim and Tomáš Lávička, pp. 257–272. London: College Publications.

SHAPIRO, Stewart. 1991. *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford Logic Guides n. 17. Oxford: Oxford University Press, doi:10.1093/0198250290.001.0001.

—. 2000. "The Status of Logic." in *New Essays on the A Priori*, edited by Paul Artin BOGHOSSIAN and Christopher PEACOCKE, pp. 333–366. Oxford: Oxford University Press, doi:10.1093/0199241279.003.0014.

STEINBERGER, Florian. 2011. "Why Conclusions should Remain Single." *Journal of Philosophical Logic* 40(3): 333–355, doi:10.1007/s10992-010-9153-3.

WILLIAMSON, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell Publishers. Second edition: Williamson (2022), doi:10.1002/9780470696675.

—. 2022. *The Philosophy of Philosophy*. 2nd ed. The Blackwell / Brown Lectures in Philosophy. Hoboken, New Jersey: Wiley-Blackwell. First edition: Williamson (2007), doi:10.1002/9781119616702.

WOODS, Jack. 2019. "Against Reflective Equilibrium in Logical Theorizing." *Australasian Journal of Logic* 16(7): 319–341, doi:10.26686/ajl.v16i7.5927.

WRIGHT, Crispin. 1986. "Inventing Logical Necessity." in *Language, Mind and Logic*, edited by Jeremy BUTTERFIELD, pp. 187–209. Cambridge: Cambridge University Press.

doi: 10.48106/dial.v74.i2

The Primacy of the Universal Quantifier in Frege's Concept-Script

Joongol Kim

This paper presents three explanations of why Frege took the universal, rather than the existential, quantifier as primitive in his formalization of logic. The first two explanations provide technical reasons related to how Frege formalizes the logic of truth-functions and the logic of quantification. The third, philosophical explanation locates the reason in Frege's logicist goal of analyzing arithmetical concepts—especially the concepts of 0 and 1—in purely logical terms.

It is a well-known fact of elementary logic that each of the universal and existential quantifier symbols, ∀ and ∃, can be defined in terms of the other, as follows:

(1) $\exists_\alpha \phi =_{\mathrm{df}} \neg\forall_\alpha\neg\phi$
(2) $\forall_\alpha \phi =_{\mathrm{df}} \neg\exists_\alpha\neg\phi$.

So one could adopt ∃ as a primitive symbol and then define ∀ in terms of it. Frege, the inventor of modern quantificational logic, did the reverse, taking the universal quantifier as primitive in his formalization of logic called the *concept-script*.[1] Thus, in his early monograph on the concept-script, *Begriffsschrift*, Frege (1967, sec. 11) introduces his universal quantifier symbol—the concavity, $\smile$ —and expresses "$\forall_\alpha\phi$" as follows:

$$\vdash\!\overset{\mathfrak{a}}{\smile}\, \Phi(\mathfrak{a})$$

Then, using the negation stroke, $\top$, Frege (1967, sec. 12) constructs the complex formula

$$\vdash\!\top\!\overset{\mathfrak{a}}{\smile}\!\top\, \Lambda(\mathfrak{a})$$

1 For a quick introduction to Frege's concept-script, see Cook (2013).

and reads it as "There are Λ," although, taken literally, it says that not all things are non-Λ. Frege's concept-script includes no special existential quantifier symbol such as the downside-up form of the concavity—the *convexity* (see Kneale and Kneale 1962, 516–517)—as an abbreviation of ┬◡┬ .

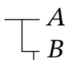
An interesting question is why Frege employed the universal, rather than the existential, quantifier as a primitive sign in his formal language. Nowhere in his writings does he address this question. Indeed, as Macbeth (2005, 4) noted, "it seems never even to occur to him that he could treat the existential quantifier as the primitive sign for generality and then define the universal quantifier in terms of it." The main purpose of this paper is to address this gap in our understanding of Frege's logical formalism by giving three possible explanations of Frege's adoption of the universal quantifier as a primitive.

The first two explanations—to be discussed in turn in sections 1, 2—offer technical reasons: given how the logic of truth-functions and the logic of quantification are formalized in the concept-script, it was natural and convenient to take the universal quantifier as primitive. The third explanation—to be given in section 3—is that Frege was forced to adopt the universal quantifier as a primitive in his pursuit of providing definitions of the numbers 0 and 1 in purely logical terms. In a well-meaning attempt to cast Frege's legacy in the most favorable light, Dummett (1981, xiii–xxv) touted his achievements in logic and its philosophical underpinnings, and downplayed his failed logicist philosophy of mathematics. Dummett (1981, xv) allowed that "Logic was, indeed, for Frege principally a tool for and a prolegomenon to the study of the philosophy of mathematics." However, if the third explanation which locates the reason for Frege's choice of the primitive quantifier symbol in his logicist account of numbers could be substantiated along the lines suggested below, that would indicate that the concept-script was not for him a mere neutral tool for studying the philosophy of mathematics but was even designed so as to serve the purposes of his logicist philosophy of arithmetic.

1 Conditionality in the Concept-Script

From a technical point of view, one notable feature of Frege's concept-script is that it has a notational device for just one binary truth-function—conditionality—and expresses the others in terms of it (with the help of the negation stroke) without introducing notational abbreviations for them. As a symbol for conditionality, Frege adopts a vertical stroke that connects two horizontal strokes; the upper and the lower horizontal stroke are respectively followed by the consequent and the antecedent of the conditional. Thus, the conditional

corresponds in modern notation to $A \rightarrow B$. Then, using the conditional stroke, Frege (1967, sec. 7) expresses conjunction ("$A \wedge B$") and disjunction ("$A \vee B$") respectively as

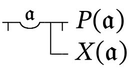
 and 

Frege (1967, sec. 7) considered the idea of introducing a sign for conjunction as a primitive and defining conditionality in terms of negation and conjunction; however, he "chose the other way because [he] felt that it enables us to express inferences more simply." He says "more simply," because by taking conditionality as a basic truth-function he was able to represent any inference with more than one premise by a single rule of inference, namely modus ponens (1967, sec. 6) (more on this shortly).

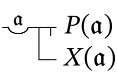
In "Boole's Logical Calculus and the Concept-script," Frege (1979) provides another reason for his choice of conditionality over conjunction as a primitive. He argues that since "it is a basic principle of science to reduce the number of axioms to the fewest possible," and since "[t]he more primitive signs you introduce, the more axioms you need," only the fewest possible primitive symbols should be introduced (1979, 36). For this purpose, "I must choose those with the simplest possible meanings," where a meaning is said to be simpler "the less it says" (1979, 36). Then he observes that the conditional stroke, which excludes only one possibility of assigning truth-values to the component sentences—the case of the antecedent being true and the consequent being false—says less than Boole's identity sign meaning "if and only if" and even less than Boole's multiplication sign meaning "and."

Now, as Frege (1979, 37) points out, there are four possible binary truth-functions each of which excludes only one truth-value assignment. One of them is disjunction expressed by the inclusive "or." Why choose conditionality over disjunction? Frege's (1979, 37) answer is: "because of the ease with which it can be used in inference, and because its content has a close affinity with the important relation of ground and consequent." The affinity between the content of conditionality and the "relation of ground and consequent" is evidenced by the fact that any consequence relationship between statements—such as that "B" is a consequence of "A or B" and "not A"—can be expressed as a conditional: if A or B, then if not A, then B. This is why "an inference in accordance with any mode of inference can be reduced to [modus ponens]" (Frege 1967, sec. 6). And "[s]ince it is therefore possible to manage with a single mode of inference, it is a commandment of perspicuity to do so" (Frege 1967, sec. 6).

The fact that Frege chose conditionality as a primitive truth-function along with negation in the concept-script provides an explanation of why he took the universal, rather than the existential, quantifier as primitive: if the conditional sign is to be the main logical operator of a truth-functional formula, then a quantified formula with a truth-functional subformula could best be symbolized in terms of a universal quantifier. For instance, consider an I-statement of the form "Some X are P." It is standardly symbolized as "$\exists x(Xx \wedge Px)$"; but if the conjunctive subformula has to be rendered in the form of a conditional, then the whole I-statement could best be analyzed as "Not everything is such that if it is X, then it is not P," and so would be expressed in the concept-script as

(3) 

Of course, it is not impossible to symbolize the I-statement in terms of an existential quantifier while keeping the conditional sign as the only binary sentential operator in its truth-functional subformula. The following will do: "$\exists x \neg(Xx \rightarrow \neg Px)$". However, (3) has an important advantage over that alternative: as is made clear by Frege's (1967, 28) diagram of "the square of the logical opposition," (3) makes explicit the contradictory relationship between the I-statement and the E-statement of the form "No X are P." The symbolization of the E-statement in the concept-script, namely

(4) 

directly contradicts (3). To be sure, this contradictory relationship between the I- and the E-statement could also be made explicit using an existential quantifier by formalizing the E-statement as "$\neg\exists x \neg(Xx \rightarrow \neg Px)$." However, this formula cries out for reanalysis as "$\forall x(Xx \rightarrow \neg Px)$," that is, (4), for the sake of simplicity and naturalness.

The upshot is that if the conditional sign is employed as the only binary truth-functional operator, then the universal quantifier is better suited than the existential quantifier to capture the logical structures of, and relationships between, quantified formulas. So Frege had a good reason to adopt the concavity as a primitive quantifier symbol in his conditionality-based concept-script.

2 Generality in the Concept-Script

Frege's 1879 monograph, *Begriffschrift*, is subtitled "*a formula language, modeled upon that of arithmetic, for pure thought*." Arithmetic, in its narrow sense, is the theory of natural numbers, but here Frege uses the term in the sense of the theory of numbers in general. In this broad sense arithmetic includes (mathematical) analysis—or better, Analysis, with a capital "A," for distinction.[2] Analysis—the theory of functions of a real variable—involves the notions of function and variable. When Frege (1967, 6) wrote that the fact that the concept-script is modeled upon the language of arithmetic "has to do with fundamental ideas rather than with details of execution," he meant that functions and variables form the core of the design of his symbolic language of logic.

To explain in more detail, first, the concept-script replaces the traditional subject-predicate analysis of a proposition with the function-argument analysis (Frege 1967, sec. 9–10). Secondly—and this is "[t]he most immediate point of contact between [his] formula language and that of arithmetic"—it adopts "the way in which letters are employed" in arithmetic (Frege 1967, 6). What Frege means by "letters" here is what mathematicians—wrongly, in Frege's (1984d, 285–288) view—refer to as variables. Arithmetic is marked partly by its use of Roman letters such as x in the formula

(5) $x^2 - 4x = x(x - 4)$.

2 In the titles of Frege's two books, *Foundations of Arithmetic* and *Basic Laws of Arithmetic*, "arithmetic" has this broad sense. This can be seen from Frege's remarks in *Grundlagen* §1 that "[i]n arithmetic, [...] it has been the tradition to reason less strictly than in geometry" and that "[t]he discovery of higher analysis"—namely, Leibniz's invention of the practical but less than rigorous method of infinitesimal calculus—"only served to confirm this tendency." Also, when he talks about "the great tree of the science of number as we know it, towering, spreading, and still continually growing" (1980b, sec. 16), he refers to arithmetic in its broad sense, including the theory of complex numbers. *Grundgesetze* contains the beginnings of an investigation of the theory of real numbers, and there is reason to think that its planned third volume was to include a treatment of complex numbers (see Dummett 1981, 241–242).

doi: 10.48106/dial.v74.i2

Here x serves as a sign of generality: it indicates that the equation holds no matter what number is put for x. By incorporating in his concept-script signs of generality (as well as of functions with an arbitrary number of arguments whose value is a truth-value), Frege was able to create a symbolic language to express the full logic of quantification.

But considering that the symbolic language of arithmetic expresses generality using Roman letters alone as in (5) and does not have separate quantifier symbols, the question arises as to why Frege also introduced the concavity sign and, therewith, German letters such as $\mathfrak{a}$ in addition to Roman letters. In *Grundgesetze* he addresses the question, and says that by means of Roman letters alone it would be impossible to delimit the scope of generality for sentences such as the following (2013, sec. 8):

(6) $-\!\!\top\!\!-\ 2 + 3x = 5x.$

(6) admits of two different readings. First, the generality sign x can be viewed as having narrow scope with respect to the negation stroke. On this reading, (6) would express the negation of a generality, namely

(7) $-\!\!\top\!\!-\!\overset{\mathfrak{a}}{\smile}\!-\ 2 + 3\mathfrak{a} = 5\mathfrak{a}$

which is true. Alternatively, the letter x can be viewed as having wide scope, in which case (6) expresses a false universal, namely

(8) $-\!\overset{\mathfrak{a}}{\smile}\!-\!\!\top\!\!-\ 2 + 3\mathfrak{a} = 5\mathfrak{a}.$

Since it is crucial for the purposes of a logical formalism to be able to capture the difference between (7) and (8), it was necessary for Frege to introduce the concavity sign as a device for delimiting the scope of Roman letters which connote generality. Thus, although the ambiguity of (6) can be removed by "stipulating that the *scope* of a *Roman letter* is to include everything that occurs in the proposition apart from the judgment-stroke" (Frege 2013, sec. 17), that is, by understanding (6) always as meaning (8), the concavity sign is still needed to express the negation of a generality such as (7).

In fact, in *Begriffschrift*, Frege (1967, sec. 11) gave the same explanation of the need for the concavity sign, albeit using slightly more complicated examples. Consider the following conditional:

(9) $\begin{array}{l} \overline{}\!\!\!\!\!\!\top\!\!\!\!-\ A \\ \;\lfloor\!\overset{\mathfrak{a}}{\smile}\!-\ X(\mathfrak{a}) \end{array}$

Frege (1967, sec. 11) emphasizes that (9) "does not by any means deny that the case in which $X(\Delta)$ is affirmed and A is denied does occur" for some object Δ. His point is that (9), a conditional formula, should not be confused with the following universal formula that says that such a case never occurs:

(10) $$\overset{\mathfrak{a}}{\smile}\!\!\begin{array}{l} \top\, A \\ \llcorner\, X(\mathfrak{a}) \end{array}$$

The difference in logical content between (9) and (10) would have been lost without the concavity. So "[t]his explains why the concavity with the German letter written into it is necessary: *it delimits the scope that the generality indicated by the letter covers*" (Frege 1967, sec. 11).

These considerations suggest another technical explanation of why Frege adopted the universal quantifier as a primitive. The concept-script was modeled on the symbolic language of arithmetic, and so Roman letters were used as a device to express generality. But as a result of such use of Roman letters, scope ambiguities arose, and the concavity was introduced to deal with them. Frege's adoption of the universal quantifier as a primitive was, then, a natural consequence of modeling his concept-script upon the symbolic language of arithmetic.

In order to avoid a possible misunderstanding, it should be noted that the fact that the concavity was introduced to delimit the scope of generality does not mean that it was intended to serve as a mere scope marker—a sort of punctuation sign—in such formulas as (7) and (8). That is, it would be a mistake to think that what expresses generality in (7) and (8) is the German letter $\mathfrak{a}$ in the formula "$2 + 3\mathfrak{a} = 5\mathfrak{a}$," with the concavity left to play the role of marking the scope of the letter. Frege (1967, sec. 11) explains the formula "$\overset{\mathfrak{a}}{\smile}\!- \Phi(\mathfrak{a})$" as meaning that "whatever we may put in place of $\mathfrak{a}$, $\Phi(\mathfrak{a})$ holds," or in modern parlance, "for any value of variable $\mathfrak{a}$, Φ is true of it." This means that in the formula "$\overset{\mathfrak{a}}{\smile}\!- \Phi(\mathfrak{a})$," generality is expressed by the quantifier "$\overset{\mathfrak{a}}{\smile}$," not by the $\mathfrak{a}$ in "$\Phi(\mathfrak{a})$." This latter $\mathfrak{a}$ always refers to something particular—namely, a given value of the variable $\mathfrak{a}$. That is Frege's point when he writes that "the horizontal stroke to the right of the concavity is the content stroke of $\Phi(\mathfrak{a})$, and here we must imagine that something definite has been substituted for $\mathfrak{a}$" (1967, sec. 11). So the concavity, with the meaning of "for any value of," is indeed a sign of generality corresponding to the modern $\forall$, and not a mere scope marker.

doi: 10.48106/dial.v74.i2

A related point to note is that the concavity is the only device in the concept-script to express generality. For Frege (1967, sec. 11), a Roman letter is an "abbreviation" for the case where "the concavity immediately follows the judgment stroke," that is, "the content of the entire judgment constitutes the scope of the German letter." Thus, despite the fact that Roman letters precede the concavity in the order of discovery, Frege saw—rightly—the explanatory primacy of the latter over the former once he had realized that Roman letters are inadequate as a device for expressing generality due to scope ambiguities.

3 The Numbers 0 and 1

Another, different kind of explanation of Frege's adoption of the universal quantifier as a primitive could be found in the roles of universal and existential quantifiers in Frege's philosophy of arithmetic. After all, as Frege (1967, 8) acknowledged in the Preface to *Begriffsschrift*, "arithmetic was the point of departure for the train of thought that led [him] to [his] [concept-script]." Not only that; he intended "to apply it first of all to that science, attempting to provide a more detailed analysis of the concepts of arithmetic and a deeper foundation for its theorems" (1967, 8). Since Frege, as a logicist, aimed to establish arithmetic as part of logic, his expressions "detailed" and "deeper" here could be understood as meaning "logical." That is, the primary applications of the concept-script were to be found in providing a logical analysis of the concepts of arithmetic and a logical foundation for its theorems. The possibility suggests itself, then, that Frege's initial attempts in that direction may have convinced him that the universal, rather than the existential, quantifier should be taken as primitive. But to support this conjecture requires evidence from Frege's early writings—early enough to have made an impact on his *Begriffsschrift* of 1879—that a logical analysis of arithmetical concepts or a logical proof of arithmetical truths compelled him to invoke the universal, rather than the existential, quantifier. Is there such evidence?

At the end of the Preface to *Begriffsschrift*, Frege (1967, 8) briefly states his future plans "to elucidate the concepts of number, magnitude, and so forth," adding that "all this will be the object of further investigations, which I shall publish immediately after this booklet." The word "immediately" here suggests that at the time of writing he was already at an advanced stage of his research about number, if not about quantity. Indeed, he reports in a letter of 1882 that "I have now nearly completed a book in which I treat the concept of number and demonstrate that the first principles of computation which up to now have generally been regarded as unprovable axioms can be proved from definitions by means of logical laws alone" (1980a, 99). The book here referred to may well be the one that Frege (2013, IX) later said he had been forced to discard due to "internal changes within the concept-script," including changing the *Begriffsschrift* triple-bar sign ≡ for identity to the usual "equals" sign =. In *Begriffsschrift* Frege used "≡" as the identity sign (of a metalinguistic kind[3]): he presents the substitutivity principle (1967, sec. 20)—that if $c \equiv d$, then if $f(c)$, then $f(d)$—as one of the two basic laws concerning the triple-bar sign along with the reflexivity principle that $c \equiv c$ (1967, sec. 21). In *Grundgesetze*, Frege adopts the "equals" sign as his new identity symbol because "I have convinced myself that in arithmetic it possesses just that reference that I too want to designate" (2013, IX). That is, in *Grundgesetze*, "I use the word 'equal' with the same reference as 'coinciding with' or 'identical with' " because he has now realized that "this is also how the equality-sign is actually used in arithmetic" (2013, IX). These remarks reveal that at the time of writing *Begriffsschrift*, Frege did not think that the "equals" sign in arithmetic has the meaning of "identical with,"[4] and hence had to choose a different symbol, ≡, to denote the relation of identity. In other words, Frege, in his early period, does not seem to have regarded arithmetic as concerned with objects (as opposed to properties, relations, or functions in general), that is, those things capable of standing in the relation of identity. These considerations suggest that Frege discarded the "nearly completed" book because of his realization that numbers must be viewed as objects.

What could Frege have thought that numbers are, in his early years, if they are not objects? What could he have thought that an equality of the form "$m = n$" means if not that m is identical with n? Clues to these questions are found in *Grundlagen*. In the beginning section of Part IV, Frege (1980b, sec. 55) first reminds the reader of the main lesson of Part III that "the content of a statement of number is an assertion about a concept," and then proceeds to give definitions of individual numbers which, as he puts it, "suggest themselves so spontaneously in the light of [the results of Part III]" (1980b, sec.

3 Frege's (1967, sec. 8) solution to the puzzle of how "$a = b$," as opposed to "$a = a$," can be informative was to take "$a \equiv b$" to talk about the names, not the objects a and b. Later he replaced it with a new solution based on the distinction between sense and meaning (1984b). For details, see Kim (2011, sec. 4–5).

4 This explains why Frege (1967) uses the "equals" sign in *Begriffsschrift* only in relation to arithmetic formulas—"$(a + b)c = ac + bc$" in §1 and "$3 \times 7 = 21$" in §5—and never in non-arithmetical contexts.

doi: 10.48106/dial.v74.i2

56). These definitions introduce the numbers 0 and 1 in the context "The number *n* belongs to a concept *F*," and so present them as properties of concepts (just as to say that wisdom belongs to Socrates is to say that wisdom is a property of Socrates). This interpretation is supported by the fact that after explaining, in §56, why those definitions must be rejected as unsatisfactory despite "suggest[ing] themselves so spontaneously,"[5] Frege (1980b, sec. 57) writes that therefore "I have avoided calling a number such as 0 or 1 or 2 a *property* of a concept" (original emphasis). It is reasonable to think that this view of numbers as properties of concepts, which he presupposes in §55 as the outcome of his initial inquiry into the concept of number only to reject it in §56, was his early view of numbers (see below for more evidence); and if so, it is also reasonable to infer that in his early period he interpreted an equality of the form "$m = n$" as an equivalence of some form such as "The number *m* belongs to a concept $F \equiv$ the number *n* belongs to *F*," where the triple bar sign is used to indicate the "identity of content" between sentences (rather than names) as in the propositions (67) and (68) of *Begriffsschrift*.

Now, given Frege's statement in *Begriffsschrift* that he will "publish immediately after this booklet" the results of his investigation into the concept of number, it seems safe to assume that while *Begriffsschrift* was being composed, Frege may have been working on—or may even have finished (as will be evidenced below)—at least a detailed outline of the "nearly completed" book he referred to in his 1882 letter quoted above. Indeed, his remark quoted at the beginning of this section—that "arithmetic was the point of departure for the train of thought that led [him] to [his] [concept-script]"—suggests that his early attempts to give logical definitions of concepts of arithmetic and to derive some of its theorems from those definitions alone led him to devise the concept-script in the first place. It is plausible, then, that the definitions of individual numbers given in *Grundlagen* §55 were part of those early attempts of Frege to give a logicist account of arithmetic, and so predated the composition of *Begriffsschrift*.

And Frege seems to have found it necessary to invoke the universal, rather than the existential, quantifier in attempting to provide logical definitions of the numbers 0 and 1. He first observes that "[i]t is tempting to define 0" as follows (1980b, sec. 55):

5 For an exposition and discussion of Frege's objections to the definitions in *Grundlagen* §55, see Kim (2013). For a defense and development of a theory of number based on similar definitions, see Kim (2015) and Kim (2020).

(11) The number 0 belongs to a concept F [or, more colloquially, there are 0 Fs] $=_{\mathrm{df}}$ no object falls under the concept F [or there are no Fs].

However, he objects that (11) "seems to amount to replacing 0 by 'no,' which means the same." That is, he raises against (11) a charge of circularity that can be leveled against an attempt to define, say, "x is an ethical action" as "x is a moral action."

One might challenge this charge of circularity by maintaining that the "no" in "There are no Fs" is short for "not any," and so that the definiens of (11) should not be viewed as replacing "0" with "no" but rather as abbreviating the following:

(12) It is not the case that there exists any F [in symbols, $\neg\exists x(Fx)$].

Thus understood, (11) would seem more similar to defining "x is single" as "x is not married" than to defining "x is an ethical action" as "x is a moral action."

The problem is that an existential statement of the form "There is an F" (or, in symbols, "$\exists x(Fx)$") has the logical meaning of "There is at least one F." Frege emphasizes this fact whenever the occasion arises. In *Begriffsschrift* he observes that "If, for example, $\Lambda(x)$ means the circumstance that x is a house, then

$$\vdash\!\!\top\!\overset{\mathfrak{a}}{\smile}\!\top\, \Lambda(\mathfrak{a})$$

reads 'There are houses or there is at least one house' " (1967, sec. 12, n15). And a moment later he points out that the expression "some" in a statement of the form "Some M are P," "must always be understood here in such a way as to include the case 'one' as well" and that "[m]ore explicitly we would say 'some or at least one' " (1967, n16). In *Grundgesetze* Frege (2013, sec. 8) is even more explicit about this, noting that the sentence

$$\vdash\!\!\top\!\overset{\mathfrak{a}}{\smile}\!\top\, 2 + 3.\mathfrak{a} = 5.\mathfrak{a}$$

"says: *there is* at least one solution for the equation '$2 + 3.x = 5.x$'," and that the sentence

$$\vdash\!\!\top\!\overset{\mathfrak{a}}{\smile}\!\top\, \mathfrak{a}^2 = 1$$

has the meaning of "*there is* at least one square root of 1." In §13, he notes that "the plural ['some'] is not to be understood as requiring that there must

doi: 10.48106/dial.v74.i2

be more than one" but as meaning "there is at least one."[6] Thus, given this fact that an existential statement has the meaning of "there is at least one ...," taking the existential quantifier as primitive and defining the number 0 as in (13)

(13) The number 0 belongs to a concept $F =_{df}$ it is not the case that there is at least one F

would have exposed Frege to the charge of defining 0 in terms of the number word "one" and so of smuggling in an arithmetical concept while attempting to give logical definitions of arithmetical concepts.

It is for that reason that Frege (1980b, sec. 55) proposes instead that "[t]he following formulation is therefore preferable: the number 0 belongs to a concept, if the proposition that *a* does not fall under that concept is true universally, whatever *a* may be." The proposal is, in effect, to define the number 0 in terms of the universal quantifier as follows:

(14) The number 0 belongs to a concept $F =_{df}$ all things are non-*F*s [in symbols, $\forall x \neg (Fx)$].

And it is also for that same reason that Frege (1980b, sec. 55) suggests the following, rather awkward definition of the number 1:

(15) The number 1 belongs to a concept $F =_{df}$ not all things are non-*F*s and if any things are *F*s, then they are the same [in symbols: $\neg\forall x \neg (Fx) \wedge \forall x \forall y((Fx \wedge Fy) \rightarrow x = y)$].

This definition could have been made simpler by replacing "not all things are non-*F*s [$\neg\forall x\neg(Fx)$]" by "there is an F [$\exists x(Fx)$]." However, that option was not open to Frege, for it meant, from his point of view, that the number 1 was defined in terms of the word "one," which means the same.

The realization that Frege was compelled to define the number 0 in terms of the universal quantifier as in (14) enables an understanding of his otherwise rather puzzling thesis about existence advanced in §53 of *Grundlagen*, namely that

> Affirmation of existence is in fact nothing but denial of the number nought.

6 For similar remarks, see also Frege (1984a, 152–153; 1979, 14, 21, 61; and 1980a, 101–102).

This might be called the *Existence-Zero thesis*, or EZ for short. EZ would seem puzzling considering how Frege (1980b, sec. 74) ultimately defined the number 0:

(16) $0 =_{\text{df}}$ the number of objects that are not self-identical.

If EZ were based on this definition of the number 0, then what it says could be formulated thus:

(17) There exists an F[7] $\leftrightarrow$ the number of Fs $\neq$ the number of objects that are not self-identical.

But (17) does not say the same as EZ. To see this, note that for Frege (1980b, sec. 73), the right-hand side of (17) says that the concept F is not equinumerous to the concept *non-self-identical object*, where two concepts G and H are said to be equinumerous just in case there is a one-one correlation between the Gs and the Hs. So what (17) says is in fact the following:

(18) There exists an $F \leftrightarrow \neg$(there is a one-one correlation between the Fs and the non-self-identical objects).

This biconditional does hold: if there exists no F, then trivially there will be a one-one correlation between the Fs and the non-self-identical objects, and *vice versa*. However, the right-hand side of (18) contains the expression "there is a one-one correlation" which is of the form "there exists an F," that is, of the same form as the left-hand side. Thus, (18) cannot be viewed as offering an explanation of what existence is, whereas that is what EZ is supposed to do: it is supposed to explain the notion of existence in terms of the number 0.

The expression "nothing but" used in the above statement of EZ indicates that for Frege, the relationship between affirmation of existence and denial of the number 0 holds by definition, that is, that EZ is true by virtue of the meaning of "exists." That would make sense if, at the time of writing *Grundlagen* §53, Frege thought that the number 0 could be defined as in (14). For, then, the following two biconditionals would hold:

(19) $\exists x(Fx) \leftrightarrow \neg\forall x\neg(Fx) \leftrightarrow \neg$ (the number 0 belongs to F).

7 This formulation of the notion of affirmation of existence is to be preferred to "Fs exist," which might be wrongly interpreted as saying that there is more than one F.

doi: 10.48106/dial.v74.i2

The first biconditional holds because, as noted above, a statement of the form "There is at least one F" or "$\exists x(Fx)$" is expressed in Frege's concept-script as "$\vdash\!\!\!\smile^{\mathfrak{a}}\!\!\!\top F(\mathfrak{a})$," or in modern notation, "$\neg\forall x\neg(Fx)$"; and the second biconditional is a corollary of (14). Thus, (19) is a simple consequence of two definitions, and to that extent, could be regarded as a definitional truth itself. Hence, affirmation of existence—"$\exists x(Fx)$"—is nothing but denial of the number 0—"$\neg$ (the number 0 belongs to F)."

Incidentally, the fact that EZ makes better sense when the number 0 is understood in the sense of (14) suggests that *Grundlagen* §53, where the thesis is advanced, reflects his early view of numbers as properties of concepts rather than his mature view of numbers as objects. This is further supported by his remarks in §53 that "existence is analogous to number" and that "existence is a property of concepts." So when Frege wrote at the beginning of §56 that the definitions in §55 "suggest themselves so spontaneously in the light of our previous results, that we shall have to go into the reasons why they cannot be reckoned satisfactory," he was renouncing his own early view of numbers as properties of concepts.

One might object that Frege's fundamental insight that a statement of number contains an assertion about a concept, which was first put forward in §46 of Part III and then reiterated at the beginning of §55 as the main lesson of Part III, continued to be upheld even in *Grundgesetze* where Frege (2013, IX) calls it "[t]he basis for my results," and that this suggests that there is no discontinuity between Frege's view of number in Part III of *Grundlagen* and his later view. But that is no objection, for that insight itself is compatible with both the early view of numbers as properties of concepts and the later view of numbers as objects. In fact, the very reason that the insight is compatible with the latter is that Frege's number-objects, as extensions of concepts, are proxies for properties of concepts.

One might also object that since in *Grundlagen* §38, Frege draws the distinction between proper names and concept words, and classifies the word "one" as a proper name, and since in §51, he declares that "The business of a general concept word"—a word "used with the indefinite article or in the plural without any article"—"is precisely to signify a concept," he must have already believed in Part III of *Grundlagen* that number words such as "one" refer to objects. But this objection assumes, wrongly, that in the earlier parts of *Grundlagen* Frege already upheld his (1984c) later dichotomy between expressions referring to objects, namely proper names, and those referring to concepts, namely predicates. Frege indeed says in §51 of Part III that "when conjoined with the definite article or a demonstrative pronoun" "[a general concept word] can be counted as the proper name of a[n object]."[8] However, in this context, "general concept word" means an expression for a first-level concept such as "satellite of the Earth." As is clear from his ensuing remark that "It is to concepts of just this kind (for example, satellite of the Earth) that the number 1 belongs," the word "number," when combined with the definite article, is meant to refer not to an object[9] but to a property that belongs to first-level concepts. In other words, since numbers are second-level properties, the word "number," when conjoined with "the," refers to a second-level property, and so does not behave like a general concept word which refers to an object when preceded by "the." Also, recall in this connection the fact that when Frege (1980b, sec. 55) gives definitions of individual numbers conceived as properties of concepts, he does so in the context "The number *n* belongs to a concept *F*," apparently thinking that expressions of the form "the number *n*" refer to properties of concepts. So Frege's (1980b, sec. 57) realization that "In the proposition 'the number 0 belongs to the concept *F*,' 0 is only an element in the predicate"—namely the second-level predicate "the number 0 belongs to"—and hence cannot denote a second-level property in its own right represents a profound break from his earlier view of number words as referring to second-level properties (despite being proper names).

In light of the above considerations it seems reasonable to hypothesize that the 1884 *Grundlagen* was not conceived and written in its entirety in response to Carl Stumpf's suggestion, in a letter dated September 9, 1882, of "explain[ing] your line of thought first in ordinary language" (Frege 1980a, 172). It is more likely that Frege set out to rewrite in ordinary language the symbolic parts of his "nearly completed" "book in which I treat the concept of number." And, while doing so, he may have come up with the objections raised in *Grundlagen* §56 to his early view of numbers as properties of concepts, and been led to the conclusion that numbers must be objects instead. The first three parts of *Grundlagen* could be the parts of the discarded book that were salvaged.

8 In the original, the word "thing [*Ding*]" is used, because the comment was made in response to Schröder's claim that abstraction "has the effect of turning what was the name of the thing into a concept applicable to more than one thing" (Frege 1980b, sec. 50).

9 Frege (1980b, sec. 45) describes the word "one" as "the proper name of an object of mathematical study," but the word "object" here does not necessarily mean what it means when he (1980b, sec. 57) concludes that numbers are objects (as opposed to properties or relations).

doi: 10.48106/dial.v74.i2

The conjecture that the first three parts of *Grundlagen* contain Frege's early reflections on number has direct textual support in the "Notes for Ludwig Darmstaedter":

> I started out from mathematics. The most pressing need, it seemed to me, was to provide this science with a better foundation. I soon realized that number is not a heap, a series of things, nor a property of a heap either, but that in stating a number which we have arrived at as the result of counting we are making a statement about a concept. [...] The logical imperfections of language stood in the way of such investigations. I tried to overcome these obstacles with my concept-script. In this way I was led from mathematics to logic. (1979, 253)

The third sentence in this quote reads like a quick summary of the first three parts of *Grundlagen*. Thus, if the narrative is to be believed, Frege had obtained all the results of those parts of *Grundlagen*, including his fundamental insight about the content of a statement of number, before he even conceived the idea of a concept-script. The concept-script was later invented as a means to overcome the obstacles he encountered while carrying out the further investigations, using ordinary language, into analysis of arithmetical concepts and proof of arithmetical truths. So Frege's claim in the 1882 letter that "I have now nearly completed a book" on number could be understood as saying that those further investigations that caused him difficulties due to the "logical imperfections of language" have been nearly completed with the help of the newly invented concept-script. The nontechnical parts of the book—Parts I–III of *Grundlagen*—had been completed before its invention.

To return to the main issue of this section, Frege's goal of providing analysis of arithmetical concepts in purely logical terms meant that he could not adopt the existential quantifier as a primitive. Since existential statements—including those of the form "Some *M* are *P*"—have the meaning of "there is at least one ...," Frege needed to paraphrase them so as to avoid making reference to the numerical notion of one. This he (1980b, sec. 55) achieved by defining the number 0 in terms of a universal negative ("∀¬"), which allowed him to paraphrase an existential statement in purely logical terms as a negative universal negative ("¬∀¬"), that is, as a "denial of the number nought" (1980b, sec. 53). Thus, the fact that for Frege, affirmation of existence is nothing but denial of the number 0 is explained by, and hence adds support to, the conjecture that he was forced to adopt the universal quantifier as a primitive by his felt need to avoid using an existential quantifier in his definitions of the numbers 0 and 1. Of course, in the end—in *Grundlagen* §56—he abandoned the definitions given in §55, including (14) and (15), and opted to define explicitly each individual number as the number of *F*s for some suitable concept *F* as illustrated in (16). However, the point remains that the definitions of *Grundlagen* §55 along with the thesis EZ of §53 are likely to have been part of his early reflections on number and so to have formed "the train of thought that led [him] to [his] [concept-script]" (Frege 1967, 8), including the decision to adopt the universal quantifier as a primitive.

4 Conclusion

The preceding sections have provided three possible explanations—two technical and one philosophical—of Frege's adoption of the universal quantifier as a primitive in his concept-script. This concluding section briefly discusses their relative merits.

As noted at the beginning of this paper, Frege nowhere says anything about why he took the universal, rather than the existential, quantifier as primitive. To that extent one could not reach a definite conclusion as to which of the three possible reasons, if any, was the real reason for Frege's adoption of the universal quantifier as a primitive. Perhaps it is more likely than not that to varying degrees all three of them contributed to and helped cement his decision.

That said, the question could be raised as to which of the three explanations provides the strongest justification for taking the universal quantifier as primitive. And from this point of view, the most satisfactory explanation seems to be the third one. Given the interdefinability of the universal and existential quantifiers, the first two explanations alone do not seem sufficient to make unavoidable the use of the universal quantifier as a primitive. Admittedly, it would have been unnatural and inefficient to use the existential quantifier as a primitive considering that the concept-script has conditionality as the sole binary truth-function; still, it was not an impossibility.

By contrast, the philosophical explanation shows that Frege had no alternative but to adopt the universal quantifier as a primitive. For, given his recurring theme that the existential quantifier involves the notion of "at least one," using it as a primitive would have conflicted with his goal of analyzing arithmetical concepts, especially the concepts of 0 and 1, in purely logical terms.

Relatedly, this explanation has an additional, decisive advantage: it renders understandable Frege's otherwise puzzling silence on the interdefinability of the universal and existential quantifiers. As noted in section 1, he addresses in detail the interdefinability of conditionality and conjunction and explains why he chose the former as a primitive (1967, sec. 7). Thus, as Macbeth (2005, 4) rightly points out, "Had he thought that there were two logically admissible quantifiers usable for the expression of generality, [...] he would have said so." But he did not say so, and this fact indicates that he did not think that the universal and existential quantifiers are equally admissible. And one can understand why given the third explanation for Frege's adoption of the universal quantifier as a primitive. Taking the existential quantifier as an equally admissible primitive would have amounted to allowing into logic what is apparently an arithmetical notion—the notion of one—which is unacceptable from his logicist viewpoint.*

Joongol Kim
Sogang University
joongolk@sogang.ac.kr

References

COOK, Roy T. 2013. "How to Read *Grundgesetze*." in *Basic Laws of Arithmetic*, pp. A1–A42. Oxford: Oxford University Press. Translated and edited by Philip A. Ebert and Marcus Rossberg, with Crispin Wright.

DUMMETT, Michael A. E. 1973. *Frege: Philosophy of Language*. London: Gerald Duckworth & Co.

—. 1981. *Frege: Philosophy of Language*. 2nd ed. Cambridge, Massachusetts: Harvard University Press. First edition: Dummett (1973).

FREGE, Gottlob. 1884. *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner. Reissued as Frege (1961).

—. 1891. *Function und Begriff. Vortrag, gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaft*. Jena: Hermann Pohle. Reprinted in Frege (2008, 1–22).

* Thanks are owed to Robert Michels—the guest editor of the special issue of *Dialectica* on the formalization of arguments—and the anonymous reviewers for helpful comments.

—. 1892a. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik NF* 100: 25–50. Reprinted in Frege (2008, 23–46).

—. 1892b. "Über Begriff und Gegenstand." *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192–205. Reprinted in Frege (2008, 47–60).

—. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, 1. Band*. Jena: Hermann Pohle. Reissued as Frege (1966).

—. 1904. "Was ist eine Funktion?" in *Festschrift Ludwig Boltzmann gewidmet zum sechzigsten Geburtstage, 20. Februar 1904*, edited by Stefan Meyer, pp. 656–666. Leipzig: Johann Ambrosius Barth. Reprinted in Frege (2008, 61–69).

—. 1950. *The Foundations of Arithmetic*. Oxford: Basil Blackwell Publishers. Translation of Frege (1884) by J.L. Austin, doi:10.4324/9781003072874.

—. 1961. *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl*. Hildesheim: Georg Olms.

—. 1966. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, 1. Band*. Hildesheim: Georg Olms. Reprografischer Nachdruck der Ausgabe Frege (1893).

—. 1967. "Begriffsschrift: A Formula Language Modeled on that of Arithmetic, for Pure Thought." in *From Frege to Gödel: A Source Book in Mathematical Logic 1879–1931*, edited by Jan van Heijenoort, pp. 1–82. Cambridge, Massachusetts: Harvard University Press. Translated by Stephan Bauer-Mengelberg.

—. 1969. *Nachgelassene Schriften*. Hamburg: Felix Meiner Verlag. Edited by Hans Hermes, Friedrich Kambartel and Friedrich Kaulbach, doi:10.28937/978-3-7873-3159-8.

—. 1976. *Wissenschaftlicher Briefwechsel*. Hamburg: Felix Meiner Verlag. Edited by Gottfried Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel and Albert Veraart, doi:10.28937/978-3-7873-2549-8.

—. 1979. *Posthumous Writings*. Oxford: Basil Blackwell Publishers. Edited by Hans Hermes, Friedrich Kambartel and Friedrich Kaulbach; translation of Frege (1969) by Peter Long and Roger Whit.

—. 1980a. *Philosophical and Mathematical Correspondence*. Oxford: Basil Blackwell Publishers. Edited by Gottfried Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel and Albert Veraart; translation of Frege (1976) by Hans Kaal and abridged by Brian McGuinness.

—. 1980b. *The Foundations of Arithmetic*. 2nd ed. Evanston, Illinois: Northwestern University Press. Translation of Frege (1884) by J.L. Austin; first edition: Frege (1950).

—. 1984a. "Function and Concept." in *Collected Papers on Mathematics, Logic, and Philosophy*, pp. 137–156. Oxford: Basil Blackwell Publishers. Translation of Frege (1891) by Peter Geach.

—. 1984b. "On Sense and Meaning." in *Collected Papers on Mathematics, Logic, and Philosophy*, pp. 157–177. Oxford: Basil Blackwell Publishers. Translation of Frege (1892a) by Max Black.

—. 1984c. “On Concept and Object.” in *Collected Papers on Mathematics, Logic, and Philosophy*, pp. 182–194. Oxford: Basil Blackwell Publishers. Translation of Frege (1892b) by Peter Geach.

—. 1984d. “What is a Function?” in *Collected Papers on Mathematics, Logic, and Philosophy*, pp. 285–292. Oxford: Basil Blackwell Publishers. Translation of Frege (1904) by Peter Geach.

—. 2008. *Funktion, Begriff, Bedeutung. Fünf logische Studien*. Göttingen: Vandenhoeck & Ruprecht. Edited and introduced by Günther Patzig.

—. 2013. *Basic Laws of Arithmetic*. Oxford: Oxford University Press. Translated and edited by Philip A. Ebert and Marcus Rossberg, with Crispin Wright.

Kim, Joongol. 2011. “Frege’s Context Principle: An Interpretation.” *Pacific Philosophical Quarterly* 92(2): 193–213, doi:10.1111/j.1468-0114.2011.01391.x.

—. 2013. “What are Numbers?” *Synthese* 190(6): 1099–1112, doi:10.1007/s11229-011-9883-y.

—. 2015. “A Logical Foundation of Arithmetic.” *Studia Logica: An International Journal for Symbolic Logic* 103(1): 113–144, doi:10.1007/s11225-014-9551-6.

—. 2020. “The Primacy of the Universal Quantifier in Frege’s Concept-Script.” *Dialectica* 74(2): 279–298, doi:10.48106/dial.v74.i2.04.

Kneale, William C. and Kneale, Martha. 1962. *The Development of Logic*. Oxford: Oxford University Press.

Macbeth, Danielle. 2005. *Frege’s Logic*. Cambridge, Massachusetts: Harvard University Press, doi:10.4159/9780674040397.

Holistic Inferential Criteria of Adequate Formalization

Friedrich Reinmuth

Peregrin and Svoboda propose an inferential and holistic approach to formalization, and a similar approach (to correctness) is considered by Brun. However, while the inferential criteria of adequacy explicitly endorsed by these authors may be holistic "in spirit," they are formulated for single formulas. More importantly, they allow the trivialization of equivalence and face problems when materially correct arguments come into play. Against this background, this paper tries to motivate holistic inferential criteria that compel us to distinguish carefully between non-trivially equivalent formalizations as well as between materially and logically correct arguments on an inferential basis.

The first section of the paper (section 1) discusses some problems faced by the inferential (and semantic) criteria of adequacy proposed by Brun (2004, 2012, 2014) and Peregrin and Svoboda (2013, 2017). According to these authors, inferential criteria are to be applied holistically. Yet, their criteria are formulated for single formulas, which leads to some application problems. More importantly, the criteria face problems that are due to their lack of syntactic sensitivity, e.g. the problem of trivialized equivalence. It is argued that postulating additional subsidiary criteria is not a satisfying option if one wants to defend an inferentialist approach to formalization and holds that there is a systematic connection between syntactic features and inferential roles. In contrast, Brun's postulate of hierarchical structure should be accepted as an important systematic constraint on our judgments of adequacy, albeit one that appears weaker than hoped for in some cases.

In section 2, I will propose holistic inferential criteria in the spirit of Peregrin and Svoboda and provide a more detailed discussion of some of the problems raised in section 1. While the criteria can be used to assess the adequacy of formalizations relative to sets of "sample arguments," they are too weak to distinguish properly between non-trivially equivalent formalizations, and face difficulties when materially correct arguments are taken into consideration. Section 3 then turns to the role of sentences in inferential contexts that are not reduced to premise-conclusion arguments, namely, to informal derivations. It is argued that if we see the development of calculi as an attempt to account for the logical correctness of arguments in a systematic way and take the distinctions in inferential roles they offer seriously, we have good reasons to strengthen our inferential criteria so that they compel us to choose between non-trivially equivalent formalizations and to distinguish carefully between materially and logically correct arguments. The last section (section 4) indicates some directions for future research.

1 Adequate Formalization, Inferential Criteria, and Trivialized Equivalence

Brun, who has provided a detailed and thorough investigation of the problems of adequate formalization (2004), and most other authors assume that a basic requirement of adequacy is that formalizations do not render intuitively incorrect arguments formally correct (correctness). Some authors, notably Baumgartner and Lampert (2008; 2010), also advocate views of different strength to the effect that adequate formalizations should not render intuitively correct arguments formally incorrect (completeness).

Peregrin and Svoboda have recently put forward an account of logic in terms of reflective equilibrium in which they promote two such criteria as "inferential" criteria of adequate formalization which they contrast with and prefer to so called "semantic" criteria which rely on comparisons of truth conditions (see 2017, esp. ch. 5 and 6). For them, adequate formalizations (logical forms) "are products of the logicians' efforts to account for the inferential structure of a language, especially to envisage the roles of individual statements within the structure" (2017, 4). Since the formalization of a sentence *S* aims at "making explicit the place of [...] *S* within the inferential structure of its natural language by means of associating *S* with a formula of a logical language" (Peregrin and Svoboda 2017, 69), inferential criteria provide the measure of success.

Before discussing the criteria, I want to introduce the main example used in the following, (the conclusion of) "an inference traditionally attributed to De Morgan" (Brun 2012, 325):

De Morgan's argument (DMA).

Every horse is an animal.
∴ Every head of a horse is a head of an animal.

For this example, Brun (2012) discusses the formalization

(P1) $\forall x(Hx \rightarrow Jx)$

of the premise

(PDM) Every horse is an animal

and the formalizations

(C1) $\forall x(Fx \rightarrow Gx)$
(C2) $\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$
(C3) $\forall x \forall y(Hy \wedge Ixy \rightarrow Jy \wedge Ixy)$
(C4) $\forall x(Hx \wedge \exists y Iyx \rightarrow Jx \wedge \exists y Iyx)$

of the conclusion

(CDM) Every head of a horse is a head of an animal

with a correspondence scheme which agrees with the following, in which entries for "a" and "b" are added:

Correspondence scheme: heads of horses.
Fx : x is a head of a horse
Gx : x is a head of an animal
Hx : x is a horse
Ixy : x is a head of y
Jx : x is an animal
a : Fury
b : Batu[1]

Note that Peregrin and Svoboda do not consider the correspondence scheme, which assigns natural language expressions to the non-logical symbols in the formalizing formula, to be part of the formalization. I will follow Peregrin

1 The argument and the formalizations (C1), (C2), and (C3) are also extensively discussed in Brun (2004), while (C4) was introduced by Lampert and Baumgartner (2010).

doi: 10.48106/dial.v74.i2

and Svoboda in this, because correspondence schemes provide a kind of formalization at the atomic level, while I want to pursue an account of the adequacy of formalizations that does not take for granted the adequacy of other formalizations.

The first inferential criterion proposed by Peregrin and Svoboda is labelled "*principle of reliability*" and provides a criterion for the correctness of formalizations:

> REL. Φ counts as an adequate formalization of the sentence *S* in the logical system **L** only if the following holds: If an argument form in which Φ occurs as a premise or as the conclusion is valid in **L**, then all its perspicuous natural language instances in which *S* appears as a natural language instance of Φ are intuitively correct arguments. (2017, 70) [2]

If we assume, for example, that De Morgan's argument is an instance of the classically valid

$$\forall x(Hx \rightarrow Jx)$$
⍰ $\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$

then it has to be intuitively correct for (C2) to be an adequate formalization of (CDM) if the logical system is classical logic (which will be the general framework in the following).

As Peregrin and Svoboda point out, (REL) is quite similar to an inferential criterion of correctness proposed by Brun (2014, 104). Peregrin and Svoboda also propose a (comparative) completeness criterion with their "*principle of ambitiousness*":

> AMB. Φ is the more adequate formalization of the sentence *S* in the logical system **L** the more natural language arguments in which *S* occurs as a premise or as the conclusion, which fall into the intended scope of **L** and which are intuitively perspicuous and correct, are instances of valid argument forms of **L** in which Φ appears as the formalization of *S*. (2017, 71) [3]

2 To simplify the following discussion, I will largely ignore the restriction to perspicuous arguments.

3 The intended scope of a logical system consists "of the arguments whose correctness is to be demonstrable by means of the [logical] language" (Peregrin and Svoboda 2017, 64–65). Peregrin

It seems clear that (AMB) is intended as a means of comparing formalizations where at least the one to be judged to be more adequate meets (REL). It also seems clear that "more natural language arguments" is to be understood in the sense of "the larger and more varied" (Peregrin and Svoboda 2017, 72). Inferential criteria such as (REL) and (AMB) that are not restricted to manageable sets of arguments can hardly be used to judge formalizations to be (more) adequate as their application obviously faces a, as Baumgartner and Lampert put it, "*termination problem*" (2008, 97).

According to Peregrin and Svoboda, the following holds:

> We can, and [...] do, base our (provisional) selection of the formalization on considering a limited number of sample arguments. Thus, a humanly manageable version of (REL) would not simply require that *all* perspicuous natural language instances of a valid argument form in which Φ occurs in place of *S* are intuitively correct, but only that this holds for those which are among the actual set of sample arguments. Similarly, we could easily reformulate (AMB) so that it (tentatively) prefers the formalization which merely reveals more intuitively correct *sample* arguments as logically correct. In such case, of course, the procedure of selecting the preferable (tentatively adequate) formalization would yield more reliable results the larger and more varied the set of sample arguments is. (2017, 72)

Moreover, they as well as Brun stress that the (intended) application of their respective criteria presupposes that "the formalizations of all sentences, save the one on which we focus our attention, is unproblematic" (Peregrin and Svoboda 2017, 70; see Brun 2014, 104).

All three authors agree that this, as Brun puts it,

> motivates a holistic approach to formalizing which proceeds by bootstrapping: as a starting point, some formalizations are presumed to be correct and used to test others, but such tests may also lead to revising some of the starting-point formalizations [...]. (2014, 104–105)

---

and Svoboda (2017, 71) relate (AMB) to the definition of the completeness of formalizations in (Baumgartner and Lampert 2008, 103).

doi: 10.48106/dial.v74.i2

However, while it may be the case that "we always test a kind of holistic structure, though we perceive it as testing the single formula" (Peregrin and Svoboda 2017, 70), the criteria are formulated for single formulas. This leads to another application problem: even if we restrict our attention to manageable sets of arguments and even if we assume certain formalizations to be adequate, we still cannot apply (REL) and (AMB) in a "humanly manageable" way. Assume, for example, that our sample set only consists of

(1) Every head of a horse is a head of an animal.
⍰ Batu is a head of an animal.

and that we consider (1) not to be intuitively correct. Assume that we want to use (REL) to assess the correctness of (C2) as a formalization of (CDM). Then, we still would have to go through all valid argument forms in which (C2) appears as the only premise and check if one of the conclusions is an adequate formalization of the conclusion of (1). Only if no such argument form exists can we judge (C2) to fulfill the criterion of correctness for the sample set. This holds even if we assume that the conclusion of (1) is adequately formalized by

$$\exists y(Jy \wedge Iby)$$

That the latter formula does not follow from (C2) does not entail that there are no adequate formalizations of the conclusion of (1) which follow from (C2). So, in order to apply (REL) (or AMB), we do not only have to assume that other formalizations are "unproblematic," but that they are "fixed" (Peregrin and Svoboda 2017, 75).

However, this makes it difficult to assess the respective merits of alternative formalizations of a sentence since we might want to rely on different formalizations of other sentences. For example, if we want to test (C1), we might want to use another formalization of the conclusion of (1), namely, "*Gb*."

Apart from facing application problems, (REL) and (AMB) are highly insensitive to the syntactic features of formalized sentences and their formalizations. Consequently, the "two principles alone [...] do not seem to be sufficient. The main problem is that they do not distinguish between very dissimilar equivalent formulas" (Peregrin and Svoboda 2017, 72). The reason is that (REL) and (AMB) only consider the validity of argument forms, which for many logical systems, e.g. classical logic, is not affected by the substitution of equivalent formulas. This failure to distinguish between equivalent formulas opens the way to “unacceptably trivial proofs for inferences involving equivalent sentences” (Brun 2014, 105). As an example, consider (C3) and (C4) and the following two sentences:

(CDM-a) Every horse that has a head is an animal that has *that* head.
(CDM-b) Every horse that has a head is an animal that has *a* head.

Since (C3) and (C4) are equivalent, they can be substituted for each other in classically valid argument forms. Now assume that (C3) is an adequate formalization of (CDM-a) and (C4) is an adequate formalization of (CDM-b). Then, (C4), being equivalent to (C3), should also be considered an adequate formalization of (CDM-a), as substituting (C4) for (C3) does not change the validity of the argument forms used to establish the adequacy of (C3). Similarly, (C3) should also be considered an adequate formalization of (CDM-b). Consequently, one could use just one of the two formulas as an adequate formalization for both sentences and “capture” the intuitive equivalence of the sentences by a trivial argument form in which the one premise is identical to the conclusion. This seems worrisome if one holds that “equivalence is subject to logical proof and should not be trivialized by simply choosing the same formalization for any two equivalent sentences” (Brun 2014, 101).

The trivialization of equivalence is a symptom of the lack of “syntactic sensitivity” of (REL) and (AMB)—and similar criteria that are formulated for premise-conclusion arguments. As Brun rightly remarks: “If there are sentences which are in a non-trivial way equivalent […], this is a matter not only of their truth-conditions but also of their syntactical features” (2014, 107). Brun, Svoboda and Peregrin also point to a desire for a compositional account of logical analysis, which seems to require some systematic sensitivity to syntactic features of the formalized sentences (Brun 2012, 328; 2014, 108; Peregrin and Svoboda 2017, 73).

In order to achieve “some kind of anchoring of the ‘logical form’ in the grammatical form of the statement of which it is a logical form” (Peregrin and Svoboda 2017, 73), they propose additional criteria such as the following:

> PT. Other things being equal, Φ is the more adequate formalization of the statement *S* in the logical system **L** the more the grammatical structure of Φ is similar to that of *S*. (2017, 72) [4]

[4] Brun gives the following examples: “the logical symbols in a formalization Φ must have a counterpart in *S*; Φ’s correspondence scheme must not include ordinary language expressions not occurring in *S*” (2012, 326–327).

However, Peregrin and Svoboda consider these criteria to be "more-or-less auxiliary" (2013, p. 2919). Brun comments:

> Rules operating on the syntactical surface implicitly guide the common practice of formalization, but if they are not to classify a great deal of standard formalizations as inadequate, they cannot be taken as strict requirements but must be interpreted very liberally or qualified by a virtually endless list of exceptions. (2014, 107)

Brun suggests that a more sophisticated grammar (and maybe also a more sophisticated logical system) is needed for precise and working syntactic criteria (2012, 328; 2014, 109). Peregrin and Svoboda seem to suggest that the very project of formalization and the development of logical systems go hand in hand with developing a (logical) syntax for the sentences in the intended scope of the logic which is projected into the syntax of the developed logical system(s) (see 2017, esp. chap. 7.3). They seem to presuppose that the non-logical symbols of logical languages are parameters that can be used to replace natural language expressions in order to arrive at (logical) forms of sentences and arguments which can then again be instantiated by natural language sentences and arguments (see 2017, esp. chap. 2.3). In this vein, they speak of "the theory of natural language syntax that has been projected into the language of predicate logic" (2017, 52).

However, if the grammatical theory we use applying (PT) is essentially a logico-syntactic theory that finds expression in the syntax of the logical system in question, applications of (PT) to formalizations of a natural language sentence *S* would presuppose that we have already settled on a formalization of *S* in order to test whether the grammatical structure of formalizations is (more) similar to the grammatical structure of *S*.

As indicated, all three authors seem to assume some connection between syntactic features and inferential roles. Given this presumed connection, one might ask why one does not try to approach syntactic features via inferential roles instead of postulating additional "rules of thumb" or hoping for a more sophisticated grammar, an approach Peregrin and Svoboda seem to advocate and which Brun seems to consider as an option (Brun 2014, 115).

If one tries to develop such an approach, one is well advised to impose systematic constraints on the choice of formalizations. Brun, who advocates systematic formalization, distinguishes two aspects, namely "formalizing analogous sentences analogously," and "formalizing step by step" (2012, 327). However, Brun is as skeptical about the strict application of these precepts as he is regarding surface rules:

> The common theme behind surface rules and the principles of analogous and step-by-step formalization is that they all become more convincing the more we can spell out in a precise and general manner how sentences are to be formalized based on some syntactic description. (2014, 109)

Again, one might ask why one should not rather use inferential criteria to determine which logico-syntactic structure one should impose on natural language sentences. Why not use inferential criteria to specify "the classes of sentences which can be formalized as instances of the same scheme" (Brun 2014, 108) and base syntactic descriptions on how sentences are to be formalized w.r.t. inferential criteria?

While the syntactic criteria and the precepts of formalizing step-by-step and analogously are, according to their authors, not strictly applicable, Brun also offers a powerful postulate (or criterion) for adequate formalizations that enforces systematic syntactic relations between non-equivalent adequate formalizations of the same sentence, the "postulate of hierarchical structure":

> PHS. If $\Phi = \langle \varphi, \kappa \rangle$ and $\Psi = \langle \psi, \kappa \rangle$ are two adequate formalizations of a sentence S in $\mathbf{L}$ then either (i) Φ and Ψ are equivalent, or (ii) Φ is more specific than Ψ, or (iii) Ψ is more specific than Φ, or (iv) there is an adequate formalization of S that is more specific than both Φ and Ψ. (2014, 109) [5]

One purpose of (PHS) is that it lets us "argue about the adequacy of formalizations by pointing out that they could (not) plausibly be the product of a systematic procedure" (Brun 2014, 109). The deeper motivation is that it ensures "that the various adequate formalizations of an inference constitute

5 Note that for Brun formalizations also contain a correspondence scheme. For this formulation of (PHS) with a fixed correspondence scheme κ, $\Phi = \langle \varphi, \kappa \rangle$ is ($\mathbf{L}$-)equivalent to $\Psi = \langle \psi, \kappa \rangle$ iff φ and ψ are ($\mathbf{L}$-)equivalent; and Φ is more specific than Ψ "iff φ can be generated from ψ by substitutions $[\alpha/\beta]$ such that either (i) α is a sentence-letter occurring in ψ and β is a formula containing at least one sentential connective or a predicate-letter, or (ii) α is an n-place predicate-letter occurring in ψ and β is an open formula with n free variables containing at least one sentential connective, quantifier or predicate-letter with more than n places" (Brun 2014, 109).

doi: 10.48106/dial.v74.i2

a certain unity" (Brun 2014, 110). Postulates like (PHS) are needed if we want adequate formalization to play a part in a systematic account of the (in)correctness of inferences, e.g. by reaching a state of reflective equilibrium, as envisaged by Brun and Peregrin and Svoboda.

Still, (PHS) explicitly allows equivalent formalizations of the same sentence. Moreover, as noted by Lampert and Baumgartner (2010, 95), (C4) and (C2) are both more specific than (C1). Thus, while one can rule out (C3) as an adequate formalization of (CDM) if (C1) is an adequate formalization of this sentence, the same does not hold for (C4). So, all (PHS) (or the equivalent criterion (HCS), which Brun uses in his 2012 paper)[6] does is that "it rules out that (C2) and (C4) are both adequate without telling us which one is inadequate" (Brun 2012, 329). Brun also holds that "(C4) and (C2) fare equally well with respect to (TC) and surface rules" (2012, 329) where (TC), Brun's semantic criterion, (with an added explanation) reads:

> TC. A formalization $\langle \varphi, \kappa \rangle$ of a sentence S in a logical system **L** is correct iff for every condition c, for every **L**-interpretation $\langle \mathcal{D}, \mathcal{I} \rangle$ corresponding to c and κ, $\mathcal{I}(\varphi)$ matches the truth value of S in c. An **L**-interpretation corresponding to a condition c and a correspondence scheme $\{\langle \alpha_1, a_1 \rangle, \ldots, \langle \alpha_n, a_n \rangle\}$ is an **L**-structure $\langle \mathcal{D}, \mathcal{I} \rangle$ with domain $\mathcal{D}$ and an interpretation function $\mathcal{I}$, such that $\mathcal{I}(\alpha_i)$ matches the semantic value of a_i in c (for all $1 \leq i \leq n$). (Brun 2014, 105; see 2014, 105–106)

This is due to the fact that "(TC) is not distinctive enough if materially *i*—valid inferences are involved" (Brun 2012, 327), i.e. informally materially correct inferences. Without going into the details of Brun's argument against the adequacy of (C4), we can note that it relies on the "strategy of analogous formalizations" (Brun 2012, 330) and is thus, according to Brun's own standards, not decisive. As we will see in the next section, inferential criteria for premise-conclusion arguments are "not distinctive enough" either if materially correct arguments are involved.

Up to now, the following picture has emerged: the inferential criteria, promoted in particular by Peregrin and Svoboda (as well as Brun's "semantic" criterion (TC) and, to some extent, (PHS)) do not incorporate the presumed

6 (HCS) reads informally: "at least one of two non-equivalent formalizations of the same sentence must be inadequate if neither is more specific than the other and there is not a third adequate formalization more specific than both" (Brun 2012, 329).

systematic relation between syntactic structure and inferential role. While this is especially obvious in the case of non-trivially equivalent formalizations, it also leads to problems when materially correct arguments are involved in the assessment of formalizations. To make up for this, the authors propose auxiliary criteria referring to syntactic features, formalizing step-by-step and the analogous formalization of analogous sentences.

This seems rather strange: if one assumes a systematic connection between syntactic features of sentences and the role they can play in inferences, then inferential criteria of adequacy should not rely on additional side-criteria of dubious applicability to ensure a systematic connection between the syntactic features of sentences and their formalizations. Rather, such a connection should result from the application of inferential criteria.

In the section after the next, I will try to outline such an inferentially oriented approach to the adequacy of formalizations. In the next section, some of the problems raised in this section will be discussed in more detail with respect to holistic inferential criteria in the spirit of (REL) and (AMB).

2 Adequacy and Premise-Conclusion Arguments

As already noted above, Peregrin and Svoboda hold that at least considerations of completeness relative to a logical system have to take into account the "*intended scope* of a logical language, consisting of the arguments whose correctness is to be demonstrable by means of the language" (2017, 64–65). They specify:

> Let us call the set of all the perspicuous arguments which characterize the behavior of *S* within the intended scope of a logical system **L** the **L**-*reference arguments for S* and any of its non-empty subsets which consists of arguments considered during a particular procedure of assessing alternative formalizations the **L**-*sample arguments for S*. (2017, 65)

Note that the intended scope of a logical system is not something given. Which arguments we consider to be (more) important reference arguments is part of the "bootstrapping" that Peregrin and Svoboda describe (2017, 74–76). The need for choosing sample arguments (and, importantly, other inferential contexts) will become clearer once the holistic inferential criteria are formulated. To do this, we need some preparatory definitions. These definitions will be

doi: 10.48106/dial.v74.i2

given for formalizations of English sentences but can easily be generalized. First, we define:

FORMALIZATION-FUNCTION. Φ *is an* **L**-*formalization function for* **S** if and only if

i) **L** is a logical system; and
ii) **S** is a non-empty set of English sentences; and
iii) Φ is a function from **S** to a set of **L**-formulas.

The following table provides examples of first-order formalization functions. The sentences in the domain are noted to the left, while the respective values are noted to the right:

Table 1: Formalization functions ($\Phi 1$), ($\Phi 2$), ($\Phi 3$), and ($\Phi 4$)

Sentences in the domain	*Values for*			
	($\Phi 1$)	($\Phi 2$)	($\Phi 3$)	($\Phi 4$)
(CDM): Every head of a horse is a head of an animal	(C1)	(C2)	(C3)	(C4)
(PDM): Every horse is an animal	(P1)			
Batu is a head of a horse	Fb	$\exists y(Hy \wedge Iby)$		
Batu is a head of an animal	Gb	$\exists y(Jy \wedge Iby)$		

The value of a formalization function Φ for a natural language sentence S will be called the *formalization of* S *w.r.t.* Φ. Thus, the four formalization functions differ in their formalizations of (CDM). They agree in their formalization of (PDM), and ($\Phi 1$) also differs from the other three formalization functions in its formalizations of the remaining two sentences.

Now we can define:

INSTANCE OF AN ARGUMENT FORM. *A is an instance of AF w.r.t. the formalization function* Φ iff there are **S** and **L** such that Φ is an **L**-formalization function for **S** and there are sentences $S_1, \ldots, S_n$ ($n \geq 1$) in **S** such that $A = \langle S_1, \ldots, S_n \rangle$ and $AF = \langle \Phi(S_1), \ldots, \Phi(S_n) \rangle$.

If A is an instance of AF w.r.t. Φ, we will call AF a *formalization of* A *w.r.t.* Φ. Let us say that A is *an argument over* **S** iff **S** is a set of English sentences and

A is a non-empty finite sequence such that every member of *A* is an element of **S**. So, for example, (DMA) and

(2) Every head of a horse is a head of an animal.
Batu is a head of a horse.
∴ Batu is a head of an animal.

are arguments over the domain of the formalization functions above.

Note that if Φ is a formalization function for a set **S** of sentences and *A* is an argument over **S**, then there is exactly one formalization of *A* w.r.t. Φ. So, for example,

(3) $\forall x(Hx \rightarrow Jx)$
∴ $\forall x(Fx \rightarrow Gx)$

is the formalization of (DMA) w.r.t. (Φ1), while

$\forall x(Hx \rightarrow Jx)$
∴ $\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$

is its formalization w.r.t. (Φ2).

If *A* is an argument over the domain **S** of an **L**-formalization function Φ, then we will say that *A is* **L**-*correct w.r.t.* Φ iff the formalization of *A* w.r.t. Φ is an **L**-valid argument form. So, for example, (DMA) is classically correct w.r.t. (Φ2), but not w.r.t. (Φ1).

Now, we will formulate relativized criteria in the spirit of (REL) and (AMB) for formalization functions. A relativization to sample classes is not only in order because it may be difficult to survey all arguments over the domain of a formalization function. It also holds—as pointed out above—that we have to decide which arguments to admit to the sample classes and which not. The criteria have the form of definitions, but they refer to the intuitive correctness of arguments and should therefore not be treated as definitions of predicates in terms of other, well-established predicates. For the correctness of formalization functions, we postulate:

COR. Φ *is a correct* **L**-*formalization function for* **S** *w.r.t.* **A** iff

i) Φ is an **L**-formalization function for **S**; and
ii) **A** is a non-empty set of arguments over **S**; and

doi: 10.48106/dial.v74.i2

iii) for every argument A in **A** it holds: if A is **L**-correct w.r.t. Φ, then A is an intuitively correct argument

So, for example, if we consider just the unit set of (2) and take classical first-order logic as the logical system, (Φ1), (Φ2), (Φ3) and (Φ4) are correct formalization functions for their common domain w.r.t. this set if we take (2) to be an intuitively correct argument (which I will assume for the following). Note that we will only consider classical first-order logic for the formal side and therefore largely omit mentioning of the logical system in the remaining part of this section.

We set for complete formalization functions:

COMP. Φ *is an* **L**-*complete formalization function for* **S** *w.r.t.* **A** iff

i) Φ is an **L**-formalization function for **S**; and
ii) **A** is a non-empty set of arguments over **S**; and
iii) for every argument A in **A** it holds: if A is an intuitively correct argument, then A is **L**-correct w.r.t. Φ.

So, for example, if we consider again just the unit set of (2) , (Φ1), (Φ2), (Φ3) and (Φ4) are all complete formalization functions w.r.t. this set. However, if we extend the set of arguments to include (DMA) (and consider it to be intuitively correct), only (Φ2), (Φ3) and (Φ4) are complete formalization functions w.r.t. the extended set.

Adequacy w.r.t. a set of arguments over the domain of a formalization function is postulated to consist in correctness and completeness w.r.t. that set:

AD. Φ *is an* **L**-*adequate formalization function for* **S** *w.r.t.* **A** iff

i) Φ is an **L**-correct formalization function for **S** w.r.t. **A**; and
ii) Φ is an **L**-complete formalization function for **S** w.r.t. **A**.

So, for example, if we consider again the set {(DMA), (2)}, (Φ2), (Φ3) and (Φ4) are all adequate formalization functions w.r.t. this set, while (Φ1) is not. Note that if a formalization function is correct, complete, or adequate w.r.t. some set of arguments, it is so w.r.t. every non-empty subset of this set.

To make comparative judgments of correctness, completeness, and adequacy, it seems natural to extend the formalization functions in question by adding new pairs of sentences and formulas and to consider different sets of arguments over the (extended) domain. Surely, it seems advisable to assume that "the procedure of selecting the preferable (tentatively adequate) formalization would yield more reliable results the larger and more varied the set of sample arguments is" (Peregrin and Svoboda 2017, 72). However, we also have to decide which "sample arguments we use to demarcate the scope of the [...] logical system" (Peregrin and Svoboda 2017, 70). The scope of a logical system is not something beyond dispute. So, for example, Lampert and Baumgartner want to use classical first-order logic to cover all kinds of intuitively correct arguments (see 2008; 2010), while Peregrin and Svoboda only want to include "*as many logically correct arguments as possible*" (2017, 71). However, they themselves hold "that no clear boundary between logically correct arguments and those that are correct but not logically correct exists in natural language" (2017, 37). Such a boundary can be drawn w.r.t. a logical system and adequate formalizations but this strategy is not straightforwardly applicable if one still has to determine which formalizations one wants to accept as adequate.

To base our discussion on richer examples, let us consider the following extensions of (Φ2), (Φ3) and (Φ4):

Table 2: Extension of (Φ2), (Φ3) and (Φ4) to (Φ2.1), (Φ3.1) and (Φ4.1)

Sentences in the domain	*Values for*		
	(Φ2.1)	(Φ3.1)	(Φ4.1)
(CDM): Every head of a horse is a head of an animal	(C2)	(C3)	(C4)
(PDM): Every horse is an animal		(P1)	
Batu is a head of a horse		$\exists y(Hy \land Iby)$	
Batu is a head of an animal		$\exists y(Jy \land Iby)$	
(CDM-a): Every horse that has a head is an animal that has that head	(C3)	(C3)	(C4)
(CDM-b): Every horse that has a head is an animal that has a head	(C4)	(C3)	(C4)
Batu is a head of Fury		Iba	
Fury is a horse		Ha	

doi: 10.48106/dial.v74.i2

Fury has a head	$\exists y Iya$
Fury is a horse that has a head	$Ha \land \exists y Iya$
Fury is a horse and Batu is a head of Fury	$Ha \land Iba$
Fury is an animal	Ja
Fury is an animal that has a head	$Ja \land \exists y Iya$
If Fury is a horse that has a head, then Fury is an animal that has a head	$Ha \land \exists y Iya \rightarrow Ja \land \exists y Iya$
Fury is an animal and Batu is a head of Fury	$Ja \land Iba$
If Fury is a horse and Batu is a head of Fury, then Fury is an animal and Batu is a head of Fury	$Ha \land Iba \rightarrow Ja \land Iba$
If Batu is a head of Fury, then Batu is a head of an animal	$Iba \rightarrow \exists y(Jy \land Iby)$
It holds for everything: if it is a horse and Batu is a head of it, then it is an animal and Batu is a head of it	$\forall y(Hy \land Iby \rightarrow Jy \land Iby)$
Everything is a head of an animal	$\forall x \exists y(Jy \land Ixy)$

The extended formalization functions have a common domain and differ only in their formalizations of (CDM), and (CDM-a) and (CDM-b), respectively.

Now consider the following arguments over the common domain of (Φ2.1), (Φ3.1), (Φ4.1):

(4) Every horse that has a head is an animal that has that head.
⍰ Every horse that has a head is an animal that has a head.

and

(5) Every horse that has a head is an animal that has a head.
⍰ Every horse that has a head is an animal that has that head.

If we assume that both arguments are intuitively correct, (Φ2.1), (Φ3.1), (Φ4.1) are adequate w.r.t. {(DMA), (2) , (4), (5)}. The difference is that (Φ3.1) and (Φ4.1) trivialize the equivalence between (CDM-a) and (CDM-b).

We can (for our purposes) define two **L**-formalization functions Φ, Φ* to be **L**-*equivalent formalization functions* iff they share the same domain **S** and it holds for every S in **S** that $\Phi(S)$ is **L**-equivalent to $\Phi^*(S)$. According to this definition, (Φ3.1) and (Φ4.1) are equivalent formalization functions w.r.t. classical logic. Thus, they render the same arguments over their common domain classically correct and are not distinguishable regarding their correctness, completeness, or adequacy by applying (COR), (COMP), and (AD). This holds in general: If **L** is a logical system that allows the substitution of **L**-equivalent formulas, e.g. classical logic, then **L**-equivalent formalization functions cannot be distinguished w.r.t. their correctness, completeness or adequacy by (COR), (COMP), and (AD).

Moreover, these criteria face difficulties when materially correct arguments come into play. To see this, let us turn to the relation between (Φ2.1) on the one hand and (Φ3.1) and (Φ4.1) on the other.

Consider the following arguments over the common domain:

(6) Every head of a horse is a head of an animal.
⍰ If Batu is a head of a horse, then Batu is a head of an animal.

(7) Every head of a horse is a head of an animal.
⍰ If Fury is a horse and Batu is a head of Fury, then Fury is an animal and Batu is a head of Fury.

(8) Every head of a horse is a head of an animal.
⍰ If Fury is a horse that has a head, then Fury is an animal that has a head.

If we assume that all three arguments are intuitively correct, (Φ3.1) and (Φ4.1) are adequate w.r.t. {(DMA), (2) , (4), (5), (6), (7), (8)}, while (Φ2.1) is only adequate w.r.t. {(DMA), (2) , (4), (5), (6)}. How could one argue that one should anyhow prefer (Φ2.1)?

First, we can note that the criterion of correctness put forward by Peregrin and Svoboda, namely

*CorArg**: An argument is correct if the step from its premises to its conclusion is a generally acceptable move in an argumentation, or if it can be reconstructed as composed from such generally acceptable moves (2017, 46)

does not clearly rule out any of the arguments as intuitively incorrect if we do not put further constraints on what moves are "generally acceptable." Given their further explanation of their notion of correctness, namely

doi: 10.48106/dial.v74.i2

> that an argument is correct iff it is safe to move from its premises to its conclusion in the sense that whoever accepts the premises cannot reject the conclusion or, more precisely, whoever *does* reject them will be taken to be either unreasonable, or not understanding the language in which they are formulated (2017, 46),

the arguments presumably have to be counted as correct since it seems hard to imagine that many competent speakers of English will hold that someone who has accepted the respective premises can reject the respective conclusion.[7]

That Peregrin and Svoboda want to include materially correct arguments amongst the correct arguments[8] is not the only reason why we cannot simply restrict "generally acceptable" to "logically acceptable" to exclude (7) and (8). Another reason is that to apply a notion of logical correctness we would need an account of logical form. If we follow Peregrin and Svoboda's explanation of formal and then logical correctness, we would have to come up with something like logical forms of these arguments and then show that these logical forms have incorrect instances.[9] However, we are just trying to determine a logical form for the arguments in question. Therefore, it seems not an admissible move to just claim that, for example, the logical form of (7) is actually

$\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$
⍰ $Ha \wedge Iba \rightarrow Ja \wedge Iba$

and that the logical form of (8) is actually

$\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$
⍰ $Ha \wedge \exists y Iya \rightarrow Ja \wedge \exists y Iya$

7 As mentioned in the preceding section, (TC) faces similar problems, as pointed out by Brun (2012, 327; 2014, 106).

8 More precisely, for Peregrin and Svoboda, correct arguments encompass logically correct, analytically correct and status quo correct arguments, where the latter are "correct due to some fixed and stable (though perhaps not eternal and unalterable) state of the world" (2017, 27).

9 That is at least what one would have to do according to the account offered in chap. 2.3 of (Peregrin and Svoboda 2017). Later, they hold that in the process of reflective equilibrium those arguments come out as logically correct whose "logical form is authorized as valid by logic" (2017, 113). In the present scenario, this would not change much since we would still face the question which logical form we are to assign to the arguments in question.

and that (apart from not being classically valid) these have clearly incorrect instances such as

> Every child of a mother is a child of a father.
> ⍰ If Martha is a mother and Rachel is a child of Martha, then Martha is a father and Rachel is a child of Martha.

and

> Every child of a mother is a child of a father.
> ⍰ If Martha is a mother that has a child, then Martha is a father that has a child.

respectively. A defender of (Φ3.1) or (Φ4.1) could rightly point out that that would just beg the question since we would simply choose the formalizations of (7) and (8) w.r.t. (Φ2.1) as the appropriate logical forms.

Moreover, a defender of (Φ3.1) or (Φ4.1) could even concede that we should formalize “analogous sentences analogously” (Brun 2012, 327) and that the incorrect instances we produced are to be formalized in line with the formalizations of (7) and (8) w.r.t. (Φ2.1): in the absence of a clear concept of “analogous sentence,” a defender of (Φ3.1) or (Φ4.1) can simply hold that the sentences in question are not analogous (see Lampert and Baumgartner 2010, 100–102).

Let us consider another argument, which is used by Lampert and Baumgartner (2010, 97–98) in their argument against Brun’s account of formalization:

> (9) Everything is a head of an animal.
> ⍰ Every head of a horse is a head of an animal.

If we assume that this argument is intuitively correct, we have an argument for whose unit set it holds that (Φ2.1) is adequate w.r.t. it, while (Φ3.1) and (Φ4.1) are not. However, Peregrin and Svoboda would probably not assume that speakers of English take this argument to be correct. They hold that “the paradoxes of material implication” lead to argument forms that “have instances that hardly any speaker of English would consider to be correct” (2017, 76). Given that the argument in question is basically a quantified version of one of the “paradoxes” (at least regarding its formalization w.r.t. Φ2.1), they would probably assume that not many speakers of English would judge it to be correct. Moreover, speakers (not already indoctrinated logically) might shy

doi: 10.48106/dial.v74.i2

away from considering it to be correct because the premise seems not simply false, but absurd.

We could of course consider more arguments, but presumably the problems already encountered would persist. In particular, (Φ2.1) cannot beat (Φ3.1) or (Φ4.1) on the completeness side if the premise position of the formalization of (CDM) is concerned. On the other hand, we have arguments such as (9) which concern the conclusion position of the formalization of (CDM) and for which (Φ2.1) beats (Φ3.1) and (Φ4.1) on the completeness side. However, such arguments will have premises that seem quite absurd. Concerning correctness, the trouble is that if the premises and the conclusions of arguments seem quite reasonable, it is unclear why competent speakers of English would hold that one can reject the conclusion if one accepts the premises.

In the next section, I will argue that we should not restrict our attention to premise-conclusion arguments but also consider how inferential relations between premises and conclusions can be accounted for inferentially by deriving conclusions from premises.

3 Adequacy and Inferential Sequences

Up to now we have only considered premise-conclusion arguments, such as

(10) Every head of a horse is a head of an animal.
Fury is a horse.
⍰ If Batu is a head of Fury, then Batu is a head of an animal.

The following is not simply a premise-conclusion argument:

(11) Assume every head of a horse is a head of an animal. Then it holds that if Batu is a head of a horse, then Batu is a head of an animal. Now assume Fury is a horse. Assume further that Batu is a head of Fury. Then Fury is a horse and Batu is a head of Fury. Thus, Batu is a head of a horse. Then Batu is a head of an animal. Thus, if Batu is a head of Fury, then Batu is a head of an animal.

Rather, (11) may be called an informal derivation. In it, the premises of (10) and an additional sentence are assumed. That last assumption is discharged in the last step, in which the conclusion of (10) is inferred, so that an informal derivation of the conclusion of (10) from the premises of (10) results.

At least from an inferential perspective, the derivation could be taken to show why the argument is logically correct by deriving its conclusion from its premises only using immediate inference steps that rely only on logico-syntactic features of the sentences involved. Such derivations can be formalized (more or less) “naturally” in natural deduction calculi such as Lemmon’s (1998):[10]

(12)

1	(1)	$\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$	Assumption (A)
1	(2)	$\exists y(Hy \wedge Iby) \rightarrow \exists y(Jy \wedge Iby)$	1 Universal quantifier elimination (UE)
3	(3)	Ha	A
4	(4)	Iba	A
3,4	(5)	$Ha \wedge Iba$	3, 4 ∧-introduction (∧ I)
3,4	(6)	$\exists y(Hy \wedge Iby)$	5 Existential quantifier introduction (EI)
1,3,4	(7)	$\exists y(Jy \wedge Iby)$	2, 6 Modus ponendo ponens (MPP)
1,3	(8)	$Iba \rightarrow \exists y(Jy \wedge Iby)$	4, 7 Conditional proof (CP)

Of course, one can view calculi as technical devices that can be used to prove that a certain formula follows from certain formulas, provided the calculi in question are correct w.r.t. the semantic consequence relation one chooses. However, one can also view logical calculi as an attempt to provide a systematic account of logical inferential relations in terms of syntactic features of formulas, and, via the “bridge” of formalization, of sentences in the scope of the logical system in question.[11] Such a view should appeal to Peregrin and Svoboda, who hold

> that language does not exist in the form of its set of sentences and a relation of inferability, but rather in the form of their generators: words and grammatical rules and basic (‘axiomatic’) instances

10 The leftmost column records the assumptions on which the formulas depend.

11 Such a view seems (at least partly) attributable to Jaśkowski and Gentzen, the founders of natural deduction, as regards their natural deduction calculi (see Jaśkowski 1934; Gentzen 1969b).

doi: 10.48106/dial.v74.i2

> of inference, plus rules of their composition. [...] the inferential competence, *viz.* the ability to tell correct inferences from incorrect ones, rests at the bottom on the knowledge of the elementary cases and in the knowledge of the ways of composition of simpler inferences into more complex ones. (2017, 159)

If taken as part of the systematic side in a reflective-equilibrium scenario, one can of course adjust the calculus, but a chosen calculus can radically constrain our commitments to the adequacy of formalizations if we also try to take the generation of logical inferability relations into account. So, for example, (Φ2.1), (Φ3.1), and (Φ4.1) all provide formalizations of the premises and the conclusion of (10) which render this argument classically correct. However, only with (Φ2.1) can we directly formalize (11) by (12).

My aim in the following is to make this idea more precise for natural deduction calculi with linear (and not tree) derivations.[12] Now we cannot speak anymore just of a logical system if a logical system is simply identified by a certain syntax and a consequence relation. Instead, we have to use something more fine-grained, namely a logical calculus w.r.t. which a given sequence of formulas is a derivation or not. I will call a sequence of formulas a *determined sequence of formulas w.r.t.* a calculus iff for every member in the sequence it is determined (for example by some form of commentary or by graphical means) if it is an assumption or an inference (in accordance with some rule). An example is the above derivation in Lemmon's system.

For the natural language side, I will speak of inferential sequences. An example is the introductory example of an informal derivation. Yet, to keep things simple, I will assume that *inferential sequences over* a set **S** of English sentences are finite non-empty sequences of expressions of the form

Assume S

and

Thus S'

12 Note that this choice is motivated by the relative ease with which informal derivations are formalized and formal derivations instantiated while the view on calculi sketched above can be applied to other types of calculi as well. Cordes and Reinmuth (2017) discuss the formalization of informal derivations in different types of linear calculi of natural deduction.

where S, S' are in **S**, with "Assume" indicating assumptions and "Thus" inferences.

I will assume that logical calculi are logical systems where an argument form is valid w.r.t. a calculus iff its conclusion can be derived from its premises in that calculus.[13] Under these assumptions, the criteria of the preceding section can be applied to logical calculi. For ease of exposition, I will restrict the following discussion to formalizations in Lemmon's system. However, they can easily be generalized or applied to other natural deduction calculi with linear derivations:

Adaptation of some terminology for Lemmon's calculus.

- Φ *is a formalization function* for* **S** iff Φ is a formalization function for **S** w.r.t. Lemmon's calculus.
- *I is an instance* of H w.r.t. the formalization function* Φ iff there is **S** such that Φ is a formalization function* for **S** and there are sentences $S_1, \ldots, S_n$ ($n \geq 1$) in **S** such that $I = \langle \ulcorner P_1 S_1 \urcorner, \ldots, \ulcorner P_n S_n \urcorner \rangle$ and H is a determined sequence of formulas of length n w.r.t. Lemmon's calculus such that for all $i \leq n$ it holds: $H_i = \Phi(S_i)$ and [[P_i = "Assume" and $\Phi(S_i)$ is assumed in line i of H] or [P_i = "Thus" and $\Phi(S_i)$ is inferred in line i of H]].
- *H is a formalization* of I w.r.t. the formalization function* Φ iff I is an instance* of H w.r.t. the formalization function Φ.

Note that I will continue to speak simply of instances, formalizations and formalization functions if I assume there is no danger of confusion. Note also that all formalization functions from the preceding section are also formalization functions w.r.t Lemmon's calculus.

First, let us consider the following inferential sequence over the domain of the formalization functions (Φ2.1), (Φ3.1), and (Φ4.1) from the previous section:

13 Of course, for the usual calculi for classical first-order logic it holds that an argument form is valid in this sense iff it is valid according to the model-theoretic definition of validity for classical first-order logic.

doi: 10.48106/dial.v74.i2

(13)

1. Assume every head of a horse is a head of an animal.
2. Assume Fury is a horse.
3. Assume Batu is a head of Fury.
4. Thus Batu is a head of an animal.
5. Thus if Batu is a head of Fury, then Batu is a head of an animal.

Obviously, this inferential sequence is a shortened version of (11). I take it that many of us would accept it as an informal derivation. However, w.r.t. the basic rules of most natural deduction calculi, its formalization would not be a derivation. While a calculus aims at covering some notion of derivability, it is intended to do so in a way which operates on the syntactic structure of formulas in a systematic way. Of course, we do not have to accept the way in which a given calculus does this. On the other hand, w.r.t. a given calculus, we have to make decisions as to which steps to count as immediate, as "the most distinctive patterns of the inferential landscape" (Peregrin and Svoboda 2017, 161). If we choose a certain formalization function, we also choose which inferential steps from natural language sentences to natural language sentences are instances* of derivations w.r.t. this formalization function and the given calculus and thus immediate in the sense that no intermediate steps are "missing."[14]

So, for example, if we choose one of the formalization functions (Φ2.1), (Φ3.1) and (Φ4.1), we also choose which of the following inferential sequences is an instance* of a derivation in Lemmon's system:

(14) An instance* of a derivation w.r.t. (Φ2.1)

1. Assume every head of a horse is a head of an animal.
2. Thus if Batu is a head of a horse, then Batu is a head of an animal.

(15) An instance* of a derivation w.r.t. (Φ3.1)

1. Assume every head of a horse is a head of an animal.
2. Thus it holds for everything: if it is a horse and Batu is a head of it, then it is an animal and Batu is a head of it.
3. Thus if Fury is a horse and Batu is a head of Fury, then Fury is an animal and Batu is a head of Fury.

---

14 The task of determining which inferences to count as immediate should be seen as integral to the project of formalization if we hold that one aim of formalization is to make "explicit the inferential properties of expressions of natural language" (Peregrin and Svoboda 2017, 109).

(16) An instance* of a derivation w.r.t. (Φ4.1)

1. Assume every head of a horse is a head of an animal.
2. Thus if Fury is a horse that has a head, then Fury is an animal that has a head.

Each of the three formalization functions renders only one of the three inferential sequences as an instance* of a derivation in Lemmon's system: (Φ2.1) does this for (14), (Φ3.1) for (15), and (Φ4.1) for (16). So, even if someone is inclined to judge each of the arguments (6), (7), and (8) from the preceding section to be intuitively correct, they ipso facto single out some inferential steps as (not) immediate if they choose one of the formalization functions.

One consideration that takes up the discussion from the preceding section is that if we want to proceed systematically and if we are interested in logical correctness, we may want to endorse inferences as immediate which seem acceptable in prima facie analogous cases. So, if we want to treat

Every child of a mother is a child of a father.

in the same way as

Every head of a horse is a head of an animal.

then only (14) seems an option w.r.t. the choice between (14), (15) and (16). Of course, as in the case of premise-conclusion arguments, such considerations are not decisive. However, they have another relevance in the new scenario. Choosing a formalization can be seen as choosing an account of how a sentence functions as a premise or conclusion. If we choose one account, we exclude others. Assume, for example, that we take (6), (7), (8) and (10) to be correct, while we have doubts about (9) and therefore choose, for example, (Φ4.1). Then we can be content in the setting of the preceding section because it renders the first four, but not the last argument correct. However, if we consider (11) an account of how and why (10) is logically correct, we cannot formalize this account if we choose (Φ4.1): if we choose (Φ4.1), we have to view (11) as an elliptical informal derivation in which certain steps are left out. Thus, if we take

doi: 10.48106/dial.v74.i2

(17)

1. Assume every child of a mother is a child of a father.
2. Thus if Rachel is a child of a mother, then Rachel is a child of a father.
3. Assume Martha is a mother.
4. Assume Rachel is a child of Martha.
5. Thus Martha is a mother and Rachel is a child of Martha.
6. Thus Rachel is a child of a mother.
7. Thus Rachel is a child of a father.
8. Thus if Rachel is a child of Martha, then Rachel is a child of a father.

to provide an account of the correctness of

Every child of a mother is a child of a father.
Martha is a mother.
⍰ If Rachel is a child of Martha, then Rachel is a child of a father.

we would also have to explain why we cannot replace "child" by "head," "mother" by "horse," "father" by "animal," "Martha" by "Fury" and "Rachel" by "Batu" to get an account of the correctness of (10).

To illustrate the need to make choices, we can consider another example. Suppose we take the inferential sequence

(18)

1. Assume every horse that has a head is an animal that has that head.
2. Thus it holds for everything: if it is a horse and Batu is a head of it, then it is an animal and Batu is a head of it.
3. Thus if Fury is a horse and Batu is a head of Fury, then Fury is an animal and Batu is a head of Fury.

to provide an account of the logical correctness of

(19) Every horse that has a head is an animal that has that head.
⍰ If Fury is a horse and Batu is a head of Fury, then Fury is an animal and Batu is a head of Fury.

and the inferential sequence

(20)

1. Assume every horse that has a head is an animal that has a head.
2. Thus if Fury is a horse that has a head, then Fury is an animal that has a head.

to account for the logical correctness of

(21) Every horse that has a head is an animal that has a head.
□ If Fury is a horse that has a head, then Fury is an animal that has a head.

Then, we can choose (Φ2.1) but neither (Φ3.1) nor (Φ4.1) as each of the latter formalization functions only offers a formalization* of one of the two informal derivations, namely (Φ3.1) of (18) and (Φ4.1) of (20).

One obvious option to make the costs in choosing one or another formalization function explicit is to reformulate the criteria of correctness, completeness, and adequacy from the preceding section directly for inferential sequences and determined sequences of formulas. Then, for example, (Φ2.1), but neither (Φ3.1) nor (Φ4.1) would be adequate w.r.t. the set {(18), (20)}. Moreover, this would also allow us to treat non-trivially equivalent formalization functions differently since they would be adequate for different sets of inferential sequences. For example, (Φ3.1) but not (Φ4.1) would be adequate w.r.t. {(18)} and (Φ4.1) but not (Φ3.1) would be adequate for {(20)} if we judge (18) and (20) to be informal derivations.

For reasons of space, I will not make this explicit, but propose an inferential version of (PHS) that takes into account the discussion so far and puts systematic inferential constraints on adequacy judgements concerning formalization functions. Still, some preparatory work is required. We can set (remember that the whole discussion is carried out for Lemmon's system):

DERARG. *H is a derivation for A w.r.t. the formalization function** Φ if and only if

i) there is **S** such that Φ is a formalization function* for **S**, and A is an argument over **S**; and
ii) H is a derivation in Lemmon's calculus such that
   a. $\{\varphi \mid \varphi$ is an undischarged assumption in $H\} = \{\varphi \mid \varphi$ is a premise in the formalization of A w.r.t. $\Phi\}$; and
   b. the conclusion of $H =$ the conclusion of the formalization of A w.r.t. Φ; and
   c. every non-logical symbol that occurs in H also occurs in the formalization of A w.r.t. Φ.

According to this definition, (12) is a derivation for (10) w.r.t. (Φ2.1). Of course, there are also derivations for (10) w.r.t. (Φ3.1) and (Φ4.1). However, these will differ from (12) and will not be formalizations of (a standardized version of) (11). Similarly, while there are derivations for (2) w.r.t. (Φ3) and (Φ4), these will differ considerably from

(22)

1	(1)	$\forall x(Fx \rightarrow Gx)$	A
1	(2)	$Fb \rightarrow Gb$	1 UE
3	(3)	Fb	A
1,3	(4)	Gb	2, 3 MPP

which is a derivation for (2) w.r.t. (Φ1). In contrast to this,

(23)

1	(1)	$\forall x(\exists y(Hy \land Ixy) \rightarrow \exists y(Jy \land Ixy))$	A
1	(2)	$\exists y(Hy \land Iby) \rightarrow \exists y(Jy \land Iby)$	1 UE
3	(3)	$\exists y(Hy \land Iby)$	A
1,3	(4)	$\exists y(Jy \land Iby)$	2, 3 MPP

which is a derivation for (2) w.r.t (Φ2), corresponds closely to (22) and to the informal

1. Assume every head of a horse is a head of an animal
2. Thus if Batu is a head of a horse, then Batu is a head of an animal
3. Assume Batu is a head of a horse
4. Thus Batu is a head of an animal

To make the notion of correspondence more precise, we set:

CORDER. *H is a derivation that corresponds to* H^* *w.r.t.* Φ^*, Φ *and* $\mathbf{S^o}$ if and only if

i) there is $\mathbf{S}$ such that Φ is a formalization function* for $\mathbf{S}$, and $\mathbf{S^o} \subseteq \mathbf{S}$; and

ii) there is **S*** such that Φ^* is a formalization function* for **S***, and $\mathbf{S^o} \subseteq \mathbf{S}^*$; and

iii) H and H^* are derivations in Lemmon's calculus and there is an n such that

a. the length of $H = n =$ the length of H^*, and

b. for every $i \leq n$ it holds:

i. if H^*_i is an assumption in line i of H^*, then H_i is an assumption in line i of H, and

ii. if H^*_i is inferred in line i of H^*, then H_i is inferred in line i of H, and

iii. for every R: if R is an inference rule of Lemmon's calculus and H^*_i can be inferred in line i of H^* in accordance with R, then H_i can be inferred in line i of H in accordance with R, and

iv. for every S in $\mathbf{S^o}$: if $H^*_i = \Phi^*(S)$, then $H_i = \Phi(S)$.

According to this definition, (23) is a derivation that corresponds to (22) w.r.t (Φ1) and (Φ2) and their common domain. Apart from the obvious correspondence on the formal side, it holds that those formulas in a certain line that are values for a sentence from the common domain of the two formalization functions are the respective values of the same sentence. Note that (23) also corresponds to (22) w.r.t. (Φ1) and (Φ2.1) and the domain of (Φ1).

On the other hand, there can be no derivation H^o in Lemmon's system such that H^o corresponds to (22) w.r.t. (Φ1) and (Φ3) or (Φ4). For example, if we tried to find a corresponding derivation for (Φ3), we would come to:

(24)

1	(1)	$\forall x \forall y(Hy \wedge Ixy \rightarrow Jy \wedge Ixy)$	A
1	(2)	?	1 UE
3	(3)	$\exists y(Hy \wedge Iby)$	A
1,3	(4)	$\exists y(Jy \wedge Iby)$	2, 3 MPP

Obviously, whatever formula we choose to infer by UE in line (2) will itself have a universal quantifier as main operator and thus be an unfit premise for the MPP in the last line. Similarly, for (Φ4), we would arrive at:

doi: 10.48106/dial.v74.i2

(24)

1	(1)	$\forall x(Hx \wedge \exists yIyx \rightarrow Jx \wedge \exists yIyx)$	A
1	(2)	?	1 UE
3	(3)	$\exists y(Hy \wedge Iby)$	A
1,3	(4)	$\exists y(Jy \wedge Iby)$	2, 3 MPP

In this case, whatever formula we choose to infer by UE in line (2) will have an antecedent that differs from the formula in line (3) and a consequent that differs from the formula in line (4) and thus again be an unfit premise for the MPP in the last line. Of course, these results for (Φ3) and (Φ4) carry over to their extensions (Φ3.1) and (Φ4.1).

Considerations concerning corresponding derivations w.r.t. different formalization functions and a subset of their domains could be used to take inferential sequences into account whose formalizations are not derivations, and which may be viewed as elliptical informal derivations. However, I will leave this for another occasion and just put forward an inferential version of (PHS) for formalization functions:[15]

> PHS-INF. If Φ is a formalization function* for **S**, and Φ* is a formalization function* for **S***, and **A** is a non-empty set of arguments over **S** ∩ **S***, then:
>
> i) Φ is not an adequate formalization function* for **S** w.r.t. **A**; or
> ii) Φ* is not an adequate formalization function* for **S*** w.r.t. **A**; or
> iii) for every *A*, every *H**: if *A* is in **A** and *H** is a derivation for *A* w.r.t. Φ*, then there is an *H* such that *H* is a derivation that corresponds to *H** w.r.t. Φ*, Φ, **S** ∩ **S***; or
> iv) for every *A*, every *H*: if *A* is in **A** and *H* is a derivation for *A* w.r.t. Φ, then there is an *H** such that *H** is a derivation that corresponds to *H* w.r.t. Φ, Φ*, **S** ∩ **S***.

This criterion extends the "unity of logical form" (Baumgartner and Lampert 2008, 95) which (PHS) is meant to ensure to the role of formalizations in derivations and thereby strongly constrains judgements of adequacy. So, for example, if we judge (Φ1) to be adequate w.r.t. {(2)}, we cannot judge (Φ3),

15 (PHS-INF) follows the (HCS)-formulation of (PHS), for which see footnote 6.

(Φ4) or any extension of either to be adequate w.r.t. any set **A** of arguments over their respective domains if $\{(2)\} \subseteq \mathbf{A}$. On the one hand, (22) is a derivation for (2) w.r.t. (Φ1) and, as shown above, there are no derivations that correspond to (22) w.r.t. (Φ1) and (Φ3) or (Φ4) and their common domain, a result which carries over to extensions of (Φ3) and (Φ4). On the other hand, we have derivations for (2) w.r.t. (Φ3) and its extensions for which there are no corresponding derivations w.r.t. (Φ3), (Φ1) and the domain of (Φ1), and the same holds for (Φ4) and its extensions. Thus, if (Φ1) is an adequate formalization function* for its domain w.r.t. {(2)}, then (Φ3), (Φ4) as well as their extensions are not.

Also, this inferential version of Brun's (PHS) "punishes" the trivialization of equivalence, because non-trivially equivalent formulas behave differently in the context of derivations. So, for example, we can show that either (Φ3.1) or (Φ4.1) is not an adequate formalization function* for their common domain w.r.t. {(19), (21)} if we accept (PHS-Inf): on the one hand, the formalization of (18) w.r.t. (Φ3.1) is a derivation for (19) w.r.t. (Φ3.1) to which no derivation corresponds w.r.t. (Φ3.1), (Φ4.1) and their common domain. On the other hand, the formalization of (20) w.r.t. (Φ4.1) is a derivation for (21) w.r.t. (Φ4.1) to which no derivation corresponds w.r.t. (Φ4.1), (Φ3.1) and their common domain. Thus, according to (PHS-Inf), at least one of the two formalization functions cannot be adequate. These results also show that (PHS-Inf) cannot be used consistently with the criteria from the preceding section. Rather, these criteria have to be adapted, e.g. by taking into account the derivations for the arguments in the respective sample sets.

If we put the discussion so far in the context of providing a systematic inferential account of logical correctness (e.g. in the context of reaching some form of reflective equilibrium), we should (or, at least, can) treat a derivation for an argument w.r.t. a formalization function as a way of accounting for the logical correctness of that argument. Choosing among formalization functions against the background of a calculus is thus a way of choosing between different ways of accounting for the supposed logical correctness of natural language arguments. Doing this, we have to make decisions and (probably) revise initial judgements. If we want a systematic account of logical correctness in terms of inferential role, where the inferential role of a sentence is tightly connected to the logical form we assign to it, then we have to make some choices.

These choices will also determine which arguments are to be counted as logically correct w.r.t. a certain logical system. If we choose a certain formalization function to be adequate w.r.t. a set **A** of arguments, we also choose which arguments in **A** come out as logically correct. Even if one does not hold that "[w]hatever is informally valid must be shown to be valid on formal grounds by means of a logical formalization involving conceptual analysis" (Baumgartner and Lampert 2008, 105), but tries to formalize intuitively logically correct arguments as logically correct w.r.t. the chosen logical system, one encounters the problem that without a notion of logical correctness, and, in turn, of logical form, one faces just a plentitude of intuitively (in)correct arguments, as described in the preceding section. But if we see the choice between formalization functions also as a choice as how to account for the logical correctness of arguments, the scenario changes, as choosing a formalization function that covers more intuitively correct arguments than another may well mean choosing a formalization function that does not cover inferential steps we take as immediate, and thus accounts of logical correctness that we want to accept.

So, for example, compared with the scenario at the end of the preceding section, (Φ3.1) and (Φ4.1) do not seem to fare better than (Φ2.1): they also offer an account of the logical correctness of arguments in which (CDM) appears as a premise. However, they cannot offer the account that (Φ2.1) provides. Moreover, if we accept (PHS-INF), we can show that either (Φ3.1) or (Φ4.1) is not adequate w.r.t. relevant sets of arguments over their domain. Thus, choosing them over (Φ2.1) does not just leave out some intuitively maybe rather dubious arguments. Concerning such arguments, we can see their logical correctness as a by-product of the process of systematization. As Brun stresses, what we want is "a *system*, not merely a list of our commitments" (2014, 113), which forces us, as Peregrin and Svoboda put it, to "impose more order on our language and our reasoning than we are able to *find* there, even at the cost of some Procrustean trimming and stretching" (2017, 102).

4 Directions for Future Research

The approach suggested in the preceding section does not aim at a merely "technical" solution to the problems encountered by inferential criteria in a premise-conclusion setting. Rather, it rests on taking seriously a notion of logical correctness in terms of inferability of the conclusion from the premises in accordance with a finite set of rules for certain expressions and ways of combining them. It would be interesting to investigate to what extent this approach allows us to use the syntax of a logical system to structure and classify natural language sentences relative to that logical system. This should include an investigation into which predicates from the metalanguage for the (syntax of) the logical system can fruitfully be adapted to describe syntactical features of the sentences in the intended scope of the logical system. For example, one could try to account for the sub-sentences of a sentence by recourse to the subformula relation for formulas.

While "an approach to logic [that] is closely allied to inferentialism in the philosophy of language and to theories underlying the so-called proof-theoretic semantics in logic" (Peregrin and Svoboda 2017, 4) should be more than compatible with taking not only premise-conclusion arguments but also inferential sequences into account when trying to determine the inferential roles of natural language sentences, it seems unclear to which extent the inferential criteria developed here fit into other approaches. This applies in particular to Baumgartner and Lampert's "*new picture* of adequate formalization" (2008, 95), according to which "the difference between informal formal and informal material validity must be dropped" (2008, 105). As the strengthened inferential criteria provide incentives to draw the line between materially and logically correct arguments more sharply, it would be interesting to assess them in the context of the debate surrounding Baumgartner and Lampert's "new picture" (see Baumgartner and Lampert 2008; Lampert and Baumgartner 2010; Brun 2012; Peregrin and Svoboda 2013).

The strengthened inferential criteria force us to make fine-grained choices and, in particular, to choose between equivalent formalizations of the same sentence. However, with this might also come the worry that the choices forced on us are too fine-grained. For example, w.r.t. the basic rules of most natural deduction calculi we have to choose between

1. Assume Fury is a horse and Fury is an animal and Batu is a head of Fury
2. Thus Fury is a horse

and

1. Assume Fury is a horse and Fury is an animal and Batu is a head of Fury
2. Thus Batu is a head of Fury

and, if we keep the order of the conjuncts, cannot choose

doi: 10.48106/dial.v74.i2

1. Assume Fury is a horse and Fury is an animal and Batu is a head of Fury
2. Thus Fury is an animal

Intuitively, all three inferences seem immediate and at least the being forced to choose between the first two seems rather strange. One option is to take this simply as the prize of a systematic account of such inferences in terms of the introduction and elimination rules for conjunction. The decision we have to make is arbitrary and comes at the price of excluding some intuitively immediate inferences, but it is, according to this option, a price we have to pay.

An alternative option is to liberalize the rules of the calculus to allow, for example, a direct formalization of the three inferential sequences as formal derivations. One direction of future research is a weighing of these options. A related line of inquiry is how one could use the notion of corresponding derivations to take also elliptical informal derivations into account. Last, but not least, one should investigate how the strengthened inferential criteria can be put to work when we deal with argumentative texts which are not readily formalizable but have first to be subjected to some form of argument analysis which may involve hermeneutical considerations (see e.g. Brun 2014; Brun and Betz 2016; Reinmuth 2014).*

Friedrich Reinmuth
University of Greifswald
reinmuthf@uni-greifswald.de

References

BAUMGARTNER, Michael and LAMPERT, Timm. 2008. "Adequate Formalization." *Synthese* 164(1): 93–115, doi:10.1007/s11229-007-9218-1.

BRUN, Georg. 2003. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. Logos n. 2. Heusenstamm b. Frankfurt: Ontos Verlag. Second edition: Brun (2004), doi:10.1515/9783110323528.

—. 2004. *Die richtige Formel: Philosophische Probleme der logischen Formalisierung*. 2nd ed. Heusenstamm b. Frankfurt: Ontos Verlag. First edition: Brun (2003).

* I would like to thank Geo Siegwart and the members of the Greifswald advanced seminar on theoretical philosophy for their helpful comments. Particular thanks are due to Moritz Cordes, Karl Christoph Reinmuth and two anonymous referees for *Dialectica* for helpful suggestions and discussion.

—. 2012. "Adequate Formalization and de Morgan's Argument." *Grazer Philosophische Studien* 85: 325–335, doi:10.1163/9789401208338_017.

—. 2014. "Reconstructing Arguments: Formalization and Reflective Equilibrium." in *Theory and Practice of Logical Reconstruction: Anselm as a Model Case*, edited by Friedrich Reinmuth, Geo Siegwart, and Christian Tapp, pp. 94–129. Logical Analysis and History of Philosophy n. 17. Münster: Mentis Verlag, doi:10.30965/9783957439468_008.

Brun, Georg and Betz, Gregor. 2016. "Analysing Practical Argumentation." in *The Argumentative Turn in Policy Analysis. Reasoning about Uncertainty*, edited by Sven Ove Hansson and Gertrude Hirsch Hadorn, pp. 39–78. Logic, Argumentation & Reasoning n. 10. Cham: Springer, doi:10.1007/978-3-319-30549-3_3.

Cordes, Moritz and Reinmuth, Friedrich. 2017. "Commentary and Illocutionary Expressions in Linear Calculi of Natural Deduction." *Logic and Logical Philosophy* 26(2): 163–196, doi:10.12775/llp.2017.002.

Gentzen, Gerhard. 1935. "Untersuchungen über das logische Schließen." *Mathematische Zeitschrift* 39: 176–210, 405–431. Republished as Gentzen (1969a), doi:10.1007/BF01201353.

—. 1969a. *Untersuchungen über das logische Schliessen*. Darmstadt: Wissenschaftliche Buchgesellschaft.

—. 1969b. "Investigations into Logical Deduction." in *The Collected Papers of Gerhard Gentzen*, pp. 68–131. Studies in Logic and the Foundations of Mathematics n. 55. Amsterdam: North-Holland Publishing Co. Translation of Gentzen (1935) by M.E. Szabo.

Jaśkowski, Stanisław. 1934. "On the Rules of Suppositions in Formal Logic." *Studia Logica (wydawnictwo poświęcone logice i jej historji)* 1: 5–32. Reprinted in McCall (1967, 232–258), with considerable change in notation, https://www.logik.ch/daten/jaskowski.pdf.

Lampert, Timm and Baumgartner, Michael. 2010. "The Problem of Validity Proofs." *Grazer Philosophische Studien* 80: 79–109, doi:10.1163/18756735-90000872.

Lemmon, Edward John. 1965. *Beginning Logic*. London: Thomas Nelson and Sons Ltd.

—. 1998. *Beginning Logic*. Boca Raton, Florida: CRC Press. Republication of Lemmon (1965).

McCall, Storrs, ed. 1967. *Polish Logic 1920–1939*. Oxford: Oxford University Press.

Peregrin, Jaroslav and Svoboda, Vladimír. 2013. "Criteria for Logical Formalization." *Synthese* 190(14): 2897–2924, doi:10.1007/s11229-012-0104-0.

—. 2017. *Reflective Equilibrium and the Principles of Logical Analysis: Understanding the Laws of Logic*. London: Routledge, doi:10.4324/9781315453934.

Reinmuth, Friedrich. 2014. "Hermeneutics, Logic and Reconstruction." in *Theory and Practice of Logical Reconstruction: Anselm as a Model Case*, edited by Friedrich Reinmuth, Geo Siegwart, and Christian Tapp, pp. 152–190. Logical Analysis and History of Philosophy n. 17. Münster: Mentis Verlag, doi:10.30965/9783957439468_010.

Considerations on Logical Consequence and Natural Language

Gil Sagi

In a recent article, "Logical Consequence and Natural Language," Michael Glanzberg (2015) claims that there is no relation of logical consequence in natural language. The present paper counters that claim. I shall discuss Glanzberg's arguments and show why they don't hold. I further show how Glanzberg's claims may be used to rather support the existence of logical consequence in natural language.

Contemporary logic is studied using the tools of formal languages that have been developed during the past two centuries. Logicians often approach natural language with some apprehension: natural language is complex and messy, studied fragment by fragment by a variety of methods that hardly seem to provide any sense of unity. This is by contrast to formal languages, that are neat, manageable and simple (to the extent that the logician devises them to be). Is there even a logic in natural language? If we are to move beyond first impressions, we should make precise what we mean by this question, and specifically, what we mean by "logical consequence," "natural language" and logical consequence being "in" natural language.

In a recent paper, "Logical Consequence and Natural Language," Michael Glanzberg (2015) confronts this issue head-on. While the literature is not short of remarks on the question of the relation between logic and natural language, Glanzberg's important contribution is a paper-long discussion of what may be meant by the question and an extensively argued response. It is therefore worthwhile to consider the details of Glanzberg's arguments, and thus further the discussion on this fundamental topic. This contribution is thus dedicated to discussing Glanzberg's stance, and to criticising the arguments he puts forward. Now, if we are to present a critique of Glanzberg's argumentation, it would be most fruitful to do so on Glanzberg's terms: on his understanding of the question of logic in natural language. However, we shall be critical not only of his response to the question at hand but also of the particular constraints that he imposes which lead him to his response. Taking up some basic assumptions from Glanzberg, we are led to very different conclusions than his. Before I delve into Glanzberg's reasoning, let me start with a broader introduction to help us orient ourselves in the discussion.

Here, together with Glanzberg, we shall treat natural language as a natural phenomenon—as the object of study of empirical linguistics. Logical consequence will be taken to be a relation between sets of sentences (constituting premises) and sentences (serving as conclusions) in the relevant language. This relation holds if the conclusion necessarily follows from the premises by virtue of the form of the sentences. We shall elaborate on this condition later on, but for now let us note that formal systems studied by logicians can be taken to be displaying, or modelling logical consequence in natural language. Our understanding of what this relation might be will be tied to the options exhibited by formal systems. That the relation of logical consequence is *in* natural language will be explained to mean that the appropriate formal systems for logic serve as good models for a phenomenon in natural language.

The formal systems we shall refer to are the products of a tradition starting with Frege's *Begriffsschrift*, which set as a primary aim to provide a methodology for the sciences. At the base of this tradition we have first and second order predicate logic—and as the aims varied and developed through the twentieth century, so did the formal systems that were used. Examples of other partakers in this traditional project, who upheld the same primary aim, are Tarski, Carnap and Quine. The virtues they sought in logical systems had to do with their uses in scientific reasoning—whether in deductive sciences (Tarski's primary target) or beyond (as we can see in Carnap and Quine).

Formal systems have as their first and foremost virtues rigour and mathematical precision. Further virtues, which can be attributed to the basic systems (first order logic and possibly some of its extensions), would include simplicity and restrictiveness. If, for example, we consider Frege's foundational project, we see that the epistemological motivations of placing arithmetic on a secure ground lead invariably to a restrictive stance towards logic.[1] Other members of the traditional project held a similar attitude, each in their own way.[2]

The formal systems devised by Frege and his successors have found their way to a variety of applications and uses, where different emphases called for different virtues. Relevant to our discussion is the *linguistic project*, which

1 Frege helped himself to second order logic, which, following Glanzberg, will be considered as restrictive for the purpose of this paper.

2 I discuss the traditional project in length in Sagi (2020, 2021).

we can see developing from the midst of the twentieth century onwards, where formal systems of logic are used in the study of natural language (as in Chomsky 1957; Davidson 1967, 1970; Davidson and Harman 1972; Montague 1974).

The traditional project has distinctive normative aspects, at least insofar as it is methodological. The linguistic project, by contrast, is wholly descriptive. Natural language, disregarded by members of the traditional project as inadequate for scientific research, here becomes the main focus. Natural language, as the subject matter of linguistic theory, is treated like any other natural phenomenon. The formal systems devised in the traditional project become useful tools for the formal study of natural language syntax and semantics. Rather than a medium for formulating scientific theories, now the formal systems become mathematical models for the study of natural language.

It is very clear, however, that the restrictive systems of the traditional project are much too coarse, and are inadequate in capturing a wide array of natural language phenomena. First or second order predicate logic may be suitable for foundational purposes, but it is hardly a good fit for linguistic study. This mismatch is where the suspicion arises that logic and natural language lie on very different grounds. Glanzberg's arguments are essentially based on the observation that standard predicate logic fails to be a good fit for the study of natural language, and he therefore concludes that natural language, on certain assumptions, does not have a genuine consequence relation.

Before moving on, I'd like to pause on the relation between the formal systems provided by a linguistic theory and the phenomenon which is the subject matter of investigation. Cook (2002) gives us a way of assessing this relation. Cook (2002, 234) presents us with three rough options. We can take the formal system to be a description of natural language and its logical properties: on this view, every aspect of the formalism corresponds (at least roughly) to a feature of the phenomenon being formalized. On the other end of the spectrum, we can view the formalism as completely instrumental: it might help us in predictions on the phenomenon at hand, but the details of formalism provide us with no insight or explanation of the inner-workings of the phenomenon. These two options lay a spectrum of possible views, where somewhere in the middle we can find the view of logic-as-modelling. In this view (see also Shapiro 1998), the formalism serves as a mathematical model of the phenomenon at hand. Some aspects or elements of the model correspond to features of the phenomenon (these are *representors* in Shapiro's terminology), and others (*artefacts*, in Shapiro's terminology) do not: they

doi: 10.48106/dial.v74.i2

help keep the model simple and easy to handle. It seems that the extremes of the spectrum are either impractical or unhelpful, and that a reasonable approach would be to aim for some place in the middle.

In the present context, when we ask whether there is a logical consequence relation in natural language, one way to approach the issue would be to see whether formal systems that satisfy basic conditions we would expect from systems for logic are good models for some phenomenon in natural language. I shall claim that Glanzberg himself provides the basis for the position that formal systems of logic are indeed models of natural language phenomena.

The plan of the paper is as follows. In section 1, I present the thesis of logic in natural language as understood through Glanzberg's terms, and I articulate the basic assumptions and observations that are essential for Glanzberg's reasoning. Glanzberg presents three arguments against the thesis of logic in natural language. I review and counter these arguments, each in turn, in section 2–section 4. Besides the negative arguments, Glanzberg also presents a positive proposal of how a logical consequence relation can be obtained by modifying natural language. In section 5, I shall argue that the process described by Glanzberg is that of modelling, and it thus serves to rather substantiate the thesis that there is a logic in natural language.

1 Making Sense of the Question: Glanzberg's Analysis

Glanzberg argues that natural language does not have a logical consequence relation. More specifically, he argues that when logic is understood in the appropriate restrictive way, the following thesis is false:

> *The logic in natural language thesis*: a natural language, as a structure with a syntax and a semantics, thereby determines a logical consequence relation. (2015, 75)

Glanzberg explains that *logic* can be understood either restrictively or permissively. The more restrictive the logic, the less inferences it accepts as valid. Basically, standard, classical first or second logic are of the restrictive sort by Glanzberg's lights, and the variety of "non-standard" and "non-classical" logics include the permissive sort (2015, 78). According to Glanzberg, the arguments he presents show that natural language does not determine a restrictive logical consequence relation, and strongly suggest that it also does not determine a permissive logical consequence relation.

We shall deal with Glanzberg's arguments in the following sections. First, however, let us lay out the claims that serve as the basis for Glanzberg's arguments.

First, we note that Glanzberg analyses logical consequence as a *necessary* and *formal* relation (2015, 76). It is necessary in the sense that a valid argument is an argument where truth is preserved from premises to conclusion over all relevant possibilities. It is formal in the sense that it holds by virtue of the forms of the sentences involved. There is, of course, much more to say, but this should suffice at present.

Now, Glanzberg (2015, 79) crucially assumes a model-theoretic account of logical consequence, and that such an account is most likely to lead to a logical consequence relation in natural language. I do not object to this assumption, but it would be helpful to see what it is based on. Glanzberg's model-theoretic approach builds on three observations. The first one is that post-Tarskian model-theoretic consequence is necessary and formal as required (Glanzberg 2015, 77). Secondly, model-theoretic consequence appears to be a good explication of logical consequence—understood as necessary and formal (notwithstanding well-known criticisms like Etchemendy 1990).

The third observation which bases the model-theoretic approach is that in the study of natural language, we find a family of related notions, among which are implications and entailments. According to Glanzberg, *implication* is a wide notion, covering relations that are either logical or of looser connections, including those based on defeasible reasoning. Within the category of implications, we have the narrow notion of logical consequence, that which aligns with the restrictive view of logic (see Glanzberg (2015, 80); apparently even though logical consequence is a subspecies of implication, it is not really a relation in natural language—more on this in what follows). And included in implications we have *entailment*, which is understood as a truth-conditional connection: p entails q if the truth conditions of p are included in the truth conditions of q (Glanzberg 2015, 80). Entailments include analytic connections, such as "Max is a bachelor, therefore Max is unmarried," and they may include also "metaphysical" connections, such as "x is water, therefore x is H_2O." That is if truth conditions are metaphysically possible worlds, and one accepts the Kripke-Putnam views of natural kind terms (Glanzberg 2015, 80).

In sum, we have on the one hand model-theoretic consequence, which fits the analysis of the notion of logical consequence. On the other hand, we have relations in natural language that come structurally close to, and even include as a subset the relation of logical consequence thus conceived.

doi: 10.48106/dial.v74.i2

Glanzberg's argumentation from this point onwards serves to draw a divide between model-theoretic consequence and the broader relations we find in natural language.

Another crucial assumption made by Glanzberg is that the way to determine whether there is a relation of logical consequence in natural language is through looking at current practices in linguistics, and more specifically, those of contemporary natural language semantics. To a certain extent, I find this assumption justified: linguistics is the science that studies natural language. If the state of the art in linguistics either enforces or undermines the existence of a certain phenomenon in natural language, we should certainly take that into primary consideration. Glanzberg, however, seems to draw more from contemporary semantic theory, and we shall review this issue in due course.

Glanzberg presents three arguments to support his conclusion: the first leans on the assumptions we spelled out above, and the other two have additional assumptions which will be brought up in our further discussion. In the following sections, I shall give an outline of the arguments and present my criticism. The outcome will be that Glanzberg's arguments are not as strong as they aim to be, and do not give sufficient basis to refute the logic in natural language thesis.

2 The Argument From Absolute Semantics

The first and main argument Glanzberg puts forward is the *argument from absolute semantics*. It is the most general of the three arguments, and it concerns the use of model theory in natural language semantics. The gist of the argument is that natural language semantics is *absolute*, and in fact does not use the range of models that model theory offers.

One of the basic ideas, adopted from Lepore, is that model theory defines only *relative* truth conditions. It gives us the notion of truth in a model. It says, for instance, whether the sentence "Snow is white" is true in some model. Semantic theory, if apt, should give conditions of truth *simpliciter*, i.e. tell us when "Snow is white" is true. Davidsonian *absolute* statements of truth conditions tell you that the sentence "Snow is white" is true if and only if snow is white, which, according to Glanzberg, is what we wanted.

Glanzberg claims that even semantic theories that use model theory, stemming from the Montagovian tradition, are, at bottom, providing absolute semantics. Glanzberg writes:

> What is characteristic of most work in the model-theoretic tradition is the assignment of semantic values to all constituents of a sentence, usually by relying on an apparatus of types (cf. Chierchia and McConnell-Ginet 1990; Heim and Kratzer 1998). Thus, we find in model-theoretic semantics clauses such as:[3]
>
> (1) a. $[\![\text{Ann}]\!] = \text{Ann}$
> b. $[\![\text{smokes}]\!] = \lambda x \in D_e.x \text{ smokes}$
>
> […] [These clauses] provide absolute statements of facts about truth and reference […] We see that the value of "Ann" is Ann, not relative to any model. (2015, 89)

Semantics of natural language, according to Glanzberg, is the study of speakers' linguistic competence, and more specifically, of knowledge of meaning. Arguably, truth conditions are what a speaker knows when they understand a sentence. The relevant study must then be directed at the absolute values presented in the clauses above. By contrast, Glanzberg explains, in order to understand the logical properties of a sentence, we look at the values of the sentence across a range of models. But since semantics of natural language is absolute, it is blind to what happens across any non-trivial range of models (2015, 91). To sum: whether natural language has a logical consequence relation will be determined by whether current semantic theory appeals to a non-trivial range of models in explaining speakers' competence. Since it doesn't, natural language, according to the argument from absolute semantics, does not have a logical consequence relation. Later on in the article, Glanzberg concedes that a range of models is explicitly appealed to in the study of determiners, but, he explains, at this point semantic theory goes beyond its proper terrain. We shall reach this point in due course.

Is natural language semantics really absolute? Here are some considerations to the contrary. Note that while the semantic value of "smokes" is a function which determines for every object in the specified domain whether it smokes, semantics does not tell us what this function is—what its values are. Indeed, all that semantics gives us is the *condition* for obtaining the value 1 from this function. And so, all we have, in extensional semantics, are truth conditions

3 Glanzberg explains: "In common notation, $[\![\alpha]\!]$ is the semantic value of α. I write $\lambda x \in D_e.\phi(x)$ for the function from the domain D_e of individuals to the domain of values of sentences (usually truth values)" (2015, 89).

doi: 10.48106/dial.v74.i2

of a sentence such as "Ann smokes" rather than an absolute truth value. Heim and Kratzer explain that the semanticist cannot, and also should not, provide the function in extension: "We do not know of every existing individual whether or not (s)he smokes. And that is certainly not what we have to know in order to know the meaning of 'smoke' " (1998, 21). Reference is not what a speaker knows. While the meaning of an expression determines a reference, what the speaker knows does not pick out the reference. This indeterminacy makes room for a range models.

Thus, despite the form of the clauses above, when we look at the practice of natural language semantics, we do find a range of models. In Zimmermann (1999) it is claimed that a range of models is a part of natural language semantics, and that it reflects linguists' ignorance. Linguists can't point out the extension of every expression in natural language. If they could, it would be determined by natural language semantics whether there are white ravens or whether Ann smokes, merely by giving the extensions of "white," "ravens," "Ann" and "smokes." If we are interested only in one model, then the relation between extensions is completely determined.[4] Now one might insist that natural language semantics does require an absolute semantics, and that the range of models is a byproduct of less than ideal theorising, not indicative of any real phenomenon in natural language. But note that the ignorance of linguists is not (at least not always) expected to be overcome, as we see from the quote of Heim and Kratzer. It is not part of linguistic competence whether Ann smokes—or on which possible worlds Ann smokes. It is not only the linguist's ignorance that a range of models may signify, but also that of competent speakers themselves.

Indeed, another recent article by Glanzberg suggests that the explanatory power of semantic theory is limited where absolute items such as (3a-b) are involved, and that such clauses contain pointers to other cognitive faculties. "[S]emantics, narrowly construed as part of our linguistic competence, is only a partial determinant of content" (2014, 259). We need further conceptual resources to fully determine the extension of every expression in a language.

Now, while I take the above considerations to undercut the absoluteness of natural language semantics, I submit that the argument from absolute semantics fails even if we accept that natural language semantics is absolute.

4 If we use possible world semantics, the extensions of expressions may vary from world to world, but then the modal profile of the term's extensions would have to be known if a single model is used. Moreover, in such semantics there's usually an "actual world" singled out which would have to match the actual extensions of terms.

Let us review Glanzberg's reasoning: Natural language semantics should indicate whether natural language has a genuine logical consequence relation; the subject matter of natural language semantics is linguistic competence; a key aspect of linguistic competence is knowledge of truth conditions; truth conditions do not require a range of models; a genuine logical consequence relation requires a range of models; therefore, there is no genuine logical consequence relation in natural language. It seems to me that all that this reasoning establishes is that the study of truth conditions in natural language is not identical to the study of logical consequence in natural language, a mark of the difference is that one uses a range of models and the other does not. Glanzberg begs the question when he looks for logical consequence in natural language by looking at a discipline which he defines through its subject matter, which is not logical consequence.

In Glanzberg's words: "semantics of natural language—the study of speakers' semantic competence—cannot look at [a range of models] and still capture what speakers understand" (2015, 91). The present claim would thus be that while a range of models would not give you all that is understood by speakers, it is what it takes to give a logical consequence relation in natural language.

This is not to claim that natural language semantics is the wrong place to look for logical consequence. We are still left with the possibility that there is a sub-phenomenon that can be identified as a logical consequence relation. Now, entailment, which is a phenomenon studied by natural language semantics, is a wider category than logical consequence according to Glanzberg. So if it is the putative narrower phenomenon of logical consequence in natural language that we were to study, we would need to adjust our toolkit accordingly. We would need to appeal to a range of models. Acknowledging this is not to dispute that natural language semantics, as the study of truth conditions and entailment, is absolute—it is merely to distinguish another, related (indeed—narrower) phenomenon.

We should add that looking at a range of models does not require more information on words' extensions beyond what natural language semantics gives us. Defining "Ann" as a singular term whose extension varies between models requires less information than giving its absolute extension. And so, natural language semantics contains all the information that is needed for the range of models involved. We may thus still agree with Glanzberg that natural language semantics is the place where we should look for a relation of logical consequence in natural language, if such exists—and we may even find it there. If it is the range of all entailments with which a native speaker is

doi: 10.48106/dial.v74.i2

competent, then they are *inter alia* competent with the subset of entailments that are logical. If a competent speaker knows truth conditions of sentences most generally, then they also have the specific knowledge that is required for merely the logical entailments, as the latter is contained in the former. This point is also relevant to Glanzberg's *argument from lexical entailments*, to which we turn next.

At this point, however, we might be accused of overlooking an important piece of information required for moving to a range of models: we need to be able to distinguish between the logical and the nonlogical vocabulary. That is because, when moving to a range of models, we let the extensions of nonlogical expressions vary (according to their semantic category), while the extensions of the nonlogical vocabulary remain fixed. It might then be claimed that the distinction between logical and nonlogical expressions is not provided by natural language semantics, and that it extends the phenomena that can be found in natural language. Indeed, this is Glanzberg's argument from logical constants—which we address in section 4.

3 The Argument From Lexical Entailments

Next, Glanzberg presents the *argument from lexical entailments*. While natural language semantics does not require a range of models, it does look at the range of possibilities that account for truth conditions. The nearest thing to logical consequence that we find, then—according to Glanzberg—are entailment relations. However, entailment, as we have seen, is presumably much broader than a restrictive notion of logical consequence, since it includes analytic and metaphysical implications. Furthermore, entailments seem to completely forgo formality—many entailments depend on lexical components of sentences. Here enters an additional assumption made by Glanzberg, concerning formality. What determines the forms of sentences are *logical constants*, and logical consequence holds in virtue of their properties (Glanzberg 2015, 77). The meanings of the nonlogical vocabulary are abstracted away. Indeed, as we've mentioned, the standard model-theoretic conception of logical consequence has us completely fix the meanings of some of the vocabulary (the logical terms) and maximally vary, in line with semantic category, the meanings of the rest of the vocabulary (the nonlogical terms). On this common conception, if an argument is accepted as valid, and the validity of an argument depends on the specific meaning of an expression appearing in it, that expression must be treated as logical, and its meaning should be fixed across models.

The logical vocabulary, on this conception, constitutes a small, distinguished subset of the whole vocabulary. In standard first order logic we include the truth-functional connectives and the universal and existential quantifiers. Glanzberg mentions that logical constants normally have certain criteria imposed on them, such as topic-neutrality or permutation or isomorphism invariance. We shall mention criteria for logical vocabulary in the next section. Here, we may note that a choice of logical vocabulary determines a consequence relation. Moreover, the stricter we are with respect to logicality of expressions, the more restrictive is the consequence relation that results.

Now, entailment is a phenomenon in natural language, and, as implicated by Glanzberg, it is the most reasonable candidate for being natural language's logical consequence relation. Entailments, however, according to Glanzberg, depend on the meanings of nonlogical expressions.

Glanzberg provides the following examples of entailments to prove his point:

> (1) a. We loaded the truck with hay.
> ENTAILS
> We loaded hay on the truck.
> b. We loaded hay on the truck.
> DOES NOT ENTAIL
> We loaded the truck with hay.
>
> (2) John cut the bread.
> ENTAILS
> The bread was cut with an instrument.
>
> [...]
> These entailments are fixed by aspects of the meanings of words like "load" and "cut". (2015, 93–94)

The words "load" and "cut" are noncontroversial examples of *non*logical expressions—in a reasonably restrictive model-theoretic consequence relation they would not be fixed. One can presumably, on a permissive view of logic, study the logic of words like "load" and "cut," and so consider them as logical constants. But, according to Glanzberg, lexical entailments permeate language too far for us to have anything like a strict separation between logical and nonlogical constants. Practically every word would have to be considered as logical—that is since practically every word has lexical entailments that depend on its meaning. Furthermore, the lexical items above obviously do not fulfil accepted criteria for logicality.

The argument from lexical entailments may be objected to on two counts: one regarding the assumption that all lexical entailments as the examples above would have to be included in natural language's logical consequence relation, and another regarding the assumed conception of formality. As for the first: recall that according to Glanzberg, logical consequence is a narrower relation than that of entailment, and it is included in it. Above, we have examples of members of the difference between entailment and logical consequence. Entailments that are also logically valid would depend for their validity only on the meanings of the distinguished logical vocabulary (whatever that may be). What prevents us from taking these special entailments and marking them members of the logical consequence relation of natural language? Logical consequence, according to Glanzberg, is not a totally alien relation to natural language. Indeed, it is a subset of an accepted relation in natural language. What is to prevent us from marking it as its own phenomenon, in natural language?

Here is one way to respond. Take an accepted natural phenomenon, say that of organic compounds, studied in organic chemistry. Among the organic compounds, we have those liked by Sara the chemist. We thus have a subset of a chemical phenomenon that can hardly be considered as its own chemical phenomenon. So, while the items exemplifying the phenomenon fall squarely within the subject matter of the relevant science, what distinguishes them—being liked by Sara—is not a feature relevant to the science. Do we have the same case with logical consequence? Is its distinguishing feature a matter of the scientific study of language, and in particular, of natural language semantics?

In the previous section, I claimed that if logical consequence is a sub-phenomenon of entailment, then surely it calls for a proper adjustment of the toolkit for studying it, including a range of models rather than an absolute semantics. The argument from absolute semantics does not refute the existence of this sub-phenomenon. However, now we confront an intriguing question, for which I don't claim to have a definite answer: which distinctions are relevant to the subject matter of natural language, and which are not? We could aim at a principled definition of the subject matter involved to arbitrate the matter, or we might aim at more social considerations, and see whether work of researchers in the relevant field employ such distinctions. Observing the discipline of natural language, Glanzberg claims that while entailment is marked as a self-standing studied phenomenon, logical consequence is not. Now, on the assumption of formality, the matter turns on whether the distinction between logical and nonlogical expressions is relevant, whether it is one that can mark a phenomenon in natural language. This is the issue tackled in Glanzberg's *argument from logical constants*, with which we deal in the next section. There I shall object to Glanzberg's exclusion of the distinction between logical and nonlogical terms from the realm of natural language.

I've mention another line of objection to the argument from lexical entailments, having to do with the assumption of formality. Admittedly, formality is a widely accepted a condition on logical consequence (Beall, Restall and Sagi 2019).[5] Glanzberg can certainly not be blamed for assuming the common conception of formality on which to base his conclusion against the existence of a restrictive logical consequence relation in natural language. However, for the sake of the more general discussion, I'd like to mention an alternative approach to logical consequence, which may still accept the examples of entailments above as logical validities without trivializing formality. Note that in order to capture the above entailments, all that is needed is some restriction on the meaning of the words "load" and "cut" or their meanings' relations with the meanings of other words. Indeed, one need not completely fix the extension of these words in order to obtain these entailments. In previous work, I have proposed a model-theoretic framework for logical consequence where there is no strict division of the vocabulary into logical and nonlogical: terms are fixed in various manners and to various degrees using *semantic constraints*—restrictions on admissible interpretations of terms (Sagi 2014). As we have clauses in standard first order logic fixing the interpretation of the logical vocabulary, we may have clauses only restricting the interpretations of terms without fixing them completely.[6] Without pursuing this line any fur-

5 Notwithstanding some exceptions, *debunkers* by the terminology of MacFarlane (2015), by whom logical consequence is not defined as formal, even if logicians avail themselves with formal tools to study this relation (see Read 1994; and other references in MacFarlane 2015).

6 These clauses may remind of *meaning postulates*, as in Carnap (1952); Montague (1974). An important difference is that while for Carnap and Montague the clauses for the logical vocabulary are treated as basic, onto which meaning postulates are added, in the framework of semantic constraints all kinds of constraints (whether those completely fixing the meaning of a term or those akin to meaning postulates, only restricting meanings of terms) are treated on a par, and they determine the forms of sentences—and thus the formality of the obtained consequence relation is upheld.

doi: 10.48106/dial.v74.i2

ther,[7] we may take note that there are alternative approaches to formality, on some of which the logical validity of the arguments above does not entail that "load" and "cut" need to be fixed as logical terms. On such approaches, it may very well turn out that entailment itself is a formal relation, and constitutes the logical consequence of natural language.

4 The Argument From Logical Constants

Finally, Glanzberg presents the *argument from logical constants*. We have mentioned the criterion for logical terms of invariance under isomorphisms. The idea is the following. Logical terms are general, and they do not make distinctions between elements of the domain. Therefore, their extension remains constant under permutations of the domain: switching between members of the domain cannot entail a difference in the extension of a logical term. For example, the extension of the first-order existential quantifier is taken to be the set of all nonempty subsets of the domain, and so it is invariant under isomorphisms: no permutation of the domain can transform a nonempty set into an empty one, or vice versa. Similarly, logical terms are indifferent to switching between members of the domain and members of other domains, and are therefore invariant under isomorphisms.

We shall leave technicalities aside to the extent that we can.[8] Here it would suffice to acknowledge the role of the criterion of invariance under isomorphisms in a current conception of logical consequence. This criterion has been defended extensively in the literature (Sher 1991, 1996) or at least accepted as a necessary condition for logicality. By this criterion, the standard quantifiers and identity relation of first order logic are logical, but in addition, so are the variety of generalized quantifiers, such as *Most* and *There are infinitely many*. Thus, one might think that the grammatical category of determiners in natural language includes logical constants that would salvage formality and the feasibility of a logical consequence relation in natural language. For instance, let us observe the semantic clause for the determiner "most" (cf. Glanzberg 2015, 98):

a. Local: $[\![\text{most}]\!]_M = \{\langle A, B\rangle \subseteq \mathcal{P}(M)^2 : |A \cap B| > |A \backslash B|\}$
b. Global: function from M to $[\![\text{most}]\!]_M$

7 I intend to explore applications of the framework of semantic constraints to natural language semantics in future work.

8 For a detailed survey, see Westerståhl (1989).

The semantic clause has a *local*, absolute, part, which, given a (or rather, "the") domain, returns pairs of subsets of the domain satisfying the condition. The second part of the clause generalizes over all model domains, making the operator *global*. According to Glanzberg, all that semantic theory requires is the local condition: this condition suffices for accounting for truth conditions of sentences involving "most." Why then do we have the global extension? Glanzberg (2015, 99) contends that the global condition serves as a useful abstraction, that goes beyond the needs of semantic theory. And so, some properties of determiners can only be captured through their global definition:

a. CONSERV (local): For every $A, B \subseteq M$, $Q_M(A, B) \Leftrightarrow Q_M(A, B \cap A)$
b. UNIV (global): For each M and $A, B \subseteq M$, $Q_M(A, B) \Leftrightarrow Q_A(A, A \cap B)$

It is claimed that natural language determiners satisfy these conditions, and thus they in fact express restricted quantification. The global property UNIV is generally stronger (see also Westerståhl 1985), and it requires a range of models. Glanzberg explains that at this point we depart from natural language semantics:

> In looking at this sort of global property, we are not simply spelling out the semantics of a language. Rather, we are abstracting away from the semantics proper—the specification of contributions to truth conditions—to look at a more abstract property of an expression. (2015, 100)

On Glanzberg's approach, what decides whether some phenomenon is part of natural language is its relevance to the determination of truth conditions. Glanzberg raises the option of still viewing isomorphism invariant determiners as logical constants, since they have a property accepted by many as a distinguishing feature of logical constants. But according to Glanzberg, what is distinctive of such expressions is that they are amenable to extensive mathematical treatment—a property held by non isomorphism invariant terms as well. "So," Glanzberg concludes, "natural language will not hand us a category of logical constants identified by having a certain sort of mathematically specifiable semantics." And "Is there anything else about language—anything about its grammar, semantics, etc.—that would distinguish the logical constants from other expressions? No" (2015, 101). By more permissive lights, not limited to isomorphism invariance, we might accept the greater class of functional categories as including the logical expressions of a language,

doi: 10.48106/dial.v74.i2

which is distinguished grammatically. But if we remain within the restrictive viewpoint, we see, according to Glanzberg, that logicality is not part of natural language.

To the argument from logical constants too I object on two counts. Natural language contains expressions that satisfy accepted criteria for logicality, such as "most" and "more." Specifically, these expressions are invariant under isomorphisms. Glanzberg claims that this criterion does not latch onto a natural phenomenon, and the category of logical constants is not recognized by natural language. Now, while isomorphism invariance might not delineate a standard grammatical category, it does, arguably, spell out a property that distinguishes some expressions from others. An expression that is invariant under isomorphisms arguably does not distinguish the identity of individuals (Sher 1991, 43). This is a property that, in this view of logicality, makes these terms logical. If there is a phenomenon such as logical consequence in natural language, and logical consequence is analysed as requiring a distinguished set of logical terms, then this distinction would be made in its theory. So if invariance under isomorphisms is accepted as the distinguishing criterion, and there are expressions in natural language that satisfy it, what else do we need in order to say that there is a category of logical expressions in natural language?

Now, echoing the discussion from section 3, one might not be satisfied with this response. Perhaps, still, this distinction is artificial, and logical consequence is thus forced on natural language. It is unclear what makes a distinction external or artificial, but we can claim that in this case, indeed, one can defend the distinction and argue further against the putative artificiality. Moreover, whether or not natural language distinguishes between logical and non-logical terms is not a settled matter in the literature. Glanzberg takes the work in Westerståhl (1985) to go beyond natural language semantics, perhaps because of its highly abstract, mathematical nature. But we can find the relevant distinction in more empirically-oriented, mainstream natural language semantics. In some recent studies in linguistics it has been proposed that language does indeed separate between logical and other entailments. Gajewski (2002) argues for a category of sentences that are L-analytic—true or false in virtue of form—as a special case of ungrammaticality, based on speakers' intuitions. Presumably, his account can be extended to include entailments. Fox (2000) and Fox and Hackl (2006) argue that the cognitive system contains a deductive system in which sentences are evaluated and ruled out if they can be proven to be contradictory. Fox's characterization of the deductive system, as well as Gajewski's characterization of the L-analytic sentences employ a distinction between logical and non-logical words, where logical words correspond roughly to the logical terms in standard first order logic. Chierchia builds on these ideas to develop a full-fledged theory of the relation between logicality and grammar. According to Chierchia, it may be that logic and grammar are distinct computational systems, yet they are interfaced with each other. Logic, in any such case, is a natural phenomenon, and its notions play a central role in grammar (Chierchia 2013). If contemporary semantic theory sets the standard, then there is a basis for distinguishing a class of logical expressions.

5 Modelling Logical Consequence in Natural Language

Glanzberg indicates two ways that can lead us to accept a version of the thesis of logical consequence in natural language. One is by considering logical consequence from a more permissive perspective. We shall not discuss this option. The other is by a process of stepping away from semantics proper to obtain a logic. The process is threefold. We first identify the logical vocabulary, by whichever criterion we choose to employ—which (if minimally restrictive) will already at this point take us beyond natural language semantics (according to Glanzberg). Next, we abstract away from the meanings of the nonlogical expressions and allow for a range of domains—and in this way we obtain a range of models that will move us away from absolute semantics. And then, we idealize: natural language is full of exceptions and grammatical complications absent in logical systems. The outcome would be much more similar to a consequence relation in a formal language than what we seemed to have started out with. Indeed, Glanzberg contends that the result of this process is a logical consequence relation, and moving away from natural language makes it possible.

Now, let us consider the process Glanzberg describes, that we briefly delineated above. I'd like to argue that this process enforces the stance that there *is* a relation of logical consequence in natural language, and that through the said process we can model this phenomenon. Recall our discussion in the introduction. When we use a formalism to model a natural phenomenon, it will include representors and artefacts: aspects or elements that will correspond to features of the phenomenon modelled, and those that do not. We invariably idealize and abstract away from many of the features of the phenomenon. Does this mean that what we describe was not really out there, and was made

doi: 10.48106/dial.v74.i2

possible by the process of modelling? Rarely in science does a phenomenon simply jump out at us through a microscope: modelling is part and parcel of the study of complex phenomena. Glanzberg himself relates the process he describes to modelling in science:

> Idealization, as it figures here, is a familiar kind of idealization in scientific theorizing that builds idealized models. One way to build idealized models is to remove irrelevant features of some phenomenon, and replace them with uniform or simplified features. A model of a planetary system is such an idealized model: it ignores thermodynamic properties, ignores the presence of comets and asteroids, and treats planets as ideal spheres (cf. Frigg and Hartmann 2012). When we build a logic from a natural language, I suggest, we do just this. We ignore irrelevant features of grammar, and replace them with uniform and simplified logical categories. (2015, 113f)

Is a planetary system not a natural phenomenon, and part of the subject matter of astronomy? Is it merely a product of modelling, or is it the target phenomenon of a highly abstract model? Inasmuch as the planetary system is a natural phenomenon, and relevantly analogous to logical consequence in natural language, then logical consequence in natural language too is a natural phenomenon.

How could one still question that the process yields a model of logical consequence as a part of natural language? The only stage that can raise doubts is that of identification. Abstraction and idealization are no doubt a part of modelling. The question is whether we are identifying any real phenomenon. If not, then there is nothing that would tie our model to empirical reality. In the arguments from lexical entailment and from logical constants, Glanzberg relies on a certain conception of the formality of logic, that includes the following two assumptions: that a sharp division of the vocabulary into logical and nonlogical is material for the determination of the relation of logical consequence, and that invariance under isomorphisms is a good criterion to be considered in this discussion. So what needs to be identified here is the category of logical constants. Another way to put the question is to ask whether logical constants in our theory are representors or merely artefacts. We have already disputed Glanzberg's arguments against their identification capturing something real. So, given some assumptions accepted by Glanzberg, we see that the process of identification, abstraction and idealization, rather than bringing into natural language something new, reveals a feature of it by way of modelling. We may conclude this section with the claim that the logic in natural language thesis is still a viable one.

6 Conclusion

Let us take stock. The question of logical consequence in natural language is a fundamental one. In order to make any kind of progress, we must explicate the question, and give a clear understanding of what either a positive of a negative response would entail. Michael Glanzberg gives us, besides arguments for a specific response, a basis on which this question can be discussed and understood. The present critique is meant to pick up the discussion, and hopefully move it forward.

We've seen that there are reasons to doubt that natural language semantics is absolute, as claimed by Glanzberg. We've also seen that even if it is absolute, this does preclude the study of phenomena in natural language from appealing to a range of models. We take it on board that the putative phenomenon of logical consequence in natural language would constitute a relation that is included in that of entailment. Now, as we've briefly mentioned, one might find a way to define logical consequence as a formal relation so that it *coincides* with the relation of entailment. Admittedly, that would take a permissive approach to logical consequence by Glanzberg's lights. Alternatively, we might distinguish a subset of entailments as the relation of logical consequence in natural language. While entailments may depend on the meanings of any expressions in the language, logical validities depend only on the logical vocabulary. So in order to distinguish logical consequence as a relation in natural language, we need to identify the logical vocabulary. The logical vocabulary may be characterized by a widely accepted criterion of invariance under isomorphisms.

The question is then whether this feature is one that falls within the purview of natural language. If natural language semantics is the relevant discipline to be studying the putative relation of logical consequence in natural language, the question is whether the distinction between logical and nonlogical terms is relevant to natural language semantics. Logical terms, characterized by isomorphism invariance, are general in that they make no distinction among individuals in a given domain. I see no reason why natural language semantics should not help itself to such a property. Indeed, we've cited linguists who

doi: 10.48106/dial.v74.i2

appeal to this property as an integral part of their work—are they all not studying natural language anymore, when they appeal to this property, but are rather doing something else? This, I would take to be a contentious claim. In sum, the logic in natural language thesis has not been refuted by Glanzberg's arguments.

The thesis of logic in natural language is reinforced when we consider how the relation of logical consequence can be identified and studied through a process of modelling. Glanzberg contends that we can obtain a relation of logical consequence in natural language through a process of identification (of the logical vocabulary), abstraction and idealization. I have suggested that as long that we are identifying something real—as long as our model in the end contains representors of a real phenomenon—what we obtain through the delineated process is a model of a real phenomenon. Specifically, if logical constants in the formalism we use do indeed represent a feature of natural language, then through the formalism we obtain a model of a bona fide linguistic phenomenon.

There is a long-standing sentiment that logic and natural language are disparate entities, and that it is a mistake to associate one with the other. Glanzberg gives substance to this sentiment through meticulous analysis and argumentation. However, I have argued that Glanzberg's approach may very well lead us to *accept* the thesis of logic in natural language. This leaves us with a negative option and with a positive option: either find what it is that may still drive logic and natural language apart that goes beyond Glanzberg's assumptions,[9] or use the tools of natural language semantics or empirical linguistics more generally figure out what the logic of natural language just is.[*]

Gil Sagi
0000-0002-7101-9927
University of Haifa
gsagi@univ.haifa.ac.il

9 It seems to me that a characterization of logic as a normative discipline, e.g. along the lines of the traditional-methodological project delineated in the introduction might provide a basis for the claim that there is no logical consequence in natural language.

* Versions of this paper were presented at the Hebrew University, the University of St Andrews, the University of Leeds and the University of Bergen. I thank the audiences there for a fruitful discussion. I also thank David Kashtan, Ran Lanzet, Jack Woods and two anonymous reviewers for helpful comments.

References

Beall, J. C., Restall, Greg and Sagi, Gil. 2019. "Logical Consequence." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/spr2019/entries/logical-consequence/.

Carnap, Rudolf. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 1st ed. Chicago, Illinois: University of Chicago Press. Second edition: Carnap (1956).

—. 1952. "Meaning Postulates." *Philosophical Studies* 3(5): 65–73. Reprinted in Carnap (1956, 222–229), doi:10.1007/bf02350366.

—. 1956. *Meaning and Necessity: A Study in Semantics and Modal Logic*. 2nd ed. Chicago, Illinois: University of Chicago Press. Enlarged edition of Carnap (1947).

Chierchia, Gennaro. 2013. *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199697977.001.0001.

Chierchia, Gennaro and McConnell-Ginet, Sally. 1990. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, Massachusetts: The MIT Press. Second edition: Chierchia and McConnell-Ginet (2000).

—. 2000. *Meaning and Grammar: An Introduction to Semantics*. 2nd ed. Cambridge, Massachusetts: The MIT Press. First edition: Chierchia and McConnell-Ginet (1990).

Chomsky, Noam. 1957. *Syntactic Structures*. Den Haag: Mouton.

Cook, Roy T. 2002. "Vagueness and Mathematical Precision." *Mind* 111(442): 225–246, doi:10.1093/mind/111.442.225.

Davidson, Donald. 1967. "Truth and Meaning." *Synthese* 17(1): 304–323. Reprinted in Davidson (1984, 17–36), doi:10.1007/BF00485035.

—. 1970. "Semantics for Natural Languages." in *Linguaggi nella società e nella tecnica. Museo Nazionale della Scienza e della Tecnica Milano, 14-17 ottobre 1968*, edited by Bruno Visentini, pp. 177–188. Saggi di cultura contemporanea n. 87. Milano: Edizioni di Communità. Reprinted in Davidson and Harman (1975, 18–24).

—. 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, doi:10.1093/0199246297.001.0001.

Davidson, Donald and Harman, Gilbert H., eds. 1972. *Semantics of Natural Language*. Synthese Library n. 40. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-010-2557-7.

—, eds. 1975. *The Logic of Grammar*. Encino, California: Dickenson Publishing Co.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Cambridge, Massachusetts: Harvard University Press.

Fox, Danny. 2000. *Economy and Semantic Interpretation*. Linguistic Inquiry Monographs n. 35. Cambridge, Massachusetts: The MIT Press.

doi: 10.48106/dial.v74.i2

Fox, Danny and Hackl, Martin. 2006. "The Universal Density of Measurement." *Linguistics and Philosophy* 29(5): 537–586, doi:10.1007/s10988-006-9004-4.

Frigg, Roman and Hartmann, Stephan. 2012. "Models in Science." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/fall2012/entries/models-science/.

Gajewski, Jon Robert. 2002. "On Analyticity in Natural Language." Unpublished manuscript, https://jon-gajewski.uconn.edu/wp-content/uploads/sites/1784/2016/08/analytic.pdf.

Glanzberg, Michael. 2014. "Explanation and Partiality in Semantic Theory." in *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman, pp. 259–292. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199669592.003.0011.

—. 2015. "Logical Consequence and Natural Language." in *Foundations of Logical Consequence*, edited by Colin R. Caret and Ole Thomassen Hjortland, pp. 71–120. Mind Association Occasional Series. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198715696.003.0003.

Heim, Irene and Kratzer, Angelika. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell Publishers.

MacFarlane, John. 2015. "Logical Constants." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language and Information, https://plato.stanford.edu/archives/sum2015/entries/logical-constants/.

Montague, Richard. 1974. *Formal Philosophy: Selected Papers*. New Haven, Connecticut: Yale University Press. Edited and with an introduction by Richmond H. Thomason.

Read, Stephen. 1994. "Formal and Material Consequence." *Journal of Philosophical Logic* 23(3): 247–265, doi:10.1007/bf01048482.

Sagi, Gil. 2014. "Formality in Logic: From Logical Terms to Semantic Constraints." *Logique et Analyse* 57(227): 259–276, doi:10.2143/LEA.227.0.3053506.

—. 2020. "Logic and Natural Language: Commitments and Constraints." *Disputatio* 12(58): 377–408, doi:10.2478/disp-2020-0014.

—. 2021. "Logic as a Methodological Discipline." *Synthese* 199(3-4): 9725–9749, doi:10.1007/s11229-021-03223-3.

Shapiro, Stewart. 1998. "Logical Consequence: Models and Modality." in *The Philosophy of Mathematics Today*, edited by Matthias Schirn, pp. 131–156. Oxford: Oxford University Press, doi:10.1093/oso/9780198236542.003.0005.

Sher, Gila Y. 1991. *The Bounds of Logic: A Generalized Viewpoint*. Cambridge, Massachusetts: The MIT Press.

—. 1996. "Did Tarski Commit 'Tarski's Fallacy'?" *The Journal of Symbolic Logic* 61(2): 653–686, doi:10.2307/2275681.

WESTERSTÅHL, Dag. 1985. "Logical Constants in Quantifier Languages." *Linguistics and Philosophy* 8(4): 387–413, doi:10.1007/bf00637410.

—. 1989. "Quantifiers in Formal and Natural Languages." in *Handbook of Philosophical Logic, Volume IV: Topics in the Philosophy of Language*, edited by Dov M. GABBAY and Franz GUENTHNER, pp. 1–131. Synthese Library n. 167. Dordrecht: D. Reidel Publishing Co. Reprinted in revised form as Westerståhl (2007), doi:10.1007/978-94-009-1171-0_1.

—. 2007. "Quantifiers in Formal and Natural Languages." in *Handbook of Philosophical Logic, Volume XIV*, edited by Dov M. GABBAY and Franz GUENTHNER, 2nd ed., pp. 223–338. Dordrecht: Springer. First publication as Westerståhl (1989), doi:10.1007/978-1-4020-6324-4_4.

ZIMMERMANN, Thomas Ede. 1999. "Meaning Postulates and the Model-Theoretic Approach to Natural Language Semantics." *Linguistics and Philosophy* 22(5): 529–561, doi:10.1023/A:1005409607329.

‘Unless’ is ‘Or,’ Unless ‘$\neg A$ Unless A’ is Invalid

Roy T. Cook

The proper translation of “unless” into intuitionistic formalisms is examined. After a brief examination of intuitionistic writings on “unless,” and on translation in general, and a close examination of Dummett’s use of “unless” in Elements of Intuitionism (1975b), I argue that the correct intuitionistic translation of “A unless B” is no stronger than “$\neg B \rightarrow A$.” In particular, “unless” is demonstrably weaker than disjunction. I conclude with some observations regarding how this shows that one’s choice of logic is methodologically prior to translation from informal natural language to formal systems.

The topic of this essay is a methodological principle at work within both pedagogical and theoretical contexts—one which is widely accepted, albeit for the most part implicitly and uncritically. The assumption in question is that the translation of informal, natural language claims into one or another formal language is logic neutral. This assumption underwrites our standard logical practices—evidenced both within the classroom and within the peer-reviewed research paper—whereby we first formalize natural language claims into a favored artificial language and only then pronounce judgement on this single, univocal formalization from the perspective of this or that logic.

Here we will see that this methodology is deeply flawed. On the contrary, we must *first* decide which logic (classical, intuitionistic, dialethic, quantum, etc.) is at work, and only then can we provide adequate translations of informal, everyday natural language expressions into whatever formal language is in play. The reason is simple to state, although defending it will require a bit of work: the same natural language expressions should be translated differently with respect to different background logics.

The argument that translation of natural language claims into formal language is neither prior to, nor independent of, our choice of one or more logics as “correct” (or, at least, as the logic currently under consideration) will focus on a particularly interesting and, in this author's opinion, under-examined example: "unless." As we shall see, the natural language connective "unless" turns out to be a particularly clear case of the phenomenon in question, since the intuitionist should translate claims involving "unless" very differently from the standard rule:

unless = (inclusive) disjunction

commonly taught to students and implicitly accepted in much professional work on logic (including much work on non-classical logic). The remainder of this essay will develop this argument as follows.

First, in section 1, we will look at the standard logical treatment of "unless," where natural language claims of the form "Φ or Ψ" are translated as (or as something *classically* equivalent to) "$\Phi \vee \Psi$," and we will examine the various options available in an intuitionist context, where, for example, "$\Phi \vee \Psi$," "$\neg\Phi \rightarrow \Psi$," and "$\neg\Psi \rightarrow \Phi$" are not equivalent.

In section 2 we will undertake a careful examination of a number of instances of "unless" claims found in *Elements of Intuitionism*, Michael Dummett's classic text on intuitionistic mathematics. As we will see, translating these in terms of disjunction—that is, via application of the rule typically taught to students and uncritically applied by their teachers—produces results that do not accurately capture the content of the original claims. In particular, while the classical logician should (or at least can) translate "unless" claims as disjunctions, the intuitionist should not, since from an intuitionionistic perspective "unless" is weaker than "or."

For the purposes of the remainder of the essay, all that will be needed from section 2 is the relatively weak claim that, intuitionistically at least, claims of the form "Φ unless Ψ" are weaker than claims of the form "Φ or Ψ"—and hence, intuitionists should abandon the "unless-is-or" equation. Interestingly, however, the evidence marshaled in this section supports a stronger claim: the intuitionistically correct translation of natural language claims of the form "Φ unless Ψ" is "$\neg\Psi \rightarrow \Phi$."

In section 3 we will make some additional observations about translation "unless" claims from an intuitionistic perspective, and deal with a few complications raised by the data examined in section 2, including the fact that the translation manual endorsed in that section makes "unless" claims fail to be commutative—that is, "Φ unless Ψ" is not always logically equivalent to "Ψ unless Φ."

Then, in section 4 we will use a toy version of Putnam's argument for quantum logic (1969) to show that the priority of choice of logic to translation in general, and the proper translation of "unless" in particular, is not a trivial or minor matter. In particular, the observations made in previous sections have profound ramifications regarding the shape that arguments for logical revision must take. Perhaps the most striking such consequence is that a particular counterexample to a particular logic—that is, an argument that is valid according to that logic, but which has a true premise and non-true conclusion—can never force us to give up a particular logical law (such as the law that takes us from that premise to that conclusion). Instead, it is always, at least in principle, open to us to argue that the natural language premises and conclusion have been translated incorrectly relative to the standards of the logic in question (which need not be identical to the standards appropriate to the logic with which our opponents wish to replace our own favored system).

Finally, in the concluding section 5 we will tie up some loose ends and note some consequences all of this has for the so-called communication problem: the problem of determining whether or not intuitionists and classical logicians (or any two camps accepting different logics as correct) mean the same thing by logical notions such and "or" and "unless."

1 Translating "Unless"

Consider how "unless" is usually handled in basic logic courses. In such courses, students are often initially confused with regard to how we ought to translate the natural language expression "unless." One common strategy for providing students with some basic insights regarding this translational conundrum is to point out (typically via clear examples) that "unless" seems to obey the following two rules of inference:

$$\begin{array}{c} \Phi \text{ unless } \Psi \\ \underline{\text{Not: } \Phi} \\ \Psi \end{array} \qquad\qquad \begin{array}{c} \Phi \text{ unless } \Psi \\ \underline{\text{Not: } \Psi} \\ \Phi \end{array}$$

These facts suggest that "Φ unless Ψ" could be plausibly translated as "$\neg\Phi \rightarrow \Psi$," or perhaps "$\neg\Psi \rightarrow \Phi$" (or perhaps even "$(\neg\Phi \rightarrow \Psi) \wedge (\neg\Psi \rightarrow \Phi)$" or "$(\neg\Phi \rightarrow \Psi) \vee (\neg\Psi \rightarrow \Phi)$"). The instructor then typically points out that:

$$\neg\Phi \rightarrow \Psi \dashv\vdash_C \Phi \vee \Psi$$
$$\neg\Psi \rightarrow \Phi \dashv\vdash_C \Phi \vee \Psi$$

doi: 10.48106/dial.v74.i2

Hence, the proper translation of "Φ unless Ψ" is "$\Phi \vee \Psi$" (or any of the logical equivalents mentioned above).[1]

Note, however, that all of this depends on the fact that introductory courses on formal logic are typically restricted to instruction on, and from the perspective of, classical logic. Imagine, however, that an intuitionistic logician teaches a course on basic logic (something that happens all the time) and further that she teaches her students intuitionistic logic (H) and teaches it from the perspective of an intuitionist (something that happens far less frequently).

Now, when discussing the proper translation of "unless" claims, even if the intuitionist argued, just as the classical logician did, that both of the argument patterns identified above seem valid (and, as we shall see, there are good reasons for being suspicious of the first argument pattern from an intuitionistic perspective!), she cannot follow her classical counterpart in concluding that this alone shows that "$\Phi \vee \Psi$" is a legitimate translation of "Φ unless Ψ." The reason is simple: The classical logician uses the *classical* logical equivalence of these more complex formulations and "$\Phi \vee \Psi$" to argue that the latter is the preferred, simplest formalization of the natural language expression "unless." For the intuitionist, however, "$\Phi \vee \Psi$" and "$(\neg\Phi \rightarrow \Psi) \wedge (\neg\Psi \rightarrow \Phi)$" are not equivalent. Moreover, each formula in the following diagram is classical logically equivalent to all of the others, but no two are intuitionistically equivalent (transitive closure of the arrows indicates entailment):[2]

---

1 Arguably, there is a stronger, *exclusive* reading of "unless"—that is, a reading of "Φ unless Ψ" that entails "not both Φ and Ψ"—that occurs in sentences such as:

> You will get soup unless you get salad.

This reading of unless also has multiple possible, non-equivalent translations for the intuitionist. We will leave construction and consideration of such translation manuals to the energetic reader.

2 Note that we need not restrict our attention to this handful of simple translations. There are many other interesting, disjunction-like operators definable within intuitionistic logic. Interesting examples include *pseudo-disjunction*:

$$\Phi \dot{\vee} \Psi =_{\text{df}} ((\Phi \rightarrow \Psi) \rightarrow \Psi) \wedge ((\Psi \rightarrow \Phi) \rightarrow \Phi)$$

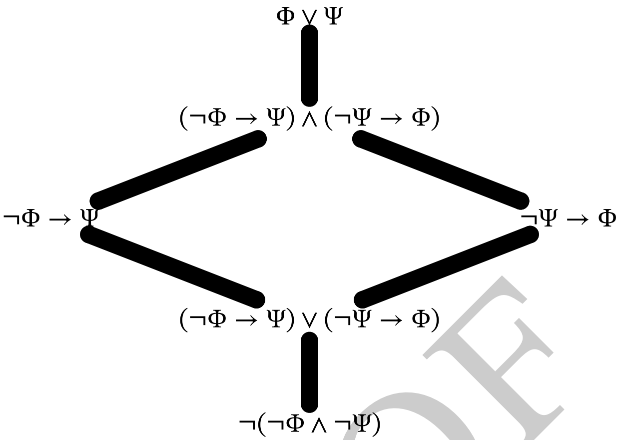
Church disjunction:

$$\Phi \ddot{\vee} \Psi =_{\text{df}} (\Phi \rightarrow \Psi) \rightarrow ((\Psi \rightarrow \Phi) \rightarrow \Phi)$$

and *Cornish disjunction*:

$$\Phi \star \Psi =_{\text{df}} (((\Phi \rightarrow \Psi) \rightarrow \Psi) \rightarrow \Phi) \rightarrow \Phi$$

These are examined in detail in Humberstone (2011)—the absolutely definitive and authoritative study of propositional connectives in classical and non-classical logics—on pages 555, 235, and 235 respectively.



Thus, there are (at least) six different rules that the intuitionist could adopt for translating "unless":[3]

1. Φ unless $\Psi = \Phi \vee \Psi$

---

3 Note that only rules [1] through [3] validate the informal inference:

$$\begin{array}{c} \Phi \text{ unless } \Psi \\ \text{Not: } \Phi \\ \hline \Psi \end{array}$$

and only rules [1], [2], and [4] validate:

$$\begin{array}{c} \Phi \text{ unless } \Psi \\ \text{Not: } \Psi \\ \hline \Phi \end{array}$$

All six rules validate variants of these rules where the conclusions are replaced by their double negations, of course. Since discussions of translation within logic courses and texts that focus on classical logic elide the difference between $\neg\neg\Phi$ and Φ (if, as the intuitionist claims, there is such a difference), then considering all of these possible translations seems wise, and is, at any rate, harmless even if in the end we accept one or both of these rules as valid on the intuitionistic understanding of "unless."

Here and below, we will speak of translation rule [1] being the "strongest" rule (and rule [6] being the "weakest" rule), as shorthand for the claim that rule [1] outputs the (intuitionistically) strongest translation of "unless" claims (and rule [6] outputs the weakest such translation) with the ordering understood as the partial ordering corresponding to our diagram.

2. Φ unless $\Psi = (\neg\Phi \rightarrow \Psi) \wedge (\neg\Psi \rightarrow \Phi)$
3. Φ unless $\Psi = \neg\Phi \rightarrow \Psi$
4. Φ unless $\Psi = \neg\Psi \rightarrow \Phi$
5. Φ unless $\Psi = (\neg\Phi \rightarrow \Psi) \vee (\neg\Psi \rightarrow \Phi)$
6. Φ unless $\Psi = \neg(\neg\Phi \wedge \neg\Psi)$

So how should an intuitionist translate "unless"? At the outset, it is worth pointing out one fact that seems like *prima facie* evidence against the claim that "Φ unless Ψ" should be translated as a disjunction: the fact that the "un-" in "unless" seems to encode a negation. Further, the "un-" seems to attach to "Ψ" in particular (as is evidenced by the slightly more pretentious, but presumably equivalent "Unless Ψ, Φ").[4] This suggests (but far from entails) that one of the other, weaker (negation-involving) formulas in the diagram above (i.e. one of rules [2] through [6]) is the best translation of the natural language expression "unless" into intuitionistic logic, and also (but perhaps more weakly) suggests that translation rule [4] is the correct intuitionistic translation of "unless."

There is, of course, another, rather simple way to obtain data relevant to determining the proper translation of "unless": we can just ask intuitionists. In a rare moment of empirical curiosity, and with this in mind, I asked Neil Tennant (via email) how he understood "unless." It turns out that he prefers the "exclusive" reading (see footnote 1), and provided (rules equivalent to) the following introduction rule:[5]

If: $\Delta, \Phi, \Psi \vdash \bot$; $\Delta, \neg\Phi \vdash \Psi$; and $\Delta, \neg\Psi \vdash \Phi$
Then: $\Delta \vdash \Phi$ unless Ψ

and elimination rules:

If: $\Delta_1 \vdash \Phi$; and $\Delta_2 \vdash \Psi$
Then: Δ_1, Δ_2, Φ unless $\Psi \vdash \bot$

4 This argument is analogous, perhaps, to the claim that the "if" in "Ψ, if Φ" attaches, in some sense, to the "Φ," which in turn helps to make vivid the equivalence between this claim and "If Φ then Ψ." Thanks to a referee for pointing this out.

5 A full and proper analysis of Tennant's response would require a bit more subtlety, since his Core Logic involves relevance constraints (e.g. transitivity fails, etc.)—see Tennant (2017). We set these complications aside here, however.

Thanks are of course owed to Tennant for permission to share the upshot of this correspondence.

If: $\Delta, \Phi \vdash \bot$
Then: Δ, Φ unless $\Psi \vdash \Psi$

If: $\Delta, \Psi \vdash \bot$
Then: Δ, Φ unless $\Psi \vdash \Phi$

Extrapolating analogues of these rules for a non-exclusive reading of “unless” is straightforward:

If: $\Delta, \neg\Phi \vdash \Psi$; and $\Delta, \neg\Psi \vdash \Phi$
Then: $\Delta \vdash \Phi$ unless Ψ

If: $\Delta, \Phi \vdash \bot$
Then: Δ, Φ unless $\Psi \vdash \Psi$

If: $\Delta, \Psi \vdash \bot$
Then: Δ, Φ unless $\Psi \vdash \Phi$

These rules clearly correspond to translation rule [2] above, where “Φ unless Ψ” is translated as “$(\neg\Phi \rightarrow \Psi) \wedge (\neg\Psi \rightarrow \Phi)$.” Thus, Tennant agrees with what will be one of the main conclusions of this paper: that translating “unless” as (or as equivalent to) “or” is incorrect.[6]

Perhaps the best way to determine how an intuitionist should translate “unless”—better even than asking them directly, given the unreliability of intuitions regarding logical form (and with apologies to Tennant!)—is to study the inferential patterns used by intuitionists when reasoning using “unless.”[7] Despite the fact that intuitionists are, sadly, few in number in comparison to their classical opponents, an exhaustive, scientifically compelling linguistic survey of most or all of their publications and pronouncements containing the expression “unless” is far beyond the scope of this essay (and the skills of its author). Thus, I will instead just present close examinations of a few striking and suggestive examples.

6 As we shall see, however, he disagrees with regard to what the *correct* translation is.

7 The point is not that our intuitions about logical form are somehow inherently suspect or are not legitimate data—on the contrary! The point, rather, is that in cases where our armchair, *a priori* philosophical intuitions about logical form conflict with the data obtained by empirically observing how the expressions are actually used (and, as we shall see, Tennant’s intuitions and the data presented in the next section are in just such conflict), it seems reasonable to privilege the linguistic data over the intuitions. And, in the interest of full disclosure (and also so Tennant doesn’t feel so alone!), my own intuitions agreed with his prior to looking at the data.

Presumably, we can find no better source for such examples than Michael Dummett's *Elements of Intuitionism* (1977). We will carry out such an examination of *Elements of Intuitionism* in the next section, where we shall see that there is a good bit of evidence in favor of translation rule [4] (and hence against [1]) as the correct intuitionistic translation of "unless"—evidence that is obtained by examining how intuitionists actually use (or, at least, how Dummett actually uses) "unless."

Before moving on, however, there is a complication that we need to deal with. It is well known that, even from a purely classical perspective, translating "unless" as "or" only works in *positive* contexts. In other words, when "unless" occurs in negative contexts, it appears to mean something different. This point is used by Higginbotham (1986), for example, to argue that "unless" is not compositional, since its meaning, and truth conditions, depend on the logical contexts within which it is embedded. Interestingly, Dummett uses "unless" in this sense at least once in *Elements of Intuitionism*:

> No account of the intuitionistic rejection of the law of excluded middle is adequate, therefore, *unless* it is based on the intuitionistic rejection of the platonistic notion of mathematical truth as obtaining independent of our capacity to give a proof. (1977, 12, emphasis added)

Even on a classical understanding of this claim, translating "unless" as "or" (or any of its logical equivalents discussed above) is inadequate, since doing so would entail that the quotation above is equivalent to:[8]

> It is not the case that there is an x such that either x is an adequate account of the intuitionistic rejection of the law of excluded middle or x is based on the intuitionistic rejection of the platonic notion of mathematical truth as obtaining independent of our capacity to give a proof.

8 Note that applying the exclusive reading of "unless" (i.e. "unless" as equivalent to exclusive disjunction) to this passage

> It is not the case that there is an x such that x is an adequate account of the intuitionistic rejection of the law of excluded middle if and only if x is not based on the intuitionistic rejection of the platonic notion of mathematical truth as obtaining independent of our capacity to give a proof

works no better.

This has the following logical form:

$$\neg(\exists x)(F(x) \lor G(x))$$

Higginbotham argues that occurrences of "unless" embedded in such negative contexts should be translated instead as "and not," resulting in something like[9]

> It is not the case that there is an *x* such that *x* is an adequate account of the intuitionistic rejection of the law of excluded middle and it is not the case that *x* is based on the intuitionistic rejection of the platonic notion of mathematical truth as obtaining independent of our capacity to give a proof

which has the following logical form:

$$\neg(\exists x)(F(x) \land \neg G(x))$$

This suggestion seems to capture at least the classical content of Dummett's claim relatively well—that is, it adequately captures how the sentence should be translated into the language of classical logic if the sentence had been uttered by a classical logician. Of course, given the occurrence of both existential quantification and negation—two notions that are central to the disagreement between classical and intuitionistic accounts of logic—in this translation, there might well be reasons to think that the intuitionistic translation of this passage should be different, similar to the reasons we shall see in the next section for objecting to the intuitionistic translation of "unless" as "or" in positive contexts.

For the sake of keeping this essay relatively short(ish) and snappy(ish), however, we will set aside the issue of translating "unless" when it occurs in the scope of negated quantifiers. The interested reader is encouraged to carry out their own textual analysis, similar to the one carried out for occurrences of "unless" in positive contexts, in order to determine if the intuitionist should adopt the same "and not" rule as the classical logician, or some intuitionistically non-equivalent (but presumably classically equivalent) formulation.

9 There is, of course, a significant literature in logic and linguistics arguing for various other ways of handling "unless" in negative contexts, including accounts that aim for a uniform approach that salvages compositionality. Since we are setting aside negative occurrences of "unless" here, we need not survey such accounts (interesting though they might be!)

doi: 10.48106/dial.v74.i2

2 "Unless" in *Elements of Intuitionism*

Let's now begin our examination of Dummett's use of "unless" in *Elements of Intuitionism*. We will not attempt to consider every occurrence of this expression in Dummett's book (we will include a footnote listing some additional examples, and explaining why they were not examined in detail here, near the end of this section). Instead, we will look at enough cases to:

- Demonstrate that translating intuitionistic utterances of "unless" as "or" is too strong—that is, we should reject translation rule [1] above in favor of something weaker (such as any of [2] through [6]).
- Construct a significant body of evidence in favor of translation rule [4] as the correct rule for translating informal "unless" claims into intuitionistic formal languages (i.e. "Φ unless Ψ" should be translated as "$\neg\Psi \rightarrow \Phi$").

Of course, the latter claim (that translation rule [4] is the correct rule) entails the former claim (that translation rule [1] is incorrect). But keeping these two claims separate in this way is useful for two reasons. First, I think it likely that many readers will find my arguments against rule [1] to be more definitive than my arguments in favor of rule [4]. As we've already noted, Tennant agrees with me about [1] being incorrect, but disagrees regarding rule [4] being the correct rule. Equally important, however, is the second reason for keeping these two claims separate: regardless of the ultimate fate of the latter, stronger claim, the incorrectness of translation rule [1] is all that is needed for the further arguments regarding logical revision that will be presented in section 4 below.

We will work through the first example in full and gory detail, and then work through additional examples somewhat more quickly and loosely. For our first such example, consider the following passage:

> A quasi-completeness proof of this kind can plainly be given only for a fragment of predicate logic within which the intuitionistically and classically provable formulas coincide (and not, as Kreisel points out, for every such fragment). As for the general case, it is evident from Theorem 5.37 that, *unless* we are prepared to accept schema (11) for primitive recursive predicates, we have no hope of proving even the quasi-completeness of any formalization of intuitionistic logic for which the extended Hauptsatz,

> which is a version of Herbrand's Theorem, holds. (Dummett 1977, 182)

In order to asses this occurrence of "unless," we need some of the mathematical background.

A logical system is *quasi-complete* if and only if every unprovable formula fails to hold on every internal interpretation (which is weaker than the requirement that there is a particular internal interpretation on which it does not hold). Schema (11) is:

$$(\forall \vec{u})(A(\vec{u}) \vee \neg A(\vec{u})) \wedge \forall \alpha \neg\neg \exists n A(\bar{\alpha}(n)) \rightarrow \neg\neg \forall \alpha \exists n A(\bar{\alpha}(n))$$

and the relevant portion of Theorem 5.37 states that, if HPC (intuitionistic predicate logic) is internally quasi-complete then all instances of Schema (11) hold where $A(\vec{x})$ is primitive recursive.[10]

So, with this in mind, how should we translate Dummett's claim that,

> [...] *unless* we are prepared to accept schema (11) for primitive recursive predicates, we have no hope of proving even the quasi-completeness of any formalization of intuitionistic logic for which the extended Hauptsatz, which is a version of Herbrand's Theorem, holds". (1977, 182)

Let's simplify Dummett's claim a bit, and instead consider the (slightly less poetic, but more precise) statement

> We are unable to prove the quasi-completeness of any formalization of HPC for which the Hauptsatz holds, unless we have reason to accept schema (11) for primitive recursive $A(\vec{x})$

and adopt the following translation manual:

> A = We are able to prove the quasi-completeness of some formalization of intuitionistic logic for which the Hauptzatz holds.

> B = We have reason to accept schema (11) for primitive recursive $A(\vec{x})$.

10 It is worth noting that Dummett also proves that, if Schema (11) holds for *all* $A(\vec{x})$ (primitive recursive or not), then ICP is internally quasi-complete for single formulas.

doi: 10.48106/dial.v74.i2

If translation rule [1], where "unless" is just disjunction, were correct, then we should formalize Dummett's claim as:

$$\neg A \vee B$$

Translating this back into natural language, this would entail that Dummett's claim is equivalent to the following:

> Either it is not the case that we are able to prove the quasi-completeness of any formalization of intuitionistic logic for which the extended Hauptsatz holds, or we have reasons to accept schema (11) for primitive recursive predicates.

Now, an intuitionist typically (and Dummett definitely) treats disjunction as determinate, in the sense that "$\Phi \vee \Psi$" is taken to be equivalent to something like:

> Φ is definitely the case, or Ψ is definitely the case.

or:

> Φ is the case, or Ψ is the case (and we can determine which).

Given this, however, the result of applying translation rule [1] to Dummett's natural language claim is immensely implausible. Earlier in the same chapter Dummett writes that:

> Unfortunately, there is no particular reason for supposing schema (11) to be intuitionistically valid; it can again be shown to be underivable in the usual systems of intuitionistic analysis, although there is not the same positive reason to suppose it invalid as there was in the case of (9). (1977, 176)

This quotation concerns, of course, schema (11) in full generality, rather than restricted to primitive recursive predicates, but the open status of (11) restricted to primitive recursive predicates is clearly expressed in a paper of Kreisel's upon which much of Dummett's discussion depends:[11]

11 "(3)" is Kreisel's label for (a principle shown by Dummett to be equivalent to) (11) restricted to primitive recursive predicates, and "weak completeness" is an alternative term for "quasi-completeness."

> […] (3) is not so implausible, and may be provable on the basis of as yet undiscovered axioms which hold for the intended interpretation (but not for the realizability interpretations). So the problem whether HPC is weakly complete is still open. (1962, 4)

Thus, it is neither the case that we definitely know (can prove) schema (11) restricted to primitive recursive predicates, nor that we can definitely refute (11) so restricted. And clearly such indecision applies to the claim about proving quasi-completeness as well: if we (i.e. Dummett, when writing the text) knew either that the internal quansi-completeness of ICP could be proven, or that it could be refuted, then surely he would have included such a proof (or at least a report of such a proof) in a chapter on completeness proofs for intuitionistic systems (or see the final sentence in the Kreisel quotation above).

But what about applying translation rules [2] through [6]? Which of the remaining translations of Dummett's "unless" claim into the language of intuitionistic logic are plausible, and which are not? If we apply our translation manual and rule [3], we obtain:

$$\neg\neg A \rightarrow B$$

which then translates back into informal prose as:

> If it is not the case that it is not the case that we have reasons to accept schema (11) for primitive recursive predicates, then we are able to prove the quasi-completeness of some formalization of intuitionistic logic for which the extended Hauptsatz holds.

This claim does not follow from Theorem 5.37 as stated, however. Theorem 5.37, as stated, amounts to:

$$A \rightarrow B$$

We can, of course, apply contraposition twice to obtain:

$$\neg\neg A \rightarrow \neg\neg B$$

which then translates back to something like:

> If it is not the case that it is not the case that we are able to prove the quasi-completeness of any formalization of intuitionistic logic for which the extended Hauptsatz holds, then it is not the case that

> it is not the case that we have reasons to accept schema (11) for primitive recursive predicates.

This does follow from Theorem 5.37. But this claim is strictly speaking weaker than the result of applying translation rule [3] (i.e. it is intuitionistically entailed by, but does not intuitionistically entail, the translation that results from applying rule [3]).

Given that Dummett asserts that the "unless" claim in question is *evident* from Theorem 5.37, this strongly suggests that translation rule [3] (and hence also against the stronger rule [2]) does not deliver the correct translation of "unless," since the result of applying this translation does not, contrary to Dummett's claim, actually follow from Theorem 5.37.[12]

Translation rule [4] fares better, however. Applying rule [4], we obtain:

$$\neg B \rightarrow \neg A$$

which translates back into prose as:

> If it is not the case that we have reasons to accept schema (11) for primitive recursive predicates, then it is not the case that we are able to prove the quasi-completeness of any formalization of intuitionistic logic for which the extended Hauptsatz holds.

This just is the contrapositive of Theorem 5.37—hence, it is clearly evident to anyone who considers that theorem and is aware of the intuitionistic validity of contraposition.[13] In addition, this translation is strictly weaker than the translation obtained via application of rule [3] (i.e. the latter intuitionistically entails the former).[14] Hence this seems like a perfectly adequate (and, given

12 It is important to note that the argument does not depend on the result of applying translation rule [3] being false (or failing to be true, etc.) The point is that the result of applying this translation rule does not result in a translation whose truth follows immediately from the theorem in question.

13 The technical term "contraposition" can refer to a number of different (classically equivalent) rules. Here we mean:

$$\Phi \rightarrow \Psi \models \neg\Psi \rightarrow \neg\Phi$$

and not, for example:

$$\neg\Phi \rightarrow \Psi \models \neg\Psi \rightarrow \Phi$$

The latter is, of course, not intuitionistically valid. Thanks are owed to an anonymous referee for suggesting this clarification.

14 Note that it is not the case in general that the translation delivered by rule [3] entails the translation delivered by rule [4]. Hence, the fact that this entailment holds with regard to the results

the options we are considering, the *strongest* adequate) way of translating this "unless" claim into the language of intuitionistic logic.

This example shows that, if we are looking for a uniform rule for translating informal "unless" claims into the formal language of the intuitionist—one that respects their actual usage of "unless"—then translating "unless" as disjunction is unacceptable, and in addition, the strongest possible such rule (at least, amongst the relatively simple rules we are considering here) that applies to all intuitionistic uses of "unless" is rule [4].[15] In order to see that this is not an isolated case, we will look at a few more examples.

Dummett writes the following in the preface to the first edition:

> Intuitionistic mathematics cannot be justified by its purely 'mathematical interest': one subject-matter may differ from another according to the degree of mathematical interest which they have; but a set of principles of mathematical reasoning, diverging in both directions from those usually accepted, is devoid of interest *unless* there is some way of understanding mathematical statements in accordance with which those principles are justified and other principles are not. (1977, ix, emphasis added)

Adopting the disjunctive rule [1], the claim in question becomes:

> For every set of principles diverging from those usually accepted, either it is devoid of interest or there is some way of understanding mathematics in accordance with which those principles are justified and others are not.

Again, as in our first example, this seems (on the intuitionistic understanding of "or") too strong: surely Dummett is not claiming that we have a *method* for determining, of each such system that diverges from classical mathematics, whether it is devoid of interest or it is justified in the way he describes.

Translation rules [3] and [4] both fare better with this example. On translation rule [3] the passage above turns out to be equivalent to:

---

of applying these rules to most of the actual instances of "unless" that occur in *Elements of Intuitionism* is an interesting fact, which we shall return to in the next section.

15 Of course, one could perhaps argue that Dummet is speaking loosely here, or is uncharacteristically misusing the expression, or… [fill in one's favorite *ad hoc* explanation for why this example is atypical]. Presumably, if one allows this strategy, then one can just cherry-pick whatever examples fit one's preconceptions about the intuitionistic meaning of "unless"– a strategy that seems neither methodologically respectable nor likely to be fruitful.

> For every set of principles diverging from those usually accepted, if it is not devoid of interest then there is some way of understanding mathematics in accordance with which those principles are justified and others are not.

And on translation rule [4] it is equivalent to:

> For every set of principles diverging from those usually accepted, if there is no way of understanding mathematics in accordance with which those principles are justified and others are not, then it is devoid of interest.

Note, however, that in this particular example (and like the previous example), the result of applying translation rule [3] in this case is *logically* stronger than the result of applying rule [4] due to the presence of an embedded negation. Let us adopt the following translation manual (somewhat loosely put):

$A(x)$ = Mathematical principles x are of some interest.

$B(x, y) = y$ is a way of understanding mathematical principles x.

Hence, "x is devoid of interest" is "$\neg A(x)$" then the result of applying translation rule [3] is:

$$(\forall x)(\neg\neg A(x) \rightarrow (\exists y)(B(x, y))$$

and the result of applying translation rule [4] is:

$$(\forall x)(\neg(\exists y)(B(x, y)) \rightarrow \neg A(x))$$

Note that the latter formula is (intuitionistically) a logical consequence of the former.

The translation we obtain by applying rule [3] says something like: If it isn't the case that a particular system is devoid of interest, then *there is* (i.e. there is a method by which we can find) a way of understanding its principles such that those principles are justified and others are not. But Dummett's original claim does not seem to imply that there is, for each such system that is not devoid of interest, a corresponding way to find a suitable interpretation of that system. If this is right, then we again have evidence that not only is rule [1] incorrect, but rule [3] (and hence rule [2]) is incorrect as well, since it produces translations that are (intuitionistically) stronger than the informal claims being translated. The translation obtained by applying translation rule [4], however, seems nicely in line with what Dummett actually seems to be saying, evaluated along intuitionistic lines.

Let’s look at another example. In his discussion of the failure of the least number principle, Dummett writes that:

> We should note, however, that the least number principle:
>
> $$\exists x A(x) \rightarrow \exists x (A(x) \wedge \forall y_{y<x} \neg A(y))$$
>
> is *not* intuitionistically valid: *unless* $A(x)$ happens to be decidable, the fact that we can find a definite number n of which we can prove that it satisfies $A(x)$ is no guarantee that we can find any number m satisfying $A(x)$ of which we can show that no smaller number satisfies it. (1977, 23, emphasis added)

Applying translation rule [1] (and reading a bit into what kind of guarantee Dummett has in mind), this claim is equivalent to:

> For any predicate $A(x)$, either $A(x)$ is decidable, or the fact that there is an x such that $A(x)$ is (on its own) no guarantee that there is a least x such that $A(x)$.

This, again, seems to be too strong, since it implies that, for any predicate $A(x)$, we have some method for determining either that $A(x)$ is decidable or that there is no guarantee that the least number principle holds for $A(x)$.[16]

Applying translation rule [3], the quotation in question turns out to be equivalent to:

> For any predicate $A(x)$, if it is not the case that:
>
> the existence of an x such that $A(x)$ is (on its own) no guarantee that there is a least x such that $A(x)$,
>
> then $A(x)$ is decidable.

There doesn’t seem to be any obvious reason to think that this claim is even true: the fact that we can refute the claim that there is no guarantee of the relevant sort seems to fall far short of being able to determine that $A(x)$ is decidable.

16 Another way of putting the worry is this: The result of applying rule [1] to this example seems to imply that whether or not $A(x)$ is decidable, for arbitrary (arithmetical) $A(x)$, is itself decidable.

doi: 10.48106/dial.v74.i2

Translation rule [4], however, makes the original quotation equivalent to something like:

> For any predicate $A(x)$, if $A(x)$ is not decidable, then the existence of an x such that $A(x)$ is (on its own) no guarantee that there is a least x such that $A(x)$.

This, unlike the result of applying rule [1] or rule [3], seems to capture exactly what Dummett's original "unless" claim was meant to express. In addition, note that, once again, the result of applying rule [3] entails the result of applying rule [4].

Here's another example. Dummett writes:

> That is not to claim that an understanding of any sentence could exist on its own, without a knowledge of any of the rest of the language: every sentence is composed of words or signs which could not be understood *unless* it were known how to use them in at least some other sentences. (1977, 255, emphasis added)

If we adopt translation rule [1], then the sentence at the end of this passage is equivalent to:

> Every sentence is composed of words such that either they cannot be understood or their use in at least some other sentences is known.

Given the intuitionistic understanding of "or," this is clearly too strong, since it implies that, for every sentence, we can decide whether we understand the words contained in it. Translation rule [3] gives us:

> Every sentence is composed of words such that, if it is not the case that they cannot be understood, then their use in at least some other sentences is known.

And translation rule [4] gives us:

> Every sentence is composed of words such that, if it is not the case that their use in at least some other sentences is known, then they cannot be understood.

Note that here (as in all of our other examples), the presence of an embedded negation (along with equating "*x* cannot be understood" with "it is not the case that *x* can be understood") makes it the case that the result of applying translation rule [3] is strictly stronger than the result of applying translation rule [4]—that is, the former logically entails the latter.

Since this passage, unlike our earlier examples, is more informal, our results will be a bit less definitive. Nevertheless, the translation rule [3] result seems odd (to the author at least)—the strange double negation construction in the antecedent does not seem to be part of the content of Dummett's informal claim. The result of applying translation rule [4], however, once again seems to capture exactly what Dummett means (and, if one disagrees with the claim that the result of applying translation rule [4] better captures Dummett's meaning than the result of applying translation rule [3], this does not affect the claim that applying translation rule [1] is just incorrect!)

Let us look at one final example. Dummett writes that:

> The upshot of our review of this second approach is that the status of mathematical objects, as existing independently of us or as the products of our own thought, is irrelevant to whether a classical interpretation of the logical constants is admissible or whether they can be interpreted only in the intuitionistic sense, *unless* the thesis that such objects are the products of our thought it understood in the most radical manner possible, namely as entailing that even primitive predicates (and ones compounded from these by the sentential operators and quantification over a finite domain) are true of them only when we have expressly recognized them to be. To what extent such a radical anti-realism with respect to the objects of mathematics is defensible, and to what extent it is compatible with realism about the contents of the physical universe, are questions left to the reader to think through. (1977, 269, emphasis added)

Applying translation rule [1] implies that the above claim is equivalent to something like:

> Either the status of mathematical objects, as existing independently of us or as the products of our own thought, is irrelevant to whether a classical interpretation of the logical constants is admissible or whether they can be interpreted only in the intuitionistic sense, or

doi: 10.48106/dial.v74.i2

> the thesis that such objects are the products of our thought must be understood in the most radical manner possible, namely as entailing that even primitive predicates (and ones compounded from these by the sentential operators and quantification over a finite domain) are true of them only when we have expressly recognized them to be.

Again, given the particularly strong reading that intuitionists attach to "or," this just seems too strong: Dummett does not seem to be claiming here that we can tell which of the two subclaims holds—in fact, the sentence that follows immediately after the "unless" claim in the original passage seems to suggest just the opposite (whatever we might suspect Dummett's actual views on these matters are).

Applying translation rule [3], we obtain something like:

> If it is not the case that the status of mathematical objects, as existing independently of us or as the products of our own thought, is irrelevant to whether a classical interpretation of the logical constants is admissible or whether they can be interpreted only in the intuitionistic sense, then the thesis that such objects are the products of our thought must be understood in the most radical manner possible, namely as entailing that even primitive predicates (and ones compounded from these by the sentential operators and quantification over a finite domain) are true of them only when we have expressly recognized them to be.

And applying translation rule [4], we obtain something like:

> If it is not the case that the thesis that such objects are the products of our thought is understood in the most radical manner possible, namely as entailing that even primitive predicates (and ones compounded from these by the sentential operators and quantification over a finite domain) are true of them only when we have expressly recognized them to be, then the status of mathematical objects, as existing independently of us or as the products of our own thought, is irrelevant to whether a classical interpretation of the logical constants is admissible or whether they can be interpreted only in the intuitionistic sense.

Again, the translation obtained by applying rule [4] seems more natural than the translation obtained via applying rule [3], although, unlike the earlier cases, I see no definitive reasons for thinking that the result of applying translation rule [3] (or translation rule [2]) in this case gives the *wrong* result.[17]

This concludes our discussion of examples that show that we should reject translation rule [1] and that, in addition, we should favor rule [4] over the rest.[18] Before moving on, however, it is worth noting that there are instances of "unless" in *Elements of Intuitionism* that could, in isolation, be read as (or as equivalent to) disjunctions. For example, in presenting the proof that there are infinitely many logically non-equivalent formulas containing a single sentence letter p, Dummett writes that:[19]

> There are denumerably many non-equivalent formulas with a single sentence-letter p, which form a highly memorable structure. Let us set $P_0 = p \wedge \neg p$, $P_1 = p$, $P_3 = \neg\neg p$, $P_4 = p \vee \neg p$, $P_5 =$

17 Although things are a bit more complicated here, the result of applying rule [3] again seems to entail the result of applying rule [4]. The former has something like:

> If (not: not: Relevant(status of math, interpretation admissible)) then (Understood Radically(thesis that math is thought))

as its logical form, while the latter has something like:

> If (not: Understood Radically(thesis that math is thought)) then (not: Relevant(status of math, interpretation admissible))

18 There are at least two other instances of "unless" in Dummett (1977) that we could consider. The first is on page 299, and the second is on page 305, and both are embedded in complicated bits of reasoning concerning choice sequences. Thus, we have left out explicit discussion of them here, since clarifying the relevant mathematics would take us too far afield and kill too many trees. The reader is encouraged, however, to consider these additional examples, and verify that in both cases translation rule [1] is inappropriate.

19 For an informal example where translation rule [1] seems compatible with the facts, consider:

> If there is a flaw at the heart of classical mathematics, then, even if the intuitionistic reconstruction of mathematics is not correct in every detail, something along those general lines must be right, *unless*, as is surely unthinkable, all but the most elementary parts of arithmetic are delusory. (1977, 250, emphasis added)

There is at least some reason to think, however, that the relative naturalness of reading this passage as an instance of disjunction (in comparison to the cases canvassed above, which cannot be so read) is that the passage is really an explicit assertion of "Φ unless Ψ" and, in addition, an implicit assertion of "it is not the case that Ψ" (indicated by "as is surely unthinkable"). Hence, if we apply translation rule [4], we obtain "$\neg\Psi \rightarrow \Phi$" which, combined with "$\neg\Psi$," entails "Φ," which in turn entails "$\Phi \vee \Psi$."

doi: 10.48106/dial.v74.i2

> $\neg\neg p \to p$, $P_6 = \neg p \lor \neg\neg p$, and, for $n > 2$, $P_{2n+2} = P_{2n-1} \to P_{2n-2}$ and $P_{2n+2} = P_{2n-3} \lor P_{2n-1}$. Then none of the formulas P_n is intuitionistially valid, and every formula with the single sentence-letter p is equivalent to P_n for some n, *unless* it is intuitionistically valid, in which case it is of course equivalent to $p \to p$; [...]. (1977, 21, emphasis added)

Given the fact that intuitionistic propositional logic is decidable, and given the fact that the construction he sketches here provides a method for identifying, for any formula containing only the single sentence letter p, the particular P_n that is its equivalent (for any propositional formula Φ in p, merely apply the decision procedure to $\Phi \leftrightarrow P_0$, then to $\Phi \leftrightarrow P_1$, then to $\Phi \leftrightarrow P_2$, and so on, until one find the true equivalence), the following claim is, in fact, intuitionistically justified:

> For any formula with the single sentence letter p, either it is equivalent to P_n for some n, or it is intuitionistically valid.

The mathematical and logical facts being consistent with this stronger reading no more implies that we should understand this instance of "unless" as a disjunction, any more than a day where the weather alternates between rain and snow implies that we should understand my assertion of:

> It will rain unless it snows.

as equivalent to the conjunction:

> It will rain and it will snow.

Thus, this example in no way throws doubt on the claim that translation rule [1] is too strong.[20]

---

20 An anonymous referee pointed out the following from Brouwer's "Points and Spaces," which was originally published in English:

> [...] the wording of a mathematical theorem has no sense unless it indicates the construction either of an actual mathematical entity or an incompatibility (e.g. the identity of the empty two-ity with an empty unity) out of some constructional condition imposed on a hypothetical mathematical system. (1954, 3)

If we apply rule [1], we obtain something like the following:

3 Some Additional Observations

Before discussing the upshot that the observations made in the previous section have for debates about logic and logical revision, there are two additional issues regarding the proper translation of "unless" that should be dealt with.

First, we should be careful regarding what, exactly, we have shown with regard to translation rule [4]. The sort of evidence presented in the previous section is merely evidence that the *strongest* rule compatible with the evidence in question is rule [4]. Of course, there are presumably good *prima facie* reasons, when translating a natural language expression into a formal language, for taking strongest translation compatible with the evidence (reasons of charity, assumptions of maximal infomativeness, etc.). But these consideration will of course compete with other (themselves *prima facie*) considerations.

One such consideration is worth mentioning here: the strong intuition that "unless" is commutative—that is, the strong intuition that whatever translation rule we adopt, it should support the following equivalence (where L is *whatever* logic we are using):

$$\Phi \text{ unless } \Psi \dashv\vdash_{\mathsf{L}} \Psi \text{ unless } \Phi$$

If we adopt translation rule [4] however, then one obvious result of this is that "unless" claims, in the mouths of intuitionists, will not, in general, be commutative. A nice example of this is given by considering various claims

> Either the wording of a mathematical theorem has no sense or it indicates the construction either of an actual mathematical entity or an incompatibility [...].

This seems stronger than what Brouwer intends here (since it entails that whether or not a theorem is sense-less or indicates an appropriate construction is decidable). Paraphrasing loosely along the lines of rule [3] gives us:

> If the wording of a mathematical theorem fails to have no sense then it indicates the construction either of [...] or [...].

This does not seem obviously too strong, but the presence of the awkward double-negation (which is absent in the original, "unless"-containing sentence) seems odd. A similarly loose application of rule [4] provides:

> If the wording of a mathematical theorem does not indicate the construction either of [...] or [...], then it has no sense.

This seems (to the author, at least) to capture exactly what Brouwer had in mind.

doi: 10.48106/dial.v74.i2

that are classically equivalent to excluded middle but are expressed in terms of "unless," such as:

$$\Phi \text{ unless } \neg\Phi$$
$$\neg\Phi \text{ unless } \Phi$$

Of course, for the classical logician, each of these will be equivalent to excluded middle (and hence a logical truth) regardless of which translation rule they adopt. But, if translation rule [4] is correct, then for the intuitionist these amount, respectively, to:

$$\neg\neg\Phi \rightarrow \Phi$$
$$\neg\Phi \rightarrow \neg\Phi$$

The first is classically but not intuitionistically valid. The second, however, is an intuitionist logical truth. Hence, they are far from being equivalent.

I myself do not have this intuition regarding the commutativity of (intuitionistic) "unless"—on the contrary, as mentioned at the beginning of this essay, I think the fact that the "un" in "Φ unless Ψ" seems (i) to indicate the presence of a negation, and (ii) to attach to "Ψ" but not to "Φ" to be evidence that "Φ" and "Ψ" are not on a par, so to speak, in "Φ unless Ψ."

Nevertheless, the reader who is convinced (for whatever reasons) that "unless" is commutative should not, given the evidence just presented, insist that this means that we should adopt translation rule [1] or translation rule [2], despite the fact that these rules deliver commutative translations of "unless" claims—after all, the examples discussed above show that applying either of these rules results in a translation that is intuitionistically stronger than the informal natural language claim being translated.

In addition, the commutativity-sympathetic intuitionist cannot adopt rule [4], but then stipulate that "unless" is, contrary to what the translation might suggest, commutative. In other words (if one wants to remain an intuitionist of some sort) one should not adopt rule [4] but then use a logic H^* where H^* is intuitionistic logic H plus the following additional rule of inference:

$$\Phi \text{ unless } \Psi \dashv\vdash_{\mathsf{H}*} \Psi \text{ unless } \Phi$$

The reason is simple: adding this rule to intuitionistic logic (combined with rule [4]) just results in classical logic. Let Φ be any formula in our formal language. Clearly $\vDash_{\mathsf{H}*} \neg\Phi \rightarrow \neg\Phi$. But, given rule [4], this is equivalent to $\vdash_{\mathsf{H}*} \neg\Phi$ unless Φ. By our commutativity rule, this gives us $\vdash_{\mathsf{H}*} \Phi$ unless $\neg\Phi$.

Applying translation rule [4] again gives $\vdash_{H*} \neg\neg\Phi \rightarrow \Phi$. Since Φ was arbitrary, it follows that $H^* = C$.

Instead, if one is absolutely committed to the commutativity of "unless"—even in intuitionistic contexts—then the correct response is to adopt translation rule [5]. On this reading, each of the claims of the form "Φ unless Ψ" discussed should be translated as:

$$(\neg\Phi \rightarrow \Psi) \vee (\neg\Psi \rightarrow \Phi)$$

Given the intuitionistic strength of disjunctions, it strikes me – intuitively, at least—that such a translation does some violence to the intended meanings of the passages quoted above. Nevertheless, there is an interesting fact that we need to take into account before putting too much weight on this observation.

In every single one of the examples discussed above (other than the final example, which was compatible with translation rule [1]), the "Φ unless Ψ" claim that we were examining was one where the Φ in question was a negated claim:[21]

- We are *unable* to prove the quasi-completeness of any formalization of HPC for which the Hauptsatz holds, unless [...].
- A set of principles of mathematical reasoning is *devoid* of interest unless [...].
- For any predicate $A(x)$, the fact that there is an x such that $A(x)$ is (on its own) *no* guarantee that there is a least x such that $A(x)$ unless [...].
- Every sentence is composed of words or signs which *could not* be understood unless [...].
- The status of mathematical objects, as existing independently of us or as the products of our own thought, is *irrelevant* to whether a classical interpretation of the logical constants is admissible or whether they can be interpreted only in the intuitionistic sense, unless [...].

In other words, each of these examples is really of the form "$\neg\Phi$ unless Ψ." And, although rule [4] and rule [5] do not deliver logically equivalent translations, they do deliver equivalent translations for cases of this sort, where the expression which is not directly after "unless" is a negated expression. In other words, although $\neg\Psi \rightarrow \Phi$ is not (intuitionistically) logically equivalent to $(\neg\Phi \rightarrow \Psi) \vee (\neg\Psi \rightarrow \Phi)$, $\neg\Psi \rightarrow \neg\Phi$ is (intuitionistically) logically equivalent

21 This is also true of the examples we did not discuss in detail, in Dummett (1977, 250, 299, 305).

doi: 10.48106/dial.v74.i2

to $(\neg\neg\Phi \rightarrow \Psi) \vee (\neg\Psi \rightarrow \neg\Phi)$. As a result, translation rule [5] will fare just as well as a translation of any of the examples discussed above as did translation rule [4].

Thus, if the intuitionist believes they have good reasons to retain the commutativity of "unless," then they can adopt rule [5] rather than rule [4]. As I have noted, I don't see good reasons for thinking that intuitinistic uses of "unless" must be commutative, and I find the translations that result from applying rule [5] to the passages examined in the previous section to be overly complicated, and to do a worse job at capturing Dummett's intended meaning, in comparison to the translations delivered by rule [4]. But for now we can set this aside, since none of the points made in the remainder of this essay depend on rule [4] being correct (or even on rules [2] and [3], much less rule [5], being incorrect): all that is required for the discussion of logical revision in the next section is that rule [1] is definitely incorrect, and nothing said here about commutativity affects our argument for that, much weaker, conclusion.

The second issue is this: why assume that there is a single, univocal, correct translation of "unless" into our formal languages in the first place? Throughout this essay we have assumed that there is such a correct translation rule, and we have then compared and contrasted rules [1] through [6] as candidates for this single, correct rule. But this might be a fallacy. After all, from an intuitionistic standpoint, the following claim:

> For any "$\Sigma(\Phi, \Psi)$" in standard propositional logic, if "$\Sigma(\Phi, \Psi)$" is the *correct* translation of the natural language expression "Φ unless Ψ" then:
>
> "$\Sigma(\Phi, \Psi)$" is no stronger than "$\Phi \vee \Psi$,"
>
> and:
>
> "$\Sigma(\Phi, \Psi)$" is no weaker than "$\neg(\neg\Phi \wedge \neg\Psi)$,"

which we quickly accepted at the very beginning of this essay, does not (intuitionistically) entail that:

> There is a"$\Sigma(\Phi, \Psi)$" in the language of propositional logic such that "$\Sigma(\Phi, \Psi)$" is the single, unique *correct* translation of the natural language expression "Φ unless Ψ."

Restricting our attention to the six competing translations rules we have explicitly discussed in this essay, we can formalize the former claim as something like:

$$(\forall x)(\mathrm{Corr}(x) \rightarrow (x = \mathrm{rule}[1] \vee x = \mathrm{rule}[2] \vee x = \mathrm{rule}[3] \\ \vee x = \mathrm{rule}[4] \vee x = \mathrm{rule}[5] \vee x = \mathrm{rule}[6]))$$

(where "Corr(x)" expresses the claim that x is the correct rule for translating "unless" into our formal language), and we can formalize the latter as:

$$\mathrm{Corr}(\mathrm{rule}[1]) \vee \mathrm{Corr}(\mathrm{rule}[2]) \vee \mathrm{Corr}(\mathrm{rule}[3]) \\ \vee \mathrm{Corr}(\mathrm{rule}[4]) \vee \mathrm{Corr}(\mathrm{rule}[5]) \vee \mathrm{Corr}(\mathrm{rule}[6])$$

The former claim does not intuitionistically entail the latter. In fact, the former claim, plus the additional claim that it is not the case that all six rules fail to be correct—that is:

$$\neg(\neg\mathrm{Corr}(\mathrm{rule}[1]) \wedge \neg\mathrm{Corr}(\mathrm{rule}[2]) \wedge \neg\mathrm{Corr}(\mathrm{rule}[3]) \\ \wedge \neg\mathrm{Corr}(\mathrm{rule}[4]) \wedge \neg\mathrm{Corr}(\mathrm{rule}[5]) \wedge \neg\mathrm{Corr}(\mathrm{rule}[6]))$$

do not jointly entail that one of the six rules must be correct.[22]

To get that conclusion, we need to assume, in addition, that some rule is, in fact, correct—that is, we need to assume:[23]

$$(\exists x)(\mathrm{Corr}(x))$$

But perhaps we should not make this additional, rather substantial assumption. We certainly have not given an argument for this claim here. Perhaps, for example, all we have justification for is the (intuitionistically weaker) claim that it can't be the case that all of rules [1] through [6] fail to be correct. After all, the failure of claims of this form to entail the corresponding disjunctions—that is, the invalidity of the relevant instance of the DeMorgan equivalences—is one of the distinctive features of intuitionistic logic. Maybe there is no single rule that correctly translates all "unless" claims (even when restricting attention to positive contexts), even though every occurrence of "unless" should be translated as no stronger than the result of applying rule [1] (or, given the arguments made above, perhaps rule [2]) and no weaker

22 A sketch of the Kripke model: There are seven worlds $w_0, w_1, w_2, w_3, w_4, w_5, w_6$. The domain of each world is $\{\mathrm{rule}_1, \mathrm{rule}_2, \mathrm{rule}_3, \mathrm{rule}_4, \mathrm{rule}_5, \mathrm{rule}_6\}$. For each n, $0 < n \leq 6$, $R(w_0, w_n)$, and for each n, $0 \leq n \leq 6$, $R(w_n, w_n)$. Corr holds of nothing at w_0, and for each n, $0 < n \leq 6$, $\mathrm{Corr}(\mathrm{rule}_n)$ at w_n.

23 Another way of making the point is that we have, until now, been assuming something like the claim that whether a particular translation rule is correct is decidable.

doi: 10.48106/dial.v74.i2

than the result of applying rule [6] (or, given the arguments made above, perhaps rule [5]).

This is a real issue, and one that deserves more attention. That being said, however, we will set it aside here, and assume for the remainder of this essay that there is a correct translation rule (and that, whatever it turns out to be, it is weaker than rule [1]). Assuming that there is a single correct rule for translating informal intuitionistic assertions containing "unless" into our formal language will simplify the discussion in the remainder of this essay. In addition, I see no reason for thinking that any of the points made below regarding logical revision depend on this assumption, but making this assumption will greatly simplify the making of these points.[24]

4 "Unless" and Logical Revision

So, what is the upshot of all of this? Why does it *matter* how an intuitionist translates "unless," and how such translations might differ from the way classical logicians translate the same bit of natural language? To begin to develop the answer to this question, we again need to think about how logic is taught in introductory formal logic courses, this time with an eye towards the *order* in which various skills are introduced.

In most introductory logic courses, and in most texts on which such courses are based, the topics in question are introduced in roughly the following order:[25]

1. Students are introduced to a particular formal language (e.g. the language of propositional logic).
2. Students are taught how to translate informal natural language sentences and arguments into the formal language, and vice versa.

24 In addition, the close connections drawn by intuitionists between the meaning of expressions and our *manifested* use of those expressions—see Dummett (1975) for a classic source—makes the assumption that there is a unique correct translation rule rather plausible.

25 Of course, in most real-world introductory courses that fit the pattern I have described, the third step involves introducing students to a single account of logic and implicitly assuming (for the sake of the course, at least) that this logic—classical logic—is correct (and hence that the translation rules given in the second step are also correct). But notice that the pattern is the same in textbooks on non-classical logics. See, for example, Sider (2010), where formalization is introduced in chapter 1, long before either classical or non-classical deductive systems or semantics are introduced (in chapters 2 and 3 respectively).

3. Students are taught how to evaluate the sentences and arguments in the formal language (e.g. in terms of logical truth/falsity, validity/invalidity) via either a deductive system or a formal semantics or both.

In short, on the way that formal logic is usually taught, the correct rules for translating natural language sentences and arguments into our formal language is prior to, and hence *must be independent of*, the introduction of the logic via which we shall evaluate those arguments.

Now, from a pedagogical perspective, this might well be the best way to introduce these topics. But once we are engaged in arguments regarding the correct logic, this gets things exactly backwards. As we have seen, the correct translation of "unless" into formal languages depends on which logic one is using—translating "unless" as "or" is perfectly acceptable if one is a classical logician, but is deeply mistaken if one is an intuitionistic logician. And—and this is the rub—this observation has ramifications for how we carry out debates regarding logical revision.

We can flesh out the point by considering a somewhat contrived variant on a classic argument for logical revision due to Hilary Putnam, based on the famous double-slit experiment.[26] In this experiment, photons are projected so that they pass through a plate with two slits cut into it and then collide with a detection screen. When the photons are projected through the plate without any observation regarding the slit through which they passed, the resulting pattern of impacts on the detection screen displays an *interference pattern* associated with wavelike behavior, and seemingly incompatible with each photon having traveled particle-like through exactly one or the other of the slits.

Given this (admittedly rather informal) description of the double-slit experiment, assume that we fire some photons, one-at-a-time, through the apparatus and we observe the expected interference pattern. Then, letting p be any one of the photons, consider the following claims:

1. p impacted the detection screen at location λ, and p passed through the first slit, unless it passed through the second slit.

26 I am merely using this example, and the physics underlying the example, to illustrate the general methodological issue I wish to raise with regard to debates about logical revision. Thus, I will describe the details briefly and somewhat simplistically. Readers interested in more a more careful discussion of Putnam's argument and assessments of its success should consult the extensive literature on this topic, which includes Gardner (1971), Dummett (1976), Gibbins (1987), Hellman (1981), and Maudlin (2005).

doi: 10.48106/dial.v74.i2

2. Either p impacted the detection screen at location λ and passed through the first slit, or p impacted the detection screen at location λ and passed through the second slit.

Putnam (in effect—he of course does not use "unless" in constructing his version of the argument) argues that physics tells us that the first claim is true, and the second claim fails to be true.[27] Let's grant this much. Now, adopting the following translation manual:

$A =_{df}$ p impacted the detection screen at location λ

$B_1 =_{df}$ p passed through the first slit

$B_2 =_{df}$ p passed through the second slit

the classical logician will formalize these claims as (or as something equivalent to):

1. $A \wedge (B_1 \vee B_2)$
2. $(A \wedge B_1) \vee (A \wedge B_2)$

Putnam also points out that the latter follows from the former in any logic L that accepts the following instance of the distributivity rule:

$$\Phi \wedge (\Psi_1 \vee \Psi_2) \vdash_{\mathrm{L}} (\Phi \vee \Psi_1) \vee (\Phi \vee \Psi_2)$$

Now, classical logic accepts the distributivity rule. Thus, if Putnam is right about the physics, then we must abandon classical logic, and replace it with a logic (such as the quantum logic Q that Putnam is endorsing) that (at a minimum) fails to validate this instance of distributivity. Since we agreed, for the sake of the example, to accept that Putnam is right about the physics, so much for classical logic. We must revise.

But what about the intuitionist? After all, the relevant distributivity law is also valid in intuitionistic logic. Does it follow that the intuitionist, like the classical logician, needs to revise their logic, abandoning intuitionistic logic for Q (or perhaps some constructive variant of it)?

27 Note the careful wording. Given that we are comparing classical logic and intuitionistic logic, we need to take care to distinguish between claims that are false and those that (in the relevant intuitionistic sense) merely fail to be true.

By this point it will surely come as no surprise to the reader to discover that the answer is "of course not." The intuitionist has another move available to her at this point. Instead of rejecting distributivity, and intuitionistic logic with it, the intuitionist can instead reject the translation manual used by the classical logician in rendering the informal claims about the physics into formal language. With the points of the previous two sections in mind, she can instead insist that we abandon the faulty rule [1], and instead adopt one of rules [2] through [6] as the proper way to translate "unless" claims. And, although we argued above that [4] (or, perhaps, [5], if one *really* wants commutativity) is the correct rule for translating intuitionistic "unless" claims, it turns out that, in this example, any of rules [2] through [6] will do. Given any of these translation manuals, the translation of the antecedent *does not* entail the translation of the consequent. Since rule [2] provides the strongest translation, it is enough to note that:

$$A \wedge ((\neg B_1 \rightarrow B_2) \wedge (\neg B_2 \rightarrow B_1)) \nvdash_{\mathrm{H}} (A \wedge B_1) \vee (A \wedge B_2)$$

Of course, this is an extremely contrived example.[28] But the lesson we can learn from it is not—it is completely general, and of deep significance, for debates about the correct logic.

Given the way that logic is taught, it is perhaps natural to think that translating informal natural language into formal languages is logic-neutral. As a result, it is tempting to think that the right way to evaluate a purported counterexample to some class of logics (i.e. an argument where the premises are true, the conclusion fails to be true, and the argument is valid according to the logics under consideration) is to first give such a univocal, logic-neutral translation into symbols, and then evaluate the validity of the resulting formal argument pattern with respect to whatever logics are under consideration, rejecting those logics that validate the argument, and accepting one (or perhaps more, if one is a pluralist of some sort) of those that do not. In short, it is natural to accept the following schema—which we shall call the *Flawed Argument for Revising Logic* (or FARL)—as correctly describing much of what goes on in debates about logical revision:

THE (FLAWED) ARGUMENT FOR REVISING LOGIC.

28 It is based upon a far less contrived example. See Cook (2018) for a general discussion of Putnam's example and translation into intuitionistic logic—a discussion that does not depend upon anything particular to "unless." The current essay can be seen as a companion piece to that essay.

(Prem$_1$) We have evidence in favor of accepting natural language claim $\Phi_{\mathcal{NL}}$.

(Prem$_2$) We have evidence in favor of rejecting natural language claim $\Psi_{\mathcal{NL}}$.

(Prem$_3$) Within the context of our current formal logic L_1, $\Phi_{\mathcal{NL}}$ is best translated as Φ_{L_1}.

(Prem$_4$) Within the context of our current formal logic L_1, $\Psi_{\mathcal{NL}}$ is best translated as Ψ_{L_1}.

(Prem$_5$) The argument from Φ_{L_1} to Ψ_{L_1} is valid in our current formal logic L_1, that is:

$$\Phi_{\mathsf{L}_1} \vdash_{\mathsf{L}_1} \Psi_{\mathsf{L}_1}$$

(Conc) We should abandon formal logic L_1 in favor of a weaker (or at least different) logic L_2 where:

$$\Phi_{\mathsf{L}_1} \nvdash_{\mathsf{L}_2} \Psi_{\mathsf{L}_1}$$

But the conclusion does not follow from the premises. After all, why should we think, as is required by the conclusion Conc, that we need to move to a new logic L_2 that does not validate the inference whose premise is the correct translation of $\Phi_{\mathcal{NL}}$, and whose conclusion is the best translation of $\Psi_{\mathcal{NL}}$, *where correctness is understood as relative to our old, now rejected, logic* L_1? Of course, if translation from natural language to formal language were logic-neutral, so that the correct translation of these claims from the perspective of L_1 just was the best translation of these claims from the perspective of L_2, then this wouldn't matter. But, as we now know, translation is not logic neutral. Thus, the conclusion of the argument pattern given above should instead be:

(Conc) We should abandon formal logic L_1 in favor of a weaker (or at least different) logic L_2 where:

$$\Phi_{\mathsf{L}_2} \nvdash_{\mathsf{L}_2} \Psi_{\mathsf{L}_2}$$

(and where Φ_{L_2} and Ψ_{L_2} are the best translations of $\Phi_{\mathcal{NL}}$ and $\Psi_{\mathcal{NL}}$, respectively, from the perspective of L_2.)

Let us call this improved argument pattern, consisting of the premises of FARL and this new conclusion, the *Corrected Argument for Revising Logic* (or CARL).

Thus, if we currently accept a particular logic L, and are then presented with a natural language argument where we accept the premises, we reject the conclusion, and the translation of the premises into our formal logic (where the correctness of the translation is judged from the perspective of our current logic L) entail the translation of the conclusion into our formal language (again, where translation is judged from the perspective of L), then we have not one but two possible strategies:

1. Switch to a logic where the offending inference is no longer valid.
2. Switch to a logic where the correct translations of the premises and conclusion are different.

In our toy example, the logician who rejects classical logic C in favor of quantum logic Q is adopting the first option (assuming that the correct translation of the premise and conclusion is the same from the perspective of C and from the perspective of Q). The classical logician who instead shifts to intuitionistic logic H (or the intuitionist logician who makes no changes to her logic) and rejects the disjunctive translation of the premises is instead adopting the second strategy.

Of course, this is, as I have emphasized repeatedly, a somewhat contrived example.[29] Nevertheless, the lesson it teaches us is deep, and can be summarized as follows:

- A particular counterexample *C* (of the sort described in the premises of FARL or CARL) can show us that a particular logic L must be rejected.
- A particular counterexample *C* (of the sort described in the premises of FARL or CARL) can never, on its own, show us that a particular inferential pattern or rule is invalid.

For any particular inference rule which seems to be challenged by a counterexample in the way that Putnam's quantum logic example seems to challenge the distributivity laws, we are (at least, in principle) free to adopt a logic that retains that rule, as long as, from the perspective of that logic, the correct translation of the premise(s) and conclusion of the purported counterexample no longer instantiate the rule in question. Of course, moving to such a logic, instead of moving to a logic where the inference rule is no longer valid, will not always be the right move, or even a plausible one (for example, it

29 To emphasize: I am not suggesting that the *right* move, for the logician faced with Putnam's purported counterexample, is the one suggested here. Instead, the point is merely that it *is* a move, and, further, there will no doubt be genuine (non-contrived) cases where it is the right move.

would be absurd for someone sympathetic to Dummett-style worries about excluded middle to retain classical logic, but argue that all natural language expressions of the form $\Phi \vee \neg\Phi$ should be translated as a random contingent sentence—e.g. Φ itself). But there will be some cases where this is the right move, and realizing this requires that one recognize that translation from natural language to formal languages (and vice-versa) is not logic-neutral.

5 Conclusion

We'll conclude the paper by explaining its title. First, we can flesh out its content a bit more:

> Unless "$\neg A$ unless A" is invalid, "A unless B" is equivalent to "A or B."

We can now make this more formal along the following lines. For the classical logician applying translation rule [1], this becomes:

$$\text{Either: } \not\models_{C} \neg A \vee A \text{ or: } A \vee B \dashv\vdash_{C} B \vee A$$

The right-hand-side of this disjunction (hence the disjunction as a whole) is obviously classically true. If the arguments given here are correct, however, the intuitionist should apply translation rule [4], and understand this claim as:[30]

$$\text{If not: } \not\models_{H} \neg A \rightarrow \neg A \text{ then: } \neg B \rightarrow A \dashv\vdash_{H} \neg A \rightarrow B$$

Now, the antecedent of this conditional is true, via an intuitionistically valid application of double negation introduction in the metalanguage to obtain:

$$\text{not: } \not\models_{H} \neg A \rightarrow \neg A$$

The consequent of this conditional is clearly false, however. Thus, the conditional as a whole is intuitionistically false.[31]

This brings up a final issue that, again, for the sake of short(ish)ness and snappy(ish)ness, we will only be able to touch on briefly here. There is a

30 Examination of the title of the paper from the perspective of rules [2], [3], [5], and [6] is left to the interested reader.

31 As a result, this is probably the first time I have given a paper a title that I believe (due to my own intuitionistic leanings) is false!

substantial debate within the philosophy of logic concerning what has come to be called the “communication problem”—that is, on determining whether intuitionistic and classical logicians mean the same thing by “and,” “or,” “not,” etc., and are just disagreeing about which claims involving these expressions are valid; or whether they mean different things by these expressions and hence are failing, in some sense, to be disagreeing (or even communicating at all) with each other.[32] I have long been sympathetic to the former understanding, and I am not alone.[33] But the arguments presented above seem to throw some doubt on that understanding of the debate. The difference between the classical and the intuitionistic understanding of the title of this paper does not seem to be merely a difference in the truth value they assign to the claim in question—on the contrary, it seems (at least, intuitively) as if they *mean* different things.

This, in turn, is explained by the fact that the intuitionist and the classical logician cannot both mean the same thing by “or” and mean the same thing by “unless.” Assume for *reductio* that they did. Then, since meaning determines truth conditions, then they would assign the same truth conditions to “or” and to “unless.” But, by the transitivity of sameness of truth conditions (and the fact that the classical logician assigns the same truth conditions to “unless” and to “or”), it should follow that the intuitionist assigns the same truth conditions to “unless” and to “or.” But as we have seen, they do not. Thus, it can’t be the case that intuitionists and classical logicians have a shared set of meanings for all of the logical expressions in natural language. Unfortunately, an in-depth examination of this issue will have to wait for another time.*

Roy T. Cook
0000-0001-7584-9197
University of Minnesota
cookx432@umn.edu

32 For a good discussion of this debate, see Hellman (1989).

33 See e.g. Tennant (1996) for an account of intuitionism that seems to depend on shared meanings.

* Thanks are owed to helpful audiences at the 2017 Workshop on Making It (Too) Precise: Ordinary Reasoning, Formalization, & Logical Modeling at the University of Geneva, Switzerland, and The 2012 Workshop on Logical Constants, Semantic Invariance, & Natural Language at the 4^{th} Indian School on Logic and Its Applications (ISLA) at Manipal University, Manipal, India. Thanks are also due to Geoffrey Hellman, Stewart Shapiro, Jos Uffink, and two anonymous referees for helpful feedback on this or related work. This article was supported by the Research Project 17ZDA024 funded by National Foundation of Social Science, China.

References

Brouwer, Luitzen Egbertus Jan. 1954. "Points and Spaces." *Canadian Journal of Mathematics* 6: 1–17, doi:10.4153/CJM-1954-001-9.

Cook, Roy T. 2018. "Logic, Counterexamples, and Translation." in *Hilary Putnam on Logic and Mathematics*, edited by Geoffrey Hellman and Roy T. Cook, pp. 17–44. Outstanding Contributions to Logic n. 9. Cham: Springer, doi:10.1007/978-3-319-96274-0_3.

Dummett, Michael A. E. 1975. "The Philosophical Basis of Intuitionistic Logic." in *Logic Colloquium '73*, edited by H. E. Rose and John C. Shepherdson, pp. 5–40. Studies in Logic and the Foundations of Mathematics n. 80. Amsterdam: North-Holland Publishing Co. Reprinted in Dummett (1978, 215–247).

—. 1976. "Is Logic Empirical?" in *Contemporary British Philosophy, 4th series*, edited by Hywel David Lewis, pp. 45–68. London: George Allen & Unwin. Reprinted in Dummett (1978, 269–289).

—. 1977. *Elements of Intuitionism*. Oxford Logic Guides n. 2. Oxford: Oxford University Press, doi:10.1093/oso/9780198505242.001.0001.

—. 1978. *Truth and Other Enigmas*. Cambridge, Massachusetts: Harvard University Press.

Gardner, Michael R. 1971. "Is Quantum Logic Really Logic?" *Philosophy of Science* 38(4): 508–529, doi:10.1086/288393.

Gibbins, Peter F. 1987. *Particles and Paradoxes: The Limits of Quantum Logic*. Cambridge: Cambridge University Press, doi:10.1017/cbo9780511570674.

Hellman, Geoffrey. 1981. "Quantum Logic and Meaning." in *PSA 1980: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Part II: Symposium Papers*, edited by Peter D. Asquith and Ronald N. Giere, pp. 493–511. East Lansing, Michigan: Philosophy of Science Association.

—. 1989. "Never say 'Never'! On the Communication Problem between Intuitionism and Classicism." *Philosophical Topics* 17(2): 47–67, doi:10.5840/philtopics19891723.

Higginbotham, James. 1986. "Linguistic Theory and Davidson's Program in Semantics." in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest LePore, pp. 29–48. Oxford: Blackwell Publishers.

Humberstone, I. Lloyd. 2011. *The Connectives*. Cambridge, Massachusetts: The MIT Press, doi:10.7551/mitpress/9055.001.0001.

Kreisel, Georg. 1962. "On Weak Completeness of Intuitionistic Predicate Logic." *The Journal of Symbolic Logic* 27(2): 139–158, doi:10.2307/2964110.

Maudlin, Tim. 2005. "The Tale of Quantum Logic." in, pp. 156–187, doi:10.1017/cbo9780511614187.006.

Putnam, Hilary. 1969. "Is Logic Empirical?" in *Proceedings of the Boston Colloquium for the Philosophy of Science 1966/1968*, edited by Robert S. Cohen and Marx W.

Wartofsky, pp. 216–241. Boston Studies in the Philosophy of Science n. 5. Dordrecht: D. Reidel Publishing Co. Reprinted as "The Logic of Quantum Mechanics" in Putnam (1975, 174–197), doi:10.1007/978-94-010-3381-7_5.

—. 1975. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press. Second edition: Putnam (1979).

—. 1979. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. 2nd ed. Cambridge: Cambridge University Press. First edition: Putnam (1975), doi:10.1017/cbo9780511625268.

Sider, Theodore. 2010. *Logic for Philosophy*. New York: Oxford University Press.

Tennant, Neil W. 1996. "The Law of Excluded Middle Is Synthetic A Priori, If Valid." *Philosophical Topics* 24(1): 205–230, doi:10.5840/philtopics19962416.

—. 2017. *Core Logic*. Oxford: Oxford University Press, doi:10.1093/oso/9780198777892.001.0001.

Assumptions, Hypotheses, and Antecedents

Vladan Djordjević

This paper is about the distinction between arguments and conditionals, and the corresponding distinction between premises and antecedents. I will also propose a further distinction between two different kinds of argument, and, correspondingly, two kinds of premise that I will call "assumption" and "hypothesis." The distinction between assumptions, hypotheses, and antecedents is easily made in artificial languages, and we are already familiar with it from our first logic courses (although not necessarily under those names, since there is no standard terminology for the distinction). After explaining their differences in artificial languages, I will argue that there are ordinary-language counterparts of these three notions, meaning that some formal properties of the artificial notions nicely capture some features of the ordinary-language counterparts and their behavior in contexts of reasoning. My next crucial claim is that these three notions often get confused in ordinary language, which leads to problems for translation into symbols. I will suggest a solution to the translation problem by pointing to some distinctive characteristics of the three notions that link them to their artificial-language counterparts. Next, I will argue that this confusion is behind some well-known philosophical problems and puzzles. I will apply the distinctions in order to explain away some famous paradoxes: the direct argument (also known as or-to-if inference), a standard argument for fatalism, and McGee's counterexample to modus ponens. As Stalnaker also solved the first two of these paradoxes by using his theory of reasonable inference, I will elucidate the similarities between our solutions, and also explain why my distinctions apply more broadly, to some cases involving indicative and counterfactuals conditionals, where reasonable inference does not apply.

Arguments that preserve truth and arguments that preserve validity have different formal properties. Based on that difference, I will consider them as two different kinds of argument and use different names for their premises ("hypotheses" and "assumptions," respectively). I will argue that the distinction between these two kinds is more useful than has been generally recognized, and that we can benefit from it in our attempts to do logic of natural language. I will also consider another old distinction: that between arguments and conditionals. There are thus three things to distinguish—two kinds of argument, and conditionals. Section 1 of this paper is about their distinctive formal properties in artificial languages, especially in classical logic and in standard conditional logics (for indicative and counterfactual conditionals). Section 2 points to a difficulty in translating arguments and conditionals from ordinary language into symbols. The "if ... then ..." construction is common to them, which means that we lack a syntactic mark to distinguish them in ordinary language, and have to find something else to guide our translation. I will suggest a new method of translating. Next, I will claim that our tendency to confuse these three things is behind a number of paradoxes. In particular, the or-to-if problem (also known as the direct argument), a standard argument for fatalism, and McGee's counterexample to modus ponens will be discussed in detail. Other, related issues, such as Kolodny and MacFarlane's rejection of modus ponens, and Yalcin's counterexample to modus tollens, will be briefly mentioned. Using my threefold distinction, I will attempt to explain away these paradoxes. Finally, I will compare my threefold distinction to Stalnaker's twofold distinction between valid and reasonable inference.

1 The Distinction in Artificial Languages

(1) $$\frac{P_1, P_2, \dots P_n}{C}$$

(2) $$\frac{\vdash P_1, \vdash P_2, \dots \vdash P_n}{\vdash C}$$

(3) $$\frac{\vDash P_1, \vDash P_2 \dots \vDash P_n}{\vDash C}$$

(4) $\{P_1, P_2, \dots P_n\} \vdash C$

(5) $\{P_1, P_2, \dots P_n\} \vDash C$

(2) claims that if the premises $P_1, P_2, \dots P_n$ are theorems, then so is the conclusion C. (3) claims that if $P_1, P_2, \dots P_n$ are valid formulae, then so is the conclusion C. (4) says that formula C is a syntactic consequence of the set of formulae $\{P_1, P_2, \dots P_n\}$, i.e. that there is a derivation of C from the set using the rules of inference, or rules and axioms, of our presupposed logical system. (5) says that C is a semantic consequence of the set of formulae $\{P_1, P_2, \dots P_n\}$, meaning that there is no interpretation (valuation, model, etc.) that makes C false and each formula from the set $\{P_1, P_2, \dots P_n\}$ true. The usual meaning of the horizontal line is truth preservation: if whatever occurs above is true, then so is the thing below. This reduces the meaning of (1) to the meaning of (5).

(6) $P_1 \wedge P_2 \wedge \dots P_n \rightarrow C$

(7) $\vdash P_1 \wedge P_2 \wedge \dots P_n \rightarrow C$

(8) $\vDash P_1 \wedge P_2 \wedge \dots P_n \rightarrow C$

(6) is a conditional with the conjunction $P_1 \wedge P_2 \wedge \dots P_n$ as its antecedent and formula C as its consequent. (7) and (8) respectively claim that (6) is a theorem and a valid formula. Among (1)–(8) only (6) is entirely in the object language. (4) and (5) are metaclaims about a relation between a set of formulae and a formula. (2) and (3) are metaclaims about a relation between a set of metaclaims and a metaclaim.

The foregoing should be familiar. Now let me point to a possible terminological confusion. We tend to use the labels “premises” or “conclusion” for the object-language formulae $P_1, P_2, \dots P_n$ and C in all of the above arguments, including (2) and (3). (I did the same above; if you didn’t notice or if it didn’t bother you, then you have the same tendency.) Strictly speaking, this is not right. The premises and the conclusion in (4) and (5) are indeed in the object language, but this is not the case in (2) and (3); what is above and below the horizontal line in (2) and (3) belongs to the metalanguage. Given the usual meaning of the line, (2) (or (3)) says that if it is true that the object-language formulae $P_1, P_2, \dots P_n$ are theorems (valid), then it is true that the object-language formula C is a theorem (valid). If we keep on calling the object-language formulae “premises” or “conclusions” as the case may be, we shall have to change the meaning of the horizontal line in (2) and (3). For, in that case, it could no longer be about truth-preservation, but about theoremhood or validity-preservation. Thus when reading (2) and (3), we have to choose between the following alternatives:

doi: 10.48106/dial.v74.i2

(9) truth-preserving line and premises/conclusions in the metalanguage, or
(10) validity/theoremhood-preserving line and object-language premises/conclusions.

Each of these can be correctly used. (9) is more common, but I will try to show later in this section that (10) may have its own merits.

Definition 1. An *assumption* is an object-language formula used as a premise in an argument of the form (2) or (3).

A *hypothesis* is an object-language formula used as a premise in an argument of the form (4) or (5).

An *argument from assumptions* has the form of (2) or (3).

An *argument from hypotheses* has the form of (4) or (5).

A *conclusion* is the whole object-language formula occurring to the right of the turnstile, or below the line in arguments of the form (2)–(5).

A *single line* is the usual truth-preserving line.

A *double line* does not indicate preservation of truth but preservation of some other special status, such as theoremhood or validity.

Having made these stipulations, I shall now comment on the choice between (9) and (10). Obviously, Definition 1 relies on (10), since all premises are said to belong to the object language. In that case, it is the line that makes the difference between the two types of arguments: whereas arguments from hypotheses claim that the conclusion inherits truth from the premises, arguments from assumptions claim that the conclusion inherits some special modal status from the premises. There is, however, no reason to restrict ourselves to only one kind of line—both are clear and both can be useful. (A third line might be introduced to stand for derivability and capture the meaning of (4), but for my present purposes two will be enough.) So, it would be better to reformulate our dilemma thus:

(11) premises/conclusions sometimes in metalanguage (2, 3) sometimes in object language (1, 5), arguments always truth-preserving, or
(12) premises/conclusions always in object language, arguments sometimes truth-preserving (1, 5), sometimes preserving special status (2, 3).

Choosing (12) over (11) might be preferable for the following reason. We apply names, such as "modus ponens" or "disjunctive syllogism" (and other such names for argument-forms) to both arguments from assumptions and arguments from hypotheses. What identifies arguments (such as modus ponens or disjunctive syllogism etc.) is their form. What identifies the form of an argument is the form of the premises and the conclusion. If this is so, choosing (12) and keeping both kinds of lines from Definition 1 enables us to say that all of the following are instances of modus ponens:

$$\frac{\vDash A, \vDash A \to C}{\vDash C}$$

$$\begin{array}{c} A, A \to C \\ \hline\hline C \end{array}$$

$$\{A, A \to C\} \vDash C$$

$$\frac{A, A \to C}{C}$$

Therefore, choosing (12) over (11) enables us to talk about different kinds of argument having the same form.

Note that this fits our informal practice in logic; although by "modus ponens" we usually mean an argument from hypotheses, we often say, for example, that the Hilbert-style axiomatization of propositional logic uses modus ponens as a rule of inference.[1] That rule (called the "rule of implication" by Hilbert and Ackermann 1950, 28) is an argument from assumptions: it says that if both a material implication and its antecedent are theorems, then so is its consequent.

Now I would like to point to certain formal properties of assumptions, hypotheses and antecedents, and I will do that in the following subsections. Before that, I will limit the types of logical systems I have in mind. Although my claims will hold for many more systems, it will be easier if we restrict our attention to a limited number. Because of the nature of the paradoxes that will be discussed in this paper, my main concern is with conditional logics, i.e. logics for indicative and counterfactual conditionals. What we might call a "typical" or "standard" conditional logic is based on some modal logic, which in turn is based on classical propositional logic (*PL*). Not any modal logic will do. The box will need to have some formal properties that capture enough features of (meta)physical or logical necessity, so usually some alethic normal modal system is used, such as *T* or *S*5, or some system between the two. Adding the so-called selection function to such a modal system gives us a typical conditional logic. The role of that function is to select desired possible

1 Here is a citation from a randomly chosen text that mentions Hilbert axiomatization: "The sole rule of a standard Hilbert axiomatics is *modus ponens*, from $\vdash A$ and $\vdash A \supset B$ to $\vdash B$" (Urbas 1996, 443).

doi: 10.48106/dial.v74.i2

worlds needed for evaluating the truth value of a conditional: $A \rightarrow C$ is true at a world α iff C is true in all of the selected worlds where A is true.[2]

Unless explicitly stated otherwise, from now on, our presupposed logical systems are PL, a modal logic based on PL, such as T, $S5$, or a system stronger than T and weaker than $S5$, and the "typical" conditional logic based on such a modal logic.

1.1 *Differences between Hypotheses and Antecedents*

Arguments and conditionals are similar. We can use "if … then …" to express either when we talk informally. However, accepting the truth of a conditional and accepting an argument are different things, like particular and universal claims. Let M be a model, or an interpretation, or a world, or a valuation. Then $M \vDash A \rightarrow C$ claims that $A \rightarrow C$ is true relative to M, while an argument with A as premise and C as conclusion is acceptable/valid if and only if there is no counterexample in any possible model (interpretation/world/valuation). Thus, we have an obvious difference between a true conditional and its corresponding argument. The validity of an argument with hypothesis A and conclusion C entails the truth of $A \rightarrow C$, but not the other way around. Conditionals can be true necessarily or contingently. Arguments are valid necessarily or not at all.

In cases where a conditional is valid, or is a theorem, the main thing that reveals the differences or similarities between conditionals and corresponding arguments and between premises and antecedents is the deduction theorem. (13) and (14) below give us the form of the theorem in the case of material implication ("⊃").

(13) If $\{P_1, P_2, \ldots, P_n\} \vdash C$ then $\{P_1, P_2 \ldots P_{n-1}\} \vdash P_n \supset C$

(14) If $\{P_1, P_2, \ldots, P_n\} \vdash C$ then $\{P_1, P_2 \ldots P_{n-1}\} \vDash P_n \supset C$

(13) and (14) are metatheorems of PL, and so is the converse of each. Before considering more general cases, let us first take $n = 1$ to compare arguments with one premise and corresponding conditionals. In the case of material implication, it is easy to pass from proven implications to arguments, and conversely:

2 Such semantics is usually called "Stalnaker-Lewis" or "standard," since it shares the main elements of the theories presented in Stalnaker (1968) and Lewis (1973, 1979b).

(15) $\{A\} \vdash C$ iff $\vdash A \supset C$, and
(16) $\{A\} \vDash C$ iff $\vDash A \supset C$

Thus, the deduction theorem and its converse inform us about the relation between antecedents of proven/valid material implications and *hypotheses*, a relation that does *not* hold between antecedents of proven/valid material implications and *assumptions*. For example, the rule of necessitation allows us to infer $\vdash \Box A$ from $\vdash A$, but $\vdash A \supset \Box A$ does not hold. Therefore, there is no significant difference between antecedents and hypotheses in (15) and (16), but there is still a significant difference between antecedents and assumptions.

The typical conditional logic defines a conditional that is stronger than material implication and weaker than strict implication, in this sense (the arrow stands for the conditional):

(17) $\vDash \Box(A \supset C) \supset (A \to C)$ and $\vDash (A \to C) \supset (A \supset C)$

The converse of (17) is not valid, i.e. the conditional does not follow from the material implication, nor does it entail the strict implication. Using (17) and the deduction theorem and its converse for "$\supset$" we can prove that an analogue of (15) and (16) holds for the conditional as well:

(18) $\{A\} \vdash C$ iff $\vdash A \to C$, and
(19) $\{A\} \vDash C$ iff $\vDash A \to C$

Thus again, there is no significant difference between the antecedents of a valid/proven conditional and the corresponding hypothesis in (18) and (19). There is still the same important difference between assumptions and antecedents of conditionals, for the same reason.

So far, we have considered cases where the number of premises $n = 1$. For an arbitrary number of premises things get more complicated, since the deduction theorem for the conditional can easily fail. Consider:

(20)

a	$\{\neg A \vee C, A\} \vDash C$	from *PL*
b	$\{\neg A \vee C\} \vDash A \to C$	from *a* by the deduction theorem for $\to$
c	$\vDash \neg A \vee C \supset (A \to C)$	from *b* by the deduction theorem for $\supset$
d	$\vDash (A \supset C) \supset (A \to C)$	from *c* by *PL*
e	$\vDash (A \to C) \supset (A \supset C)$	from 17

f	$\vDash (A \rightarrow C) \equiv (A \supset C)$	from *d* and *e* by *PL*

(20.*f*) reduces the arrow to the horseshoe and must be rejected if we want to keep the difference between the two connectives. Step (20.*e*) amounts to the claim that modus ponens is valid for the conditional. If we assume that a conditional is not a conditional without modus ponens, then (20.*e*) cannot be rejected. Rejecting any other step beside (20.*b*) would require a change in the basic (propositional or modal) logic. So, the smallest price is to reject (20.*b*).

The converse of the deduction theorem amounts to the claim that modus ponens holds for the implication or conditional in question. Since modus ponens is considered to hold trivially in typical conditional logics, so does the metatheorem that claims that modus ponens holds. Therefore, the converse of the deduction theorem holds for both horseshoe and arrow. However, since the deduction theorem for "→" does not generally hold, relations between arguments and conditionals differ from the relations between arguments and material implications. We can see that hypotheses move easily around the turnstile in the case of material implication:

$$\{P_1, P_2, \ldots, P_m, \ldots, P_n\} \vDash C$$

if and only if

$$\{P_1, P_2, \ldots, P_m\} \vDash (P_{m+1} \supset (P_{m+2} \supset \ldots (P_n \supset C) \ldots))$$

if and only if

$$\vDash (P_1 \supset (P_2 \supset \ldots (P_n \supset C) \ldots))$$

if and only if

$$\vDash P_1 \wedge P_2 \wedge \ldots P_n \supset C$$

But, if we replace "⊃" with "→," the two middle elements in this chain of equivalences have to be dropped so that only two remain:

$$\{P_1, P_2, \ldots, P_m, \ldots, P_n\} \vDash C$$

if and only if

$$\vDash P_1 \wedge P_2 \wedge \ldots \wedge P_n \rightarrow C$$

The reason for this is that whereas exportation and importation are valid for material implication, exportation is invalid for conditionals:[3]

$$\{A \to (B \to C)\} \vDash A \land B \to C \quad \text{(imp.)}$$

$$\{A \land B \to C\} \nvDash A \to (B \to C) \quad \text{(exp.)}$$

Because of this the material implication easily allows nesting in the consequent, while nesting is often problematic for conditionals. We can use our previous example to illustrate that:

$\{\neg A \lor C, A\} \vDash C$	$\{\neg A \lor C, A\} \vDash C$
$\{\neg A \lor C\} \nvDash A \to C$	$\{\neg A \lor C\} \vDash A \supset C$
$\nvDash \neg A \lor C \to (A \to C)$	$\vDash \neg A \lor C \supset (A \supset C)$
$\vDash ((\neg A \lor C) \land A) \to C$	$\vDash ((\neg A \lor C) \land A) \supset C$

Let me summarize this subsection. What is the difference between accepting a conditional and accepting an argument? We can understand this question in two ways: (a) What is the difference between accepting the *truth* of a conditional and the validity of an argument? Or (b) What is the difference between accepting the *validity* of a conditional and the validity of an argument? Let us answer first for the case of simple antecedents, i.e. arguments with only one hypothesis, and leave the more general case for later. (ad a) The validity of an argument with A as hypothesis and C as conclusion is sufficient for the truth of $A \to C$. The truth of $A \to C$ can be context-dependent and contingent, and is therefore not sufficient for the validity of the argument. (ad b) But the argument is valid if and only if the conditional is valid. Thus, in this case, the difference between antecedents and hypotheses (conditionals and arguments) is not significant. This would *not* hold if A were an assumption instead of a hypothesis. In more general cases, when we have more than one hypothesis, things are more complicated. Hypotheses cannot become antecedents by moving right from the turnstile, since the deduction theorem does not hold for conditionals. Since the converse of the deduction theorem holds, antecedents

3 When brackets are omitted, a formula is an implication or equivalence rather than a conjunction or disjunction. So "$A \land B \to C$" means "$(A \land B) \to C$."

Exportation is considered invalid because adding it to standard conditional logic causes a collapse into classical logic, i.e. that would make the arrow the same as the horseshoe. A proof can be seen in McGee (1985, 465–466). See also his footnote 7 where he relates this proof to the failure of the deduction theorem. Gibbard (1981, 234 and further) proved similar results in a different way. Unlike McGee, Gibbard did not go on to deny the validity of modus ponens.

doi: 10.48106/dial.v74.i2

can become hypotheses by moving left from the turnstile. Hypotheses can become antecedents only all at once, i.e. if the antecedent is a conjunction of all the hypotheses, and an empty set remains on the left of the turnstile.

1.2 *The Distinction between Assumptions and Hypotheses*

The decision to regard both assumptions and hypotheses as object language formulae allows us to talk about the same argument-forms for different types of argument. It also makes sense of claims like the following: "conclusion C follows from A if A is taken as an assumption, but not if A is a hypothesis"; "this form is valid for arguments from hypotheses, but not for arguments from assumptions." Often an argument-form is valid for both kinds; modus ponens, for example. Our main interest in this section is to show some forms that hold only for one kind.

1.2.1 Inferences Both Ways

The claim that two formulae are equivalent is usually expressed in symbols with a turnstile and a material biconditional: $\vDash A \equiv B$ or $\vdash A \equiv B$. Such an equivalence can also serve as a definition of one of the formulae, A or B. For later purposes, it is important to notice that if two formulae can be inferred from each other, this double inference does not always amount to equivalence.

(21) $\dfrac{\vDash A}{\vDash B}$ and $\dfrac{\vDash B}{\vDash A}$

(22) $\{A\} \vDash B$ and $\{B\} \vDash A$

(23) $\vDash A \equiv B$

From (22) we can infer (23). We just need to apply the deduction theorem to (22):

(24) $\vDash A \supset B$ and $\vDash B \supset A$

(24) follows from (22), and (23) follows from (24).

However, (23) does not follow from (21). Inferences both ways from assumptions do not amount to equivalence. Consider:

(25) $\dfrac{\vDash A}{\vDash \Box A}$ and $\dfrac{\vDash \Box A}{\vDash A}$

(26) $\vDash A \equiv \Box A$

(25) is valid, but (26) is not.

1.2.2 Validity of some Standard Rules (transitivity, contraposition, constructive dilemma)

In conditional logics, arguments from *hypotheses* in the form of transitivity (hypothetical syllogism) and contraposition typically fail:

$$\{A \to B, B \to C\} \nvDash A \to C$$

$$\{A \to C\} \nvDash \neg C \to \neg A$$

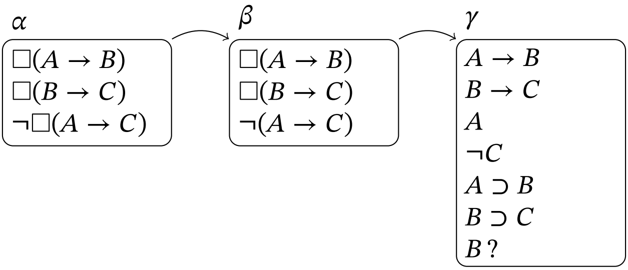
We will show that these forms hold for arguments from *assumptions*. In these proofs, we will make several suppositions about conditionals, but these suppositions are all "safe," i.e. they trivially hold in standard conditional logics. We will suppose that the converse of the deduction theorem and modus ponens hold for →, and that strict implication entails conditional (17); also, we suppose the standard truth conditions: a conditional is true in a world iff the consequent holds in all selected antecedent-worlds. We will also require that these conditions imply that if a conditional is false in a world, then there must be an accessible world where the antecedent is true and the consequent false. Below are the syntactic and semantic versions of the proof of the transitivity of "→":

(27)

a	$\vdash A \to B$	assumption
b	$\vdash B \to C$	assumption
c	$\{A\} \vdash B$	from *a* by the converse of the deduction theorem
d	$\{A\} \vdash B \to C$	from *b* by *PL* (monotonicity)
e	$\{A\} \vdash C$	from *c* and *d* by modus ponens
f	$\vdash A \supset C$	from *e* by the deduction theorem for ⊃
g	$\vdash \Box(A \supset C)$	from *f* by necessitation
h	$\vdash A \to C$	from g and 17 by modus ponens

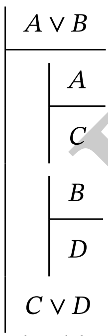
Now the semantic version of transitivity. A countermodel cannot be made:

doi: 10.48106/dial.v74.i2



The negated necessity in α requires the existence of an accessible world (say, β) where the proposition which is not necessary in α is false. The false conditional in β requires the existence of an accessible world (γ) where the antecedent is true and the consequent false. In γ the two conditionals hold (since they are necessary in a world from which γ is accessible), and they entail the two material implications (17). But then γ is an impossible world.

Thus, transitivity as an argument from assumptions holds for conditionals, and we can similarly show that contraposition holds too. However, constructive dilemma, which is a valid form for arguments from hypotheses, fails for arguments from assumptions. Consider constructive dilemma in the way it is presented in Fitch-style systems of natural deduction:



This rule, like the other introduction and elimination rules for each connective in natural deduction systems, is an argument from hypotheses. The assumption-version of constructive dilemma would require both sub-arguments and the main argument to be from assumptions. It might be more convenient to present the two kinds of argument Gentzen-style. So, the constructive dilemma as an argument from hypotheses looks like this:

$$\dfrac{A \lor B \qquad \dfrac{A}{C} \qquad \dfrac{B}{D}}{C \lor D}$$

or:

$$\dfrac{A \lor B \qquad \{A\} \vdash C \qquad \{B\} \vdash D}{C \lor D}$$

The constructive dilemma as an argument from assumptions looks like this (the turnstiles may be replaced by single turnstiles for a syntactic version):

$$\dfrac{\models A \lor B \qquad \dfrac{\models A}{\models C} \qquad \dfrac{\models B}{\models D}}{\models C \lor D}$$

or, more conveniently, using the double line:

$$\begin{array}{c} A \lor B \qquad \begin{array}{c} A \\ \hline\hline C \end{array} \qquad \begin{array}{c} B \\ \hline\hline D \end{array} \\ \hline\hline C \lor D \end{array}$$

Let us take $\neg A$ for B, $\Box A$ for C, and $\Box\neg A$ for D, and let us consider these two arguments:

$$\dfrac{A \lor \neg A \qquad \dfrac{A}{\Box A} \qquad \dfrac{\neg A}{\Box\neg A}}{\Box A \lor \Box\neg A}$$

$$\begin{array}{c} A \lor \neg A \qquad \begin{array}{c} A \\ \hline\hline \Box A \end{array} \qquad \begin{array}{c} \neg A \\ \hline\hline \Box\neg A \end{array} \\ \hline\hline \Box A \lor \Box\neg A \end{array}$$

From "It does or it does not rain" we should not be able to infer "It either necessarily rains or it necessarily does not rain." The two arguments fail for different reasons. The former, the argument from hypotheses, has a valid form but the sub-arguments are invalid. The latter, the argument from assumptions, has valid sub-arguments but an invalid form.

doi: 10.48106/dial.v74.i2

	necessitation	inference both ways gives equivalence	transitivity	contraposition	constructive dilemma
arguments from hypotheses	×	✓	×	×	✓
arguments from assumptions	✓	×	✓	✓	×

Let us also mention some cases where the two types of arguments match:

	modus ponens	modus tollens	importation	exportation
arguments from hypotheses	✓	✓	✓	×
arguments from assumptions	✓	✓	✓	×

2 Translation from Ordinary Language into Symbols

In this section, we turn from formal to natural language and look for counterparts of our three notions. We face an immediate difficulty. In formal language, we had no difficulty recognizing and distinguishing antecedents from premises, or conditionals from arguments. It was enough to be familiar with the syntax of the formal language. However, in natural language we do not have distinctive syntactic characteristics of conditionals and arguments because we often use "if … then …" for both. Rarely do we have conditionals and arguments expressed in an explicit form which tells us that it is one and not the other. Thus, we have a problem when we want to translate our if-constructions into symbols: when and why are we to translate them as conditionals, and when and why are we to translate them as arguments? How can we deal with this problem? Suppose we had a good/acceptable/not-obviously-false/adequate/true/ultimate theory of conditionals, i.e. a formal semantics. Such a theory would be an obvious candidate for a translation guide: it would tell us about the formal characteristics of conditionals, on the one hand, and arguments, on the other, and it would reveal how these differ (similar to what I tried to do in section 1). With these differences in mind, we would do our best to choose a charitable translation that makes the most sense in the given context.[4]

Let us pretend that the standard theory of conditionals, as outlined in section 1, is our theory of choice. Let us bear in mind that it is at best an outline of a theory, with huge gaps to be filled and lots of formal and informal work left to be done, and that this work must include pragmatics if we are to understand our usage of conditionals and to be able to evaluate our semantics. The outline is compatible with many formal semantics that have been proposed—some of those being very weak (in the sense that few rules involving conditionals hold), like Gabbay (1972), some being considered strong, like Stalnaker (1968). There is a chance that the reader's favorite theory might be among them. So let us pretend that we accept the standard theory sketched in section 1, and with it everything said about the formal properties and differences of conditionals and the two kinds of argument. These formal properties will be our guide in translation from ordinary language to symbols, as I suggested in the previous paragraph.

However, we need more things to guide us. We need some characteristics of the ordinary language conditionals and arguments that would link them to their symbolic counterparts. These characteristics are the main topic of this section. I believe that an adequate theory of conditionals (based on the outlined theory we pretend to accept) would imply that antecedents, hypotheses, and assumptions have the following characteristics that I list under the label:

> Thesis
> 2.1. The *antecedent* of a true indicative (counterfactual) conditional is (would be), in the given context, a sufficient condition for the truth of the consequent.

4 Here is some evidence, from randomly chosen academic literature, that "if … then …" is used for both conditionals and arguments. It is enough to show examples of arguments stated in terms of "if … then …". "Modus ponens says that if P is true, and if P implies Q, then Q must be true" (Dretske 2005, 28). "Existential generalization says that if we have found a particular object satisfying some property, then we can assert that there exists an object satisfying that property" (Wolf 2005, 20). "[…] [M]odus ponens says that if you know that p is true, and you also know that whenever p is true q is true, then you can give birth to the new baby truth, q" (Fishman 2002, 8). "But *modus tollens* is a rule of logic, too. And *modus tollens* says that if a logically correct argument leads to a false conclusion, then by God (or by Goddess!) something is wrong with the premises" (Koertge 2010, 7). I am not interested if all the details are correct in these citations, but only in the fact that they express arguments in terms of an "if …" form. Inferring from these that, for example, Dretske believed that modus ponens was a conditional would not be a charitable reading.

2.2. The conjunction of *hypotheses* of a valid argument is, in any possible context, a sufficient condition for the conclusion.
2.3. *Assumptions* of a valid argument are premises such that their special status is, in any possible context, a sufficient condition for the same status of the conclusion.

Let me explain these in turn.

My suggestion is to regard antecedents as a kind of sufficient reason for the consequent. The idea is old, but has been abandoned or forgotten. I will offer some inconclusive arguments for the claim.

First, this claim works well when applied to particular cases in the later sections of this paper.

Second, what else are antecedents if not some sufficient reasons? This is not easy to answer. As we said, the syntax of natural language cannot give the full answer as it does not distinguish premises from antecedents. We may find some help from our formal semantics and say that ordinary language antecedents are whatever is best described by the artificial language antecedents. This, however, presupposes that we already have a solution to the translation problem. In order to have a ready answer to the translation problem, a fully (or at least reasonably) developed theory with semantics and pragmatics is needed. However, many of us are still waiting for such a theory, and some are also waiting for the "right" formal semantics, even if they expect to find it within our presupposed outline from section 1.[5] So, since it seems that we currently lack the "right" theory, I suggest a shortcut—namely, to empirically test the Thesis (which I suppose would follow from the "right" theory), and see if it can be helpful to the problem of translation.

Third, the idea is compatible with our outlined theory. As we said, the outline is compatible with many different semantics, and 2.1 is stated in terms vague enough, I think, to be compatible with most of these. The outline assumes a selection function. What does it do? The role of that function is to somehow separate (what a theory takes to be) relevant from irrelevant antecedent-worlds (for each antecedent and each world of evaluation). Part or all of the meaning of "relevant" should be that all propositions that express the sufficient reason (in the given context) hold at each of the relevant antecedent-worlds. Let us use Goodman's old example with the match *m* (Goodman

5 Remember, the outline we agreed to presuppose is only a skeleton, not a particular conditional logic. Cf. Djordjević (2012) about the important differences between various semantics that fit the outline.

1947; cited from Goodman 1983). Let A = "the match m is struck," C = "the match m lights," and let both A and C be false. Let B_1 = "m is dry," B_2 = "m is well-made," B_3 = "oxygen enough is present," and B_4 = "All dry, well-made matches light when struck in the presence of enough oxygen." Let B_{1-4} be true; they describe the "given context" (or some part of it, depending on the chosen theory of conditionals). The conditional "Had m been struck, it would have lit" ($A \to C$) is true in the described situation. The proposition A is, in the given context (which is here described by B_{1-4}), sufficient for the truth of C. Our favorite theory, since it is a sensible theory, selects the relevant A-worlds in such a way that all of B_{1-4} hold at each of them (we are obviously not interested in A-worlds where the match is not properly made, where different natural laws hold, or where matches are being lit by being put in tomato juice). C would hold in each of these worlds, and our theory gives the right truth value of the conditional.

Of course, "sufficient in the given context" works differently for counterfactuals and for indicative conditionals. The latter are epistemic, and the selected A-worlds can be different, either because we use different selection functions or because one function depends on different contextual parameters for the two kinds of conditionals. Suppose we know B_{1-4}, we do not see the match, and have no beliefs about A and C. Then we would accept "If m was struck, then it lit," for the same reasons we have accepted the analogue counterfactual above. However, if we hold the match and see that it never lit, that is, we know $\neg C$, and further have no beliefs about A and B_2 but know B_1, B_3 and B_4, we would reject that indicative conditional (being convinced that no sufficient reason for the lighting could have possibly obtained) and would rather accept a contrary conditional $A \to \neg B_2$, i.e. "If m was struck, then it was not well made." In this case $\neg C$, B_1, B_3 and B_4 would hold in every selected A-world. Also, $\neg C$, B_1, B_3 and B_4 would now determine "the given context," and A is in that context sufficient for $\neg B_2$.[6]

A fourth reason in favor of 2.1 might be this. Sufficient reasons are good for explanations. If asked why conditionals have the truth value they have, the answer may convincingly be cashed in terms of sufficient conditions. For example, why is the counterfactual considered above "Had m been struck, it would have lit" true? We could offer B_{1-4} as explanation (noting that here the antecedent, together with B_{1-4}, is sufficient for the consequent). If asked

6 Similar examples, and the term "epistemic conditionals," were first discussed by Warmbrōd (1981, 1983) and Gibbard (1981).

doi: 10.48106/dial.v74.i2

why the indicative “If *m* was struck, then it was not well made” is true, we could offer $\neg C$, B_1, B_3 and B_4 as explanation. It would be good for our formal semantics if the *truth conditions* were related to *explanations of truth values*. Saying that $A \rightarrow C$ is true because C holds in the selected worlds is not an explanation, unless we know that the selection function can be interpreted as if it picks up the antecedent-worlds where the explanation holds. If we do not know that, or worse, cannot know that, then why use such a selection function? Worse still, if we do know that the explanation cannot hold in all of the selected antecedent-worlds, that would be a good reason to reject the semantics.[7]

However, I am aware that I cannot please everyone. For example, if you prefer a unified theory of conditionals that includes all or most if-constructions, you will not be pleased with my 2.1. In particular, “even if” conditionals certainly do not go well with 2.1. In addition, 2.1 is meant to work primarily for contingent antecedents and consequents. To make things simpler, I will stipulate that a conditional is vacuously true if the antecedent is impossible or the consequent necessary (which accords with standard conditional logic anyway). There have always been philosophers who do not like that, and their number seems to be growing. Still, in spite of different views we might have, hopefully you will find something of interest in my paper. Different approaches to conditionals, or theories of conditionals, may nevertheless agree about a large and important class of conditionals. There is a chance that the conditionals occurring in the paradoxes that I will discuss below belong to such a class and that we agree about them.

Let us now turn to the “special status,” which, according to the Thesis, makes the difference between assumptions and hypotheses. In Definition 1, we mentioned two special statuses of assumptions—validity and theoremhood. Both valid propositions and theorems are necessary, so we may count logical necessity as the third special status preserved by arguments from assumptions. In artificial language, arguments from hypotheses went from premises to conclusion; arguments from assumptions went from the special status of premises to the same status of the conclusion. My suggestion is that there are analogue situations in ordinary language. Sometimes we argue from premises or a premise to conclusion, say from P to C: we suppose P and claim that C follows. Sometimes, however, we do not simply suppose P; we suppose

7 These are not far-fetched possibilities. For such reasons Djordjević (2013) rejects a class of some of the most popular semantics, including Lewis’s.

that P cannot be false. Consequently, our supposition is not P itself but a claim about a modal qualification of P, that is, our supposition is that P has a certain modal status. When we suppose that P cannot be false, we rule out the possibility of $\neg P$, that is, we treat P as if it were necessary. In that case, the result of our inference has to be stronger than C—it has to be that C inherits the same modal status. Because of that, such arguments should be translated into symbols as arguments from assumptions, i.e. as necessity-preserving arguments, not as truth-preserving arguments from hypotheses.

Pragmatics teaches us that in every conversation something is taken for granted[8] and that some possibilities are ignored.[9] I am here especially interested in cases where a contingent proposition is taken for granted, and its negation is ruled out of consideration. This can happen for various reasons. The most obvious case is when we explicitly agree to suppose something, say P. As long as P holds as a supposition, in a smooth conversation we do not call it into question, nor do we consider $\neg P$ as a possibility. For that part of our conversation P is treated as if it were necessary. But P does not need to be stated explicitly in order to be treated as if it were necessary—it could be a presupposition, or a part of the common ground.[10] The negation of P might not belong among what Lewis called relevant possibilities in a conversation. Thus we can say that there are, in ordinary language, propositions whose negation is ignored and which are treated as if they were necessary. So we gain another candidate for the special status that may be preserved by the arguments from assumptions. It is epistemic necessity. The other three (validity, theoremhood, and logical necessity) are more likely to occur in an artificial language, while epistemic necessity is more suitable as a status of ordinary language assumptions.

What is the exact nature of that necessity? What are its formal properties? Can the answer to that question give a full or only partial answer to the next question (which is my main concern here): what are the formal properties of arguments that preserve that kind of necessity? I wish I could answer. These are million-dollar questions, and what I am able to offer here is far from a complete answer. Arguments that preserve different kinds of necessity may share some formal properties (for example, the rule that necessity entails truth is common to logical and physical necessity). Sometimes, they may share all their formal properties (maybe this is the case with logical and metaphysical

8 Cf. for example Stalnaker (2002, 701), Lewis (1979a; 233 in the 1983 reprint).
9 For example Lewis (1979a; 246–247 in the 1983 reprint).
10 In Stalnaker's sense, cf. (1975, 2002).

doi: 10.48106/dial.v74.i2

necessity—the system $S5$ is sometimes said to capture one, sometimes the other of these two senses of necessity).[11] Epistemic necessity might not be a "real" necessity, in a logical or (meta)physical sense. However, in the context of reasoning it might well behave as a "real" necessity. If always or only sometimes, I do not know. But here is what I suggest. Let us assume that the formal properties from the two tables at the end of section 1 are common to all arguments from assumptions that preserve different kinds of special status.[12] Next, when we realize that our ordinary language premise or if-clause is not simply P, but the claim that P has special status, we should translate our argument or if-construction into symbols using arguments from assumptions, not conditionals nor arguments from hypotheses. In general, when translating our if-constructions into symbols, we need to figure out which of 2.1, 2.2, and 2.3 is intended by our if-clause, and translate accordingly. My last suggestion is that we put the previous suggestions to the test. The proof of the pudding is in the eating. So let us test the distinction between assumptions, hypotheses, and antecedents on some paradoxes.

3 Case 1: the Direct Argument

The so-called "horseshoe-analysis" ($\supset$-analysis to be shorter) says that natural-language indicative conditionals are material implications, or that the truth conditions for indicative conditionals are the same as the truth conditions for material implication. This theory has always had its supporters, maybe since the time of Philo, but certainly since the time of Grice,[13] albeit (it seems) as a minority. The Direct Argument (DA), which allegedly supports the $\supset$-analysis, goes like this:

(DA) $A \vee B$ entails $\neg A \rightarrow B$

Stalnaker said this about DA:

> This piece of reasoning—call it the *direct argument*—may seem tedious, but it is surely compelling. Yet, if it is a valid inference, then the indicative conditional conclusion must be logically equivalent to the truth-functional material conditional [... because] the

11 For more details and subtle distinctions about $S5$ necessities see for example Hale (2012).

12 All except necessitation, which might be a bit more complicated. I will comment on it in section 6.

13 Cf. Part I of Grice (1989), especially chapter 4 "Indicative Conditionals."

> argument in the opposite direction—from the indicative conditional to the material conditional—is uncontroversially valid. [...] and *this* conclusion [i.e. the ⊃-analysis] has consequences that are notoriously paradoxical [... and] must be explained away by anyone who wants to defend the thesis that the direct argument is valid. Yet anyone who denies the validity of that argument must explain how an invalid argument can be as compelling as this one seems to be. [...] There are thus two strategies that one may adopt to respond to this puzzle: defend the [⊃-analysis] and explain away the paradoxes of the material implication, or reject the [⊃-analysis] and explain away the force of the direct argument. (1975; cited from Stalnaker 1999, 63. The square brackets have been added to the original.)

Stalnaker adopted the second strategy. I will do the same here, in a different way.

What kind of argument is DA? It is obviously supposed to be an argument from hypothesis in Stalnaker's paper, but let us consider both possibilities—DA as an argument from hypotheses (DAh), and DA as an argument from assumptions (DAa). Let us further note the fact that DAh is invalid in the standard conditional logic, and that DAa is valid. Following what Stalnaker said and implied in his paper,[14] in solving paradoxes, pointing to a mistake is the smaller part of the job. The main part is to explain why it is a mistake and why it has not been noticed. The standard logic already did the smaller part by rejecting DAh. Let us turn to the main part.

If the disjunction is understood as an assumption, i.e. if it *has* to be that either A or B is the case, and the possibility of the disjunction being false is ruled out of consideration, then it has to be that if it is not one disjunct, it is the other. So DAa sounds good. It seems strange to say: "Under the assumption that $A \vee B$, if A is false, maybe $A \vee B$ is false as well ... So it might not be the case that B is true if A is false." The strangeness may be explained by noting that it is a case of making an assumption and canceling it in the same breath. It is usually not done, because it is not clear what would be the purpose of introducing an assumption and immediately giving it up. Of course, in the dynamics of a conversation presuppositions may be introduced for some part

14 In the above citation, and also in (Stalnaker 1999, 74): "[It] is not enough to say that step x is invalid and leave it at that, even if that claim is correct. One must explain why anyone should have thought that it was valid."

doi: 10.48106/dial.v74.i2

of the conversation and then canceled. But we are now discussing the validity of an argument, and we are not interested in the part of the conversation in which our premise has been canceled. Our premise says that we are limited to considering the situations where $A \vee B$ is true, and other possibilities are being ignored. The premise can be canceled, but as long as it holds, we cannot reject the conclusion $\neg A \rightarrow B$, because the antecedent cannot bring into consideration scenarios that are outside of the presupposed limit. In terms of the formal semantics, the assumption ruled out the possible worlds where the disjunction is false, so the selection function cannot select any such world. (If the antecedent does bring in possibilities from beyond the limit, this amounts to canceling the premise, and such cases are irrelevant for evaluating DAa; formally speaking, if the conclusion is evaluated after the premise has been canceled, then the premise and the conclusion are not evaluated in the same model.)

Things are different, however, if the disjunction is understood as a hypothesis. Nothing is presupposed about the modal status of a hypothesis, so there is no limit to possible scenarios (the selection function is not limited to the possible worlds where the hypothesis is true). In considering whether $\neg A \rightarrow B$ follows from the hypothesis $A \vee B$, we might say that our antecedent might point to situations where the disjunction is not true, so it may be false that B is the case if $\neg A$ is. This does not mean that the antecedent cancels the premise (i.e. the premise and the conclusion can be evaluated in the same model). The hypothesis is about the actual situation (or about the situation in whichever the world of evaluation is) and the antecedent may (but need not) be about the actual situation. Therefore, the hypothesis $A \vee B$, even if true, is not sufficient, in every possible context, for $\neg A \rightarrow B$. This might be a justification for considering DAh invalid and DAa valid.

What does this mean for the relation between DA and ⊃-analysis? ⊃-analysis may be represented as a biconditional:

$$\vDash (A \supset B) \equiv (A \rightarrow B)$$

or, which is the same:

$$\vDash (\neg A \vee B) \equiv (A \rightarrow B)$$

or, if we substitute A for $\neg A$ for convenience:

$$(\supset\text{-a}) \quad \vDash (A \vee B) \equiv (\neg A \rightarrow B)$$

We will take ⊃-a as expressing the ⊃-analysis.

⊃-a is a biconditional consisting of two implications:

(28) $\models (A \lor B) \supset (\neg A \to B)$

(29) $\models (\neg A \to B) \supset (A \lor B)$

One half of ⊃-a, (29), is considered trivial (assuming that modus ponens is valid for the arrow). Applying the converse of the deduction theorem to (28) gives us DAh:

(DAh) $\{A \lor B\} \models \neg A \to B$

Therefore, DA is said to support the ⊃-analysis because DAh plus two trivialities (the deduction theorem for ⊃ and (29)) imply ⊃-a.

On the other hand, DAa does not support the ⊃-analysis:

$$\text{(DAa)} \quad \frac{A \lor B}{\neg A \to B}$$

$$\text{(converse DAa)} \quad \frac{\neg A \to B}{A \lor B}$$

Both DAa and its converse are valid, but this two-way inference does not entail the equivalence ⊃-a (as shown in section 1.2.1).

Thus my suggestion is that the DA problem can be explained away by pointing to an equivocation. Arguments from assumptions and arguments from hypotheses can be easily confused in ordinary language. The reason why DA may appear compelling is because it is understood as DAa. In that case, however, DA does not support the ⊃-analysis. It does support the ⊃-analysis only if understood as DAh, which is less compelling (or not at all). Therefore, DA is either not compelling (understood as DAh) or if it is compelling (understood as DAa), then it has nothing to do with ⊃-analysis. When translating DA into symbols we should pay attention to the exact intended meaning of our premise: do we suppose simply $A \lor B$ or do we suppose that anything opposing $A \lor B$ is ruled out of consideration (i.e. that $A \lor B$ must hold)? We should render DA as DAh in the first case, and as DAa in the second.

What did I exactly achieve or plan to achieve here? I have provided reasons for thinking that DAh is not compelling, but I cannot say that I have proved that DAh is invalid. One can hardly expect a conclusive proof of a thing like that. In my view, such basic rules of inference are to be evaluated together

doi: 10.48106/dial.v74.i2

with the comprehensive theories to which they belong. Opposing comprehensive theories, such as those based on the ⊃-analysis and those based on the standard theory outlined above, are to be tested empirically and evaluated according to their overall success. A “proof” of a rule of inference would then be its belonging to a more successful theory. Obviously, I did not say nearly enough to estimate which approach is more successful. So I am not here in the business of proving or disproving the ⊃-analysis. However, I believe that I have scored a point for the standard theories: having noted the fact that DAh is invalid and DAa is valid in standard logics, I argued that such theories have semantic and pragmatic means to justify that fact and to explain away the DA problem (with the aid of my distinctions and Thesis).

That completes what I have to say about the DA problem, as it is usually presented in the literature. I will add just a few words about counterfactuals. DA is said to be a problem for indicative conditionals and not for counterfactuals, because the counterfactual version of DAh is said not to be as compelling as the indicative version, or maybe not compelling at all.[15] I do not know the exact reason for that claim, but here is my guess as to what might be behind it. Analogous to the indicative versions, DAh is invalid and DAa valid for counterfactuals in standard logics. If asked to explain whether this is good or bad for standard logics, I would say that it is good. My explanation would be exactly analogous to the explanation I gave above for the indicative versions. All the details would remain the same. Whence, then, comes the difference in intuitive acceptability of the two versions? A typical indicative has an antecedent that is not known to be true or false. A typical counterfactual points to a counterfactual situation by an antecedent known to be false. For that reason, it might be easier to cancel presuppositions, assumptions, and premises by using a counterfactual than by using an indicative conditional. My guess is that the counterfactual version of DAh appears to be less compelling because its premise looks more easily cancelable by the antecedent of the conclusion, which is why the premise does not seem to ensure the truth of the conclusion.

Whether or not my guess is right, such reasoning is not correct. When evaluating an argument, we are interested in what holds under the premise. There is no point in looking at what holds after the premise has been canceled. In explaining the indicative version, I noted that the premise has not been canceled in either case: neither in the explanation of the validity of DAa nor

15 Counterfactual DAa is presumably more compelling than counterfactual DAh. But counterfactual DAa is rarely considered.

in the explanation of a possible counterexample to DAh. It can happen, of course, in some conversations that a premise gets canceled by the conclusion, but then we do not have a counterexample.

4 Case 2: A Standard Argument for Fatalism

Let us consider what Dummett (1964, 345) called a standard argument for fatalism. Stalnaker, who considered the same argument (1975; see the reprint 1999, 74f), presented it in the form of natural deduction (this means that the main argument and the sub-arguments are from hypotheses):

(30)

a	Killed ∨ ¬Killed	a. I will be killed in the air raid or I won't.
b	│ Killed	b. Suppose I will be killed.
c	│ Precautions → Killed	c. Then I will be killed even if I take precautions.
d	│ Ineffective	d. Therefore, precautions are ineffective.
e	│ ¬Killed	e. Suppose I won't be killed.
f	│ ¬Precautions → ¬Killed	f. Then I won't be killed even if I don't take precautions.
g	│ Unecessary	g. Therefore, precautions are unnecessary.
h	Ineffective ∨ Unnecessary	h. Therefore, precautions are either ineffective or unnecessary.

On the one hand, we feel that the conclusion does not follow. On the other, the argument seems valid. The main argument has the valid form of a constructive dilemma, and the first premise is logically true, so if there is a mistake, it must be in the sub-arguments. Dummett (1964, 346ff) argued that no conditional which allows the steps (30 c) and (30 f) is strong enough to allow the steps (30 d) and (30 g). Thus, he points to an equivocation of two senses of conditionals. According to Stalnaker, even if we accept Dummett's solution, there are more questions to be answered. He argues that the main task is not to point to a mistake committed in the fatalism argument, but to show why anybody would make such a mistake. Had Dummett shown that

doi: 10.48106/dial.v74.i2

there were these two senses of conditionals in ordinary language, that would have been a full solution. Stalnaker, however, does not believe that this could be done. Instead, he proposed a solution in terms of his notion of *reasonable* inference: the argument is invalid because the sub-arguments are invalid (in Stalnaker's semantics for conditionals), since (30 c) and (30 f) are invalid steps. The force of the argument comes from the fact that the sub-arguments are reasonable. The whole argument, however, is not reasonable, since the reasonableness of sub-arguments does not ensure the reasonableness of the inference from (30 a) to (30 h).

I leave the discussion of Stalnaker's reasonable inference for section 6. Here I will offer another solution. Let us first state the relevant facts from the standard conditional logic. Constructive dilemma is valid as an argument from hypotheses and invalid as an argument from assumptions (as we saw in section 1.2.2). Next, this version of *verum ex quodlibet* is not valid in standard conditional logic:

$$\frac{C}{A \to C}$$

(This was to be expected anyway once we have noticed that the deduction theorem for conditionals does not hold: see section 1.1.) We will need a name for this rule, so let us call it *hypothesis ex quodlibet*. On the other hand, the following rule is valid (call it *assumption ex quodlibet*):

$$\begin{array}{c} C \\ \hline\hline A \to C \end{array}$$

(After the assumption rules out all $\neg C$-worlds, the selection function for the conditional has nothing else to select but C-worlds.) For these reasons, the sub-arguments (30 b – 30 d) and (30 e – 30 g) are invalid as arguments from hypotheses, and valid as arguments from assumptions.

In my view, we have here once again a case of equivocation of assumptions with hypotheses. The steps (30 c) and (30 f) are only valid for the case of entailment from assumptions. If we *assume* that I will be killed, then we rule out of consideration any possibility that the opposite might happen; then it follows that I will be killed even if I take precautions. On the other hand, under the assumption that I will not be killed, it must be that it will be so, whatever I do or do not do. However, as we saw in section 1.2.2, constructive dilemma is not valid for arguments from assumptions. That is, although the sub-arguments are valid, the whole argument is not. The whole argument has a valid form as an argument from hypotheses, but then the sub-arguments are invalid. The *hypothesis* (30 b) (Killed) cannot rule out as impossible my survival. Even if it is true, (30 b) is not a sufficient condition in every context for the conditional (30 c). In general, the consequent (as a hypothesis) is not sufficient in every context for the truth of the conditional. In other words, the Thesis accords with the facts about conditional logic we pointed to, that the rule we might call *premise ex quodlibet* is valid for assumptions and invalid for hypotheses:

$$\frac{C}{A \to C}$$

$$\{C\} \not\models A \to C$$

Therefore, my view is that the alleged strength of the argument (30 a – 30 h) for fatalism comes from an equivocation. The sub-arguments might appear valid if understood as arguments from assumptions, and the whole argument looks valid when understood as an argument from hypotheses.

What exactly did I achieve or plan to achieve here? I did not prove that the steps (30 c) and (30 f), i.e. the sub-arguments, are invalid as arguments from hypotheses. I just stated the fact that they already are invalid in standard conditional logic. I also stated the fact that they are valid from assumptions. Then I tried to explain why I think that the theory has pragmatic and semantic means to justify these facts, and hence that it can explain away the paradox. My aim was not to prove or disprove fatalism; my position is not metaphysical, but logical. I argued that the fact that the argument for fatalism is poor, according to our presupposed logic, is to be justified in pragmatic terms, including the distinctions from the Thesis.

One more thing to do here is to compare the indicative and the counterfactual version. Just imagine that the conditionals in the sub-arguments (30 c) and (30 f) are not indicative but counterfactual. Some philosophers might point to what they see as a disanalogy between the two versions and see only one version as paradoxical. The problem may be stated this way. There is a disanalogy between the indicative and the counterfactual version. The indicative version might appear paradoxical, so there is a problem to solve. The counterfactual version does not appear paradoxical, it just appears invalid, so there is nothing to solve. I, however, have claimed to have "solved" both versions, in exactly the same way.

Where does the disanalogy come from? Apparently, it stems from the claim that at least one of the rules, i.e. *hypothesis ex quodlibet* or *assumption ex quodlibet*, is more compelling for indicative than for counterfactual conditionals. Suppose I will be killed. Does it follow that:

(30 c) I will be killed even if I take precautions?

Or, suppose that I was killed. Does it follow that:

(30 c-*cf*) I would have been killed even if I had taken precautions?

While the former might appear okay, the latter is clearly invalid. Or so the objection goes.

In assessing these two arguments, we first need to specify the nature of the supposition "Killed." After all, perhaps we will easily agree that *both* arguments are invalid if the supposition is a hypothesis. Also, *hypothesis ex quodlibet* would make our conditional logic collapse into classical logic, i.e. we would end up with a horseshoe-theory for both counterfactual and indicative conditionals. So, the supposition should be regarded as an assumption. That is, our premise is not only that I will be (was) killed, but also that my survival is ruled out of consideration. Hence we may reformulate the objection as saying that the above indicative instance of *assumption ex quodlibet* is more compelling than the latter counterfactual instance. But why is that so? Or, better, is it so at all?

I do not think it is so. Let us first note that both indicative and counterfactual version of *assumption ex quodlibet* are valid in standard theories. Let us further note that our instance of that rule looks acceptable—both (30 c) and (30 c-*cf*) sound good, given that my survival is out of the question (i.e. given that "Killed" is not a hypothesis but an assumption). I do not see any relevant difference between the indicative and the counterfactual version. They pass or fail together. The fact (discussed at the end of section 3) that counterfactuals, unlike indicative conditionals, are convenient tools for canceling presuppositions is not relevant here. It is true that one may deny (30 c-*cf*) and claim:

Had I taken precautions, I might not have been killed after all!

This might be perfectly rational, but still it is irrelevant to our purpose. This claim cancels our premise ("Killed"). When assessing an argument, we want to know what follows from a premise while it still holds, not after it has been canceled. Thus I think that if one denies that (30 c-*cf*) follows from the assumption "Killed," then one either understands the premise as a hypothesis or does not realize that the premise has been canceled, which in turn may happen only if one forgets that the premise is an assumption and not a hypothesis. So I believe that my solution to the indicative case, if it is any good, solves *mutatis mutandis* the counterfactual case.

5 Case 3: McGee's Counterexample to Modus Ponens

McGee (1985) proposed a counterexample to modus ponens:

> Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:
>
> M_1. If a Republican wins the election, then if it's not Reagan who wins, it will be Anderson.
>
> M_2. A Republican will win the election.
>
> Yet they did not have reason to believe:
>
> M_C. If it's not Reagan who wins, it will be Anderson.

(I have added the labels "M_1," "M_2," "M_C.") Given the background story, we believe M_1 and M_2, and we do not believe M_C because we believe in the conditional with the contrary consequent: If it is not Reagan who wins, it will be Carter. What I see as the main problem, and the point where the strength of the counterexample lies, is the fact that M_1 appears to be not only true but trivially so, even though it has a true antecedent and a false consequent.

In section 3 we talked about the smaller and bigger tasks involved in solving a paradox (finding the mistake and explaining why it is a mistake and why anybody should make it). Standard conditional logic offers the smaller part of

doi: 10.48106/dial.v74.i2

a possible solution: this is not a counterexample to modus ponens because the long premise is not true. It has a true antecedent and a false consequent, so it cannot meet the truth conditions. Now for the main task—why does M_1 appear to be trivially true?

Let us use the Thesis to consider three things—sentence M_1 translated into symbols as a conditional and two kinds of arguments:

(31) $\text{Republican} \rightarrow (\neg\text{Reagan} \rightarrow \text{Anderson})$

(32) $\{\text{Republican}\} \vDash \neg\text{Reagan} \rightarrow \text{Anderson}$

(33) $$\frac{\text{Republican}}{\neg\text{Reagan} \rightarrow \text{Anderson}}$$

The Thesis requires the antecedent of a true conditional to be sufficient, in the given context, for the consequent. In (31) this is *de facto* not the case, since the antecedent is true and the consequent is not. This is a sense in which (31) is false, which in this case may be offered as a justification for the standard truth conditions for conditionals. Since the antecedent is not sufficient for the consequent in the *given* context, it cannot be sufficient in *every* context, so (32) is invalid. On the other hand, the proposition Republican, *as an assumption*, has the strength to rule out of consideration the Democrats and Carter. Once they have been ruled out, the conclusion of (33) is perfectly acceptable (given that a Republican *has* to win, then, of course, it has to be that if it is not one of the two, it is the other). We cannot maintain that Carter will win if Reagan does not, because our assumption made us forget about Carter. Therefore our reason to reject M_C no longer exists. Thus (33) is valid. Again, the proposition Republican, *as an antecedent*, does not have the strength to rule out what opposes it; so, Carter is still in the game and, because of that, the antecedent is not sufficient in (31). My suggestion is that the way to explain away McGee's paradox is to point to a confusion between antecedents and assumptions. M_1, interpreted as (31), is false, and that is why we do not have a counterexample to modus ponens. The reason why M_1 appears to be trivially true is because we understand it as (33).

This completes the solution I propose. I would like to add few more thoughts a) to avoid possible misunderstanding, b) to emphasize the need of introducing the notion of *arguments from assumptions*, and c) to say a few words about how disputes about basic rules of inference could be resolved (this will also help me to explain better my ambitions in this paper).

a) One might object to the claim that there are arguments from assumptions in ordinary language. Why would anybody suppose that a contingent proposition (such as Republican) is necessary? That sounds unreasonable. Even if we grant that a kind of necessity is involved, am I not confusing logical and epistemic necessity? I plead not guilty. When making an assumption (e.g. Republican) we are not making a logical or metaphysical supposition about the modal status of the claim. We do not suppose that, God forbid, the Republicans necessarily win. We temporarily choose some (logical or metaphysical) possibilities as relevant, and rule out others as irrelevant to our conversation. Relevant possibilities are those compatible with our assumption, which amounts to treating the assumption as if it were necessary. This is a phenomenon routinely explained in pragmatics (rather than an unreasonable claim that something contingent is necessary). Also, I am not confusing different kinds of necessity. True, I never explained the exact nature of the necessity involved. But given that different kinds of necessity may share some formal properties, in this paper I test the supposition that the formal properties from the table in section 1 hold for arguments from assumptions (as explained in the last paragraph of section 2).

b) In “Scorekeeping in a Language Game” Lewis (1979a) introduced his notion of accommodation into pragmatics. If participants in a conversation are cooperative (in the Gricean sense), they try to give a chance of truth to what they hear, interpreting it charitably using various accommodations of presuppositions, resolving vagueness, moving the border between relevant and irrelevant possibilities, etc. McGee’s long premise M_1, as mentioned, appears to be not only true, but logically true. As such, it should be among the first candidates for accommodation and charitable reading. It cannot be simply dismissed as false. A good solution of a paradox (and, more generally, a logic of natural language) must find a right balance between being prescriptive and being descriptive. It seems to me that standard conditional logic (without Thesis and my distinctions) might be in trouble here. If interpreted as a conditional, M_1 is false, and I do not see how standard logic might render it true without giving up some of its essential features. One way of interpreting M_1 as true, without modifying the standard logic, might be to claim that the main and the embedded conditional use different selection functions.[16] This means that there is a context switch in the middle of M_1 that is guilty of

16 Based on a conversation with Stalnaker on a similar example, I believe that his solution of the McGee problem would go along these lines.

doi: 10.48106/dial.v74.i2

the mistake. Still, if this is to be a good solution, it should offer a systematic explanation of how and why such switches of the selection function happen. This explanation should provide some kind of justification for the context switch—even if it is a mistake, it is still rational people who make it. The explanation should also account for the spontaneity of the switch in M_1—since M_1 appears to be logically true, there probably must be some rule-governed pragmatic reason for the switch.

Maybe all this can be done, maybe even in a way compatible with my solution. However, instead of proceeding along these lines, I prefer to use my distinctions because they are more generally applicable—they are not limited to cases with embedded conditionals, nor to cases with at least two conditionals occurring, nor do they necessarily involve a context switch. Moreover, I do not believe that every if-construction in ordinary language must at any cost be considered a conditional (it might well be an argument). Therefore, I prefer to explain that there are two possible interpretations, and to make the two senses of M_1 clear, one in which M_1 is to be rejected (as a conditional), and the other in which it is acceptable (as an argument from assumption). Then I propose that confusing the two senses is the mistake that creates the problem. Next I explain why the mistake was easy to make, which is also why the mistake is excusable. Still, an excusable mistake is a mistake, and should be corrected.[17]

c) Even though I try to introduce a new rule for translation of ordinary language into symbols, the position I defend in this paper is rather conservative and traditional. I talk in terms of sufficient reasons and I believe that there are "sacred" basic rules of inference, such as modus ponens and modus tollens, that are constitutive of the meaning of conditionals and cannot be questioned. In that regard, I have a long tradition on my side. That incurs the risk that I might overestimate the strength of my arguments. I try to keep that in mind when considering different theories, especially those which are radically different. McGee was the first to propose a semantics where modus ponens is invalid, but there are more attacks. There are new theories dealing with the interaction between conditionals and modals. Some of these build new semantics for indicative conditionals to accommodate certain conditional

17 The last two paragraphs under b) were supposed to provide an extra reason for the importance of using the notion of argument from assumptions. Another reason might be found in the literature. Leitgeb (2011) offers a solution to a problem in belief revision (discovered by Chalmers and Hájek 2007) in terms of a distinction that, it seems to me, pretty much resembles mine between hypotheses and assumptions.

claims that are considered false by the standard theories. The victim of this approach may be modus ponens (Kolodny and MacFarlane 2010) or modus tollens (Yalcin 2012). How can we resolve the dispute between these new radical theories and the traditional approach?

Some reactions (especially the early ones) to McGee's counterexample tried to find a mistake in his argumentation, attempting to show that he overlooked something or violated some principles that he presumably also accepts or should accept. However, it seems that neither he nor the others just mentioned ever made such a mistake. I do not believe that this dispute can be solved by finding a "mistake" that one side is making. A more useful approach would be first to admit that McGee as well as MacFarlane, Kolodny and Yalcin know very well what they are doing when they oppose standard opinions. They are not working on small details. They are offering a new general approach to conditionals. These approaches are to be compared in the same way as competing scientific theories are compared. They will be eventually accepted or rejected based on their overall success. That is certainly not a matter of finding a "mistake" in some trivial sense.

I believe that I have scored a point for the traditional side. This is because I believe that the distinctions I have defended are applicable to a large field, to many problems that have often been considered separately, problems for which many different unrelated solutions have been proposed. Also, my distinctions are applicable to counterfactuals as well, and some of the paradoxes, formulated originally in terms of indicative conditionals, have their analogous counterfactual versions. The new radical theories have yet to deal with them.[18] (More about counterfactuals in the next section.)

6 Relation to Stalnaker's Reasonable Inference

The first two cases above (direct argument and fatalism) were discussed in Stalnaker's paper "Indicative Conditionals" (1975). My solution has a certain similarity to Stalnaker's solution in terms of his notion of "reasonable infer-

18 Furthermore, my solutions and distinctions are compatible with the traditional approach, and are not compatible with these new theories. This is because it is essential for my approach to keep a clear difference between antecedents and assumptions, and keep the former much weaker than the latter. Antecedents may do lots of things, change context, trigger or cancel presuppositions, introduce new possibilities etc., but they cannot rule out the possibility of what opposes them, as assumptions do. New semantics see antecedents much the same as I see assumptions. But I need another paper to discuss that properly.

ence." In this section I will try to explain where the similarities and differences come from. Comparison to Stalnaker's theory will, I believe, make my position clearer:

> An inference from a sequence of assertions or suppositions (the premises) to an assertion or hypothetical assertion (the conclusion) is *reasonable* just in case, in every context in which the premises could appropriately be asserted or supposed, it is impossible for anyone to accept the premises without committing himself to the conclusion. (1999, 65)

There are several common words in this definition that are actually Stalnaker's technical notions. We need to explain "context," "appropriateness," and "acceptance."

By "context" Stalnaker means those features of context that determine what propositions are expressed by our sentences. The most important feature, he says, is common knowledge, or presumed common knowledge, common ground, or background information that one takes for granted only if one presupposes that other participants in the conversation take it for granted (cf. Stalnaker 1999, 67; 2002, 701). The formal device that represents the common ground is *context set*, a set of worlds not ruled out by the common ground. A proposition is said to be *compatible with* or *entailed by* a context, respectively, when it is true at some or all the worlds from the context set. Contexts can change during our conversation, even by the conversation itself. Any *accepted* assertion changes the context by becoming an additional presupposition of subsequent conversation. That is, accepted assertions express propositions that rule out of the old context set the worlds where they do not hold, and then these propositions hold throughout the new context set. The *appropriateness* condition states that one cannot appropriately assert a proposition in a context incompatible with it. Applied to conditionals, the condition leads to the rule that one can appropriately assert a conditional only if its antecedent is compatible with the context. A typical counterfactual has an antecedent presumed to be false, so the rule is meant for indicative conditionals only.

Stalnaker defines entailment in the usual way: "A set of propositions (premises) *entails* a proposition (the conclusion) just in case it is impossible for the premises to be true without the conclusion being true as well" (1999, 65). Using my terminology, this is the relation between the set of hypotheses and the conclusion. Reasonable inference, on the other hand, corresponds to my arguments from assumptions. The reason for this is that the premises, once asserted and accepted, change the context and hold throughout the resulting context, i.e. they are entailed by the new context. Thus negations of accepted premises become inappropriate; we may say that they are ruled out of consideration. Accordingly, the premises have the status of necessity (relative to the context set), the same status that all other presuppositions from the common ground have. The conclusion of a reasonable argument is then entailed by the context, and it inherits the special status from the accepted premises. Thus, reasonable arguments are about preservation of that special status, not about preservation of truth. Because of that the formal properties of reasonable inference match those of arguments from assumptions, and do not match those of arguments from hypotheses. From Stalnaker's paper we learn that transitivity and contraposition are reasonable (1999, 73) and constructive dilemma is not (1999, 74f). We also learn that the direct argument is reasonable, and it is easy to see that the converse (from conditional to disjunction) is also reasonable (1999, 72f). Therefore, reasonable inference both ways does not amount to equivalence (Stalnaker rejects the ⊃-analysis).

	necessitation	inference both ways gives equivalence	transitivity	contraposition	constructive dilemma
arguments from hypotheses	×	✓	×	×	✓
arguments from assumptions	✓	×	✓	✓	×
reasonable inference	?	×	✓	✓	×

This is the same table from the end of section 1, with one additional row for reasonable inference. The only difference between the last two rows is in the case of necessitation. I put the question mark because both answers are possible, depending on the meaning of the box, i.e. the modal operator. If the box stands for logical necessity, then necessitation is not reasonable. If the box

doi: 10.48106/dial.v74.i2

stands for the epistemic necessity of the same kind that a premise gains by being accepted and becoming part of the common ground, then necessitation is reasonable.

This relation between entailment and arguments from hypotheses on the one side, and reasonable inference and arguments from assumptions on the other, makes Stalnaker's and my solutions to cases 1 and 2 similar. The direct argument is invalid but its strength comes from its being reasonable according to Stalnaker's explanation, while I called it invalid as an argument from hypothesis and explained its alleged strength by pointing to the validity of the corresponding argument from assumption. The fatalism argument has the valid form and invalid sub-arguments, and unreasonable form and reasonable sub-arguments, again analogous to the solution I defended in section 4. Why, then, do I look for new distinctions?

I believe that my distinctions point to a more basic phenomenon and are applicable to more kinds of cases. Solutions in terms of my distinctions match those of Stalnaker's solutions in terms of reasonable inference, but my distinctions apply more broadly, because they are not limited by the appropriateness condition. First, a typical counterfactual has an antecedent presumed to be false, which makes the conditional inappropriate, so the notion of reasonable inference is not meant for this class of conditionals. Second, the notion of reasonable inference cannot be applied to arguments involving indicative conditionals that do not meet the appropriateness condition. For that reason, Stalnaker's notion cannot be used to resolve McGee's case. Reagan's winning may well be a part of the common ground and hold throughout the context set. Reagan's not winning occurs twice in McGee's counterexample, so neither the premises nor the conclusion meets the appropriateness condition.[19]

Consider the McGee case again. Sometime after the elections we could imagine such a conversation:

---

19 There is a possibility that common ground includes Reagan's winning, and it is not a far-fetched one. This is important for my argumentation, and I will try to show it in more detail. We can modify McGee's example by adding some more information. Let the opinion poll results be 69%, 30%, 1% for Reagan, Carter and Anderson, respectively. Imagine a conversation where participants believe that the margin of error is ±3%, which they understand as meaning that the actual voting results cannot differ from the opinion poll results more than 3%. Through several meetings and conversations on similar topics, this belief became part of the common ground for the group. Reagan's winning is entailed by their common ground, so it is part of it.

Another example. I think we will easily agree that there once were or still are conversations where part of the common ground is that Reagan won the 1980 elections. Now consider a past tense version of McGee's example:

A: Had a Republican won, then, had it not been Reagan, it would have been Anderson.

B: Yes, but a Republican did win (you missed the news).

A: So, had Reagan not won, Anderson would have.

Consider also the fatalism case (30 a)–(30 h) again. It pertains to some period and some person. Suppose that a few years later we are presented with this argument, which also pertains to that same person and same period:

a	Killed ∨ ¬Killed
b	Killed (assumption)
c	Precautions → Killed
d	Ineffective
e	¬Killed (assumption)
f	¬Precautions → ¬Killed
g	Unecessary
h	Ineffective ∨ Unnecessary

a. He was killed in the air raid or he was not.
b. Suppose he was killed.
c. Then it would have been so even if he had taken precautions.
d. Therefore, precautions are ineffective.
e. Suppose he was not killed.
f. Then it would have been so even if he had not taken precautions.
g. Therefore, precautions are unnecessary.
h. Therefore, precautions are either ineffective or unnecessary

It is difficult to argue that these examples talk about something different than the original examples, and that these counterfactuals say something different

---

If a Republican won the election, then if it was not Reagan, it was Anderson.
A Republican won.
So, if it was not Reagan who won, it was Anderson.

Here the appropriateness condition would not be met, but the example would pose the same problem as the original version. This version may not usually be properly assertable, but semantics must be able to evaluate it anyway. For example, this might not be what the participants in the conversation are saying to each other, but it could be that they are merely estimating something said or written by another person.

from what was said by the analogous indicative conditionals.[20] Thus, these examples present the same puzzles as the original versions already discussed in previous sections. My solutions to them would be exactly analogous to the solutions I proposed for the indicative versions. For these reasons, I believe that the distinction between antecedents, hypotheses and assumptions is more broadly applicable than Stalnaker's notion of reasonable inference.

This is not a critique of Stalnaker's theory, but a comparison that helps me emphasize and clarify my points. There is no conflict between our solutions—they go along within the appropriateness limit (as in DA and the original fatalism case), and the reason for that match has been explained in this section. In addition, my distinctions apply to some cases involving inappropriate indicative conditionals (which may occur in McGee-style counterexamples) and to some cases involving counterfactuals (like the two past-tense versions of McGee's counterexample and the fatalism argument).

There is another more subtle difference between Stalnaker's solution and mine, and that is a difference in emphasis, stress, or, let us say, accent. It comes from the choice of terminology. There is a positive component of the meaning of the word "reasonable." It suggests something laudatory or commendable. Within the expression "invalid but reasonable" it suggests something justifiable or forgivable. Within my terminology, what is justifiable or forgivable is never the use of an invalid argument. Invalidity is a mistake, and is therefore bad. Justification is to be looked for elsewhere. In Stalnaker's case, an argument, for example DAh, can be invalid and reasonable. In my case, it is not the same argument that is good in one sense and bad in another, but two different arguments: one good and the other bad (for example, DAa and DAh). So, I do not need to say that there is something justifiable in using invalid arguments, i.e. in the mistake itself. We both look for an excusing factor that would explain why the mistake was easy to make (in Stalnaker's case, because the invalid argument may be reasonable; in my case, because assumptions, hypotheses, and antecedents may be hard to distinguish in

20 Similar examples were made by Strawson (1986), from the (1997) reprint, p. 163:

> (1) Remark made in the summer of 1964: "If Goldwater is elected, then the liberals will be dismayed."—(2) Remark made in the winter of 1964: "If Goldwater had been elected, then the liberals would have been dismayed." It seems obvious that about the least attractive thing that one could say about the difference between these two remarks is that it shows that … the expression "if … then …" has a different meaning in one remark from the meaning which it has in the other.

ordinary language). Therefore, whereas for Stalnaker it is the argument stated in the formal language that can be bad and excusable (e.g. DAh), in my case what may be bad and excusable is never an argument expressed in symbols, but the *translation* of ordinary language if-constructions into symbols.*

Vladan Djordjević
University of Belgrade
vladan@ualberta.ca
vladan.djordjevic@gmail.com

References

Chalmers, David J. and Hájek, Alan. 2007. "Ramsey + Moore = God." *Analysis* 67(2): 170–172, doi:10.1111/j.1467-8284.2007.00670.x.

Djordjević, Vladan. 2012. "Goodman's Only World." in *Between Logic and Reality: Modeling Inference, Action and Understanding*, edited by Majda Trobok, Nenad Miščević, and Berislav Žarnić, pp. 269–280. Logic, Epistemology, and the Unity of Science n. 25. Dordrecht: Springer, doi:10.1007/978-94-007-2390-0_15.

—. 2013. "Similarity and Cotenability." *Synthese* 190(4): 681–691, doi:10.1007/s11229-012-0198-4.

Dretske, Fred I. 2005. "The Case against Closure." in *Contemporary Debates in Epistemology*, edited by Ernest Sosa and Matthias Steup, 1st ed., pp. 24–49. Contemporary Debates in Philosophy n. 3. Boston, Massachusetts: Blackwell Publishers. Second edition: Steup, Turri and Sosa (2014, 27–39).

Dummett, Michael A. E. 1964. "Bringing about the Past." *The Philosophical Review* 73(3): 338–359. Reprinted in Dummett (1978, 333–350), doi:10.2307/2183661.

—. 1978. *Truth and Other Enigmas*. Cambridge, Massachusetts: Harvard University Press.

Fishman, Mark B. 2002. "Teaching AI Epistemology to Humans." in *Proceedings of the Thirty-Fourth Annual Meeting of the Florida Section of the Mathematical Association of America*, edited by David Kerr and Bill Rush. Florida Gulf Coast

* This research has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia (project "Logico-epistemological Foundations of Science and Metaphysics," no. 179067).
For very helpful comments and discussions, I thank Milos Arsenijevic, Hanoch Ben-Yami, Nate Charow, Alan Hájek, Andrej Jandric, Maiya Jordan, Chris Lepock, Jelena Ostojic, Majda Trobok, Timothy Williamson, and late Berislav Zarnic. I also thank the audiences at the CEU Summer course "Conditionals: Philosophical and Linguistics Issues" in Budapest, UC Davis colloquium, "Mind, World, and Action" IUC Dubrovnik, 2nd Belgrade Conference on Conditionals, Graz – Belgrade Philosophy Meeting. I am especially indebted to Brian Leahy, Adam Sennet, and two anonymous referees. Thanks to Ljiljana Ognjenovic who helped me proofread the final version.

University: Florida Section of the MAA, http://sections.maa.org/florida/proceedings/2001/fishman.pdf.

Gabbay, Dov M. 1972. "A General Theory of the Conditional in Terms of a Ternary Operator." *Theoria* 38(3): 97–104, doi:10.1111/j.1755-2567.1972.tb00927.x.

Gibbard, Allan F. 1981. "Two Recent Theories of Conditionals." in *Ifs: Conditionals, Belief, Decision, Chance, and Time*, edited by William L. Harper, Robert C. Stalnaker, and Glenn Pearce, pp. 211–247. The University of Western Ontario Series in Philosophy of Science n. 15. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-009-9117-0_10.

Goodman, Nelson. 1947. "The Problem of Counterfactual Conditionals." *The Journal of Philosophy* 44(5): 113–128. Incorporated into Goodman (1955, 13–36), doi:10.4324/9780429495687-23.

—. 1955. *Fact, Fiction and Forecast*. Cambridge, Massachusetts: Harvard University Press.

—. 1983. *Fact, Fiction and Forecast*. 4th ed. Cambridge, Massachusetts: Harvard University Press. First edition: Goodman (1955).

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

Hale, Bob. 2012. "What is Absolute Necessity?" *Philosophia Scientiae. Travaux d'histoire et de philosophie des sciences* 16(2): 117–148, doi:10.4000/philosophiascientiae.743.

Harper, William L., Stalnaker, Robert C. and Pearce, Glenn, eds. 1981. *Ifs: Conditionals, Belief, Decision, Chance, and Time*. The University of Western Ontario Series in Philosophy of Science n. 15. Dordrecht: D. Reidel Publishing Co., doi:10.1007/978-94-009-9117-0.

Hilbert, David and Ackermann, Wilhelm. 1928. *Grundzüge der theoretischen Logik*. Berlin: Springer Verlag.

—. 1938. *Grundzüge der theoretischen Logik*. 2nd ed. Berlin: Springer Verlag. First edition: Hilbert and Ackermann (1928), doi:10.1007/978-3-662-41928-1.

—. 1950. *Principles of Mathematical Logic*. New York: Chelsea Publishing Company. Translation of Hilbert and Ackermann (1938) by Lewis M. Hammond, George G. Leckie and F. Steinhardt, with revisions, corrections and added notes by Robert E. Luce.

Koertge, Noretta. 2010. "The Feminist Critique [Repudiation] of Logic." Unpublished manuscript, https://philpapers.org/rec/KOETFC.

Kolodny, Niko and MacFarlane, John. 2010. "Ifs and Oughts." *The Journal of Philosophy* 107(3): 115–143, doi:10.5840/jphil2010107310.

Leitgeb, Hannes. 2011. "God – Moore = Ramsey (A Reply to Chalmers and Hájek (2007))." *Topoi* 30(1): 47–51, doi:10.1007/s11245-010-9088-x.

Lewis, David. 1973. *Counterfactuals*. Cambridge, Massachusetts: Harvard University Press. Cited after republication as Lewis (2001).

—. 1979a. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8(3): 339–359. Reprinted in Lewis (1983, 233–249), doi:10.1007/BF00258436.

—. 1979b. "Counterfactual Dependence and Time's Arrow." *Noûs* 13(4): 455–476. Reprinted, with a postscript (Lewis 1986b), in Lewis (1986a, 32–51), doi:10.2307/2215339.

—. 1983. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

—. 1986a. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1986b. "Postscript to Lewis (1979b)." in *Philosophical Papers, Volume 2*, pp. 52–66. Oxford: Oxford University Press, doi:10.1093/0195036468.003.0002.

—. 2001. *Counterfactuals*. Oxford: Blackwell Publishers. Republication of Lewis (1973).

McGee, Vann. 1985. "A Counterexample to Modus Ponens." *The Journal of Philosophy* 82(9): 462–471, doi:10.2307/2026276.

Sosa, Ernest and Steup, Matthias, eds. 2005. *Contemporary Debates in Epistemology*. 1st ed. Contemporary Debates in Philosophy n. 3. Boston, Massachusetts: Blackwell Publishers. Second edition: Steup, Turri and Sosa (2014).

Stalnaker, Robert C. 1968. "A Theory of Conditionals." in *Studies in Logical Theory*, edited by Nicholas Rescher, pp. 98–112. American Philosophical Quarterly Monograph Series n. 2. Oxford: Basil Blackwell Publishers. Reprinted in Harper, Stalnaker and Pearce (1981, 41–55).

—. 1975. "Indicative Conditionals." *Philosophia: Philosophical Quarterly of Israel* 5(3): 269–286, doi:10.1007/978-94-009-9117-0_9.

—. 1999. *Context and Content: Essays on Intensionality, Speech and Thought*. Oxford: Oxford University Press, doi:10.1093/0198237073.001.0001.

—. 2002. "Common Ground." *Linguistics and Philosophy* 25(4/5): 701–721, doi:10.1023/a:1020867916902.

Steup, Matthias, Turri, John and Sosa, Ernest, eds. 2014. *Contemporary Debates in Epistemology*. 2nd ed. Contemporary Debates in Philosophy n. 3. Malden, Massachusetts: Wiley-Blackwell. First edition: Sosa and Steup (2005), doi:10.1002/9781394260744.

Strawson, Peter Frederick. 1986. "'If' and '⊃'." in *Philosophical Grounds of Rationality: Intentions, Categories, Ends*, edited by Richard E. Grandy and Richard Warner, pp. 228–242. Oxford: Oxford University Press. Reprinted in Strawson (1997, 162–178).

—, ed. 1997. *Entity and Identity. And Other Essays*. Oxford: Oxford University Press, doi:10.1093/0198250150.001.0001.

Urbas, Igor. 1996. "Dual-Intuitionistic Logic." *Notre Dame Journal of Formal Logic* 37(3): 440–451, doi:10.1305/ndjfl/1039886520.

WARMBRŌD, Ken. 1981. "An Indexical Theory of Conditionals." *Dialogue. Revue canadienne de philosophie / Canadian Philosophical Review* 20(4): 644–664, doi:10.1017/s0012217300021399.

—. 1983. "Epistemic Conditionals." *Pacific Philosophical Quarterly* 64(1): 249–265, doi:10.1111/j.1468-0114.1983.tb00198.x.

WOLF, Robert S. 2005. *A Tour through Mathematical Logic*. The Carus Mathematical Monographs n. 30. Washington, D.C.: Mathematical Association of America.

YALCIN, Seth. 2012. "A Counterexample to Modus Tollens." *Journal of Philosophical Logic* 41(6): 1001–1024, doi:10.1007/s10992-012-9228-4.

Published by *Philosophie.ch*

Verein philosophie.ch
Fabrikgässli 1
2502 Biel/Bienne
Switzerland
dialectica@philosophie.ch

https://dialectica.philosophie.ch/

ISSN 0012-2017

Dialectica is supported by the Swiss Academy of Humanities and Social Sciences.

Abstracting and Indexing Services

The journal is indexed by the Arts and Humanities Citation Index, Current Contents, Current Mathematical Publications, Dietrich's Index Philosophicus, IBZ — Internationale Bibliographie der Geistes- und Sozialwissenschaftlichen Zeitschriftenliteratur, Internationale Bibliographie der Rezensionen Geistes- und Sozialwissenschaftlicher Literatur, Linguistics and Language Behavior Abstracts, Mathematical Reviews, MathSciNet, Periodicals Contents Index, Philosopher's Index, Repertoire Bibliographique de la Philosophie, Russian Academy of Sciences Bibliographies.

Contents