Reconstructing Arguments: Formalization and Reflective Equilibrium

Georg Brun

University of Zurich, Centre for Ethics, Zollikerstrasse 117, 8008 Zurich, Switzerland, Georg.Brun@philos.uzh.ch

Abstract   Traditional logical reconstruction of arguments aims at assessing the validity of ordinary language arguments. It involves several tasks: extracting argumentations from texts, breaking up complex argumentations into individual arguments, framing arguments in standard form, as well as formalizing arguments and showing their validity with the help of a logical formalism. These tasks are guided by a multitude of partly antagonistic goals, they interact in various feedback loops, and they are intertwined with the development of theories of valid inference and adequate formalization. This paper explores how the method of reflective equilibrium can be used for modelling the complexity of such reconstructions and for justifying the various steps involved. The proposed approach is illustrated and tested in a detailed reconstruction of the beginning of Anselm's *De casu diaboli*.

Zusammenfassung   Traditionelle logische Argumentrekonstruktionen zielen darauf ab, die Gültigkeit umgangssprachlicher Argumente zu beurteilen. Das erfordert es, Argumente in Texten zu ermitteln, komplexe Argumentationen in einzelne Argumente zu zerlegen, Argumente in Standardform zu bringen und mit Hilfe logischer Formalismen die Argumente zu formalisieren und deren Gültigkeit nachzuweisen. Diese Aufgaben sind durch eine Reihe von teilweise gegenläufigen Zielen geleitet, sie interagieren in verschiedenen Rückkopplungen, und sie sind mit der Entwicklung von Theorien der logischen Gültigkeit und der adäquaten Formalisierung verwoben. Dieser Artikel untersucht, wie die Methode des Überlegungsgleichgewichts verwendet werden kann, um die Komplexität von Argumentrekonstruktionen zu modellieren und die beteiligten Schritte zu rechtfertigen. Dieser Vorschlag wir in einer exemplarischen Rekonstruktion des Anfangs von Anselms *De casu diaboli* illustriert und getestet.

The well-known logical investigations of Anselm's ontological arguments exemplify a tradition in which logic is a cornerstone of argument analysis and evaluation. Such projects of logical analysis combine informal techniques of analysing argumentative texts with formalizations. Thus, a distinctive feature of this practice is the use of logical formulas, which are assigned to an ordinary language text and employed to scrutinize it logically. The focus of this paper is an analysis of this practice itself. I side with those who argue that logical analysis of arguments and formalization in particular are best understood as forms of reconstruction. This includes Quine's formalization as regimentation (1960, ch. V) as well as the view that formalizing is a form of explication (e.g. Blau 2008, 143, 145; see also Brun 2004, ch. 8.2). At the centre of these views is the claim that formalization is a form of reconstruction since it aims at more precision in representing logical forms by eliminating ambiguities and other troublesome features of ordinary language argumentation.

In what follows, this basic idea will be explored by introducing an explicit methodological framework based on the method of reflective equilibrium. More specifically, I will analyse the "core business" of traditional logical analysis, which aims at proving the validity of arguments with the help of logical formalisms. The guiding question is how we can justify the various reconstructive elements involved in showing that an argument is valid. This question is addressed in sections 4 and 5, where I give a general account of the method of reflective equilibrium and explain how it is applied to the justification of logical systems, specifically to explications of logical validity and to theories of formalization. Section 6 then uses a sample of Anselm's writings to discuss how the method of reflective equilibrium may be applied to the reconstruction of particular arguments. This case study provides examples for most points which are discussed in general terms in sections 1–5.

But first of all, we need a clearer picture of the various components of logical reconstruction (sect. 1), of their goals and of the challenges that result for justifying this practice (sect. 2). Since formalization, an essential part of logical reconstruction, is poorly covered in the literature, I will provide a short introduction to theories of formalization as a background in section 3.

1.   Setting the scene: showing validity by reconstruction

The heart of the traditional logic-oriented project of argument analysis is determining the arguments a text presents and evaluating their validity. Consequently, such an analysis is organized towards proofs of validity in some system of formal logic, paradigmatically zero- or first-order logic. Since I will defend the view that logical argument analysis is best understood as a form of reconstruction, I will speak of "argument reconstruction" and use "argument analysis" in a narrower sense. Figure 1 is a provisional scheme for reconstructing arguments (cf. Rosenberg 1994, 162):[1]
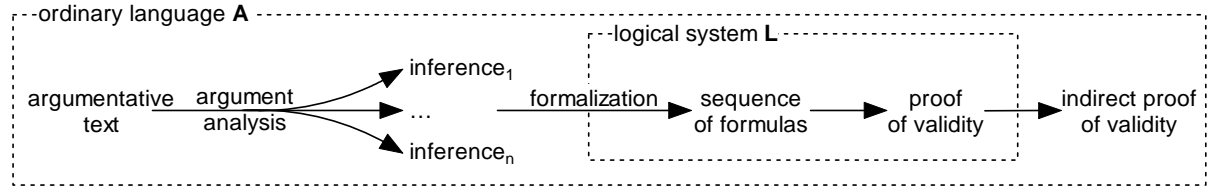
Figure 1: Simplified structure of argument reconstruction

In this scheme, four steps (represented by arrows) can be distinguished:

1. Argument analysis deals with a text in an ordinary[2] language **A**, identifies one or more arguments, analyses their relation and reconstructs them as inferences. An inference is an argument in **A** which is in the standard form of a (finite) sequence of at least two sentences $\langle A_1, \ldots, A_n \rangle$, where $A_1, \ldots, A_{n-1}$ are the premises and A_n is the conclusion. Argument analysis involves four types of reconstructive manipulations: deleting irrelevant elements, adding, reformulating and reordering premises and conclusion (see Brun/Hirsch Hadorn 2009, ch. 8.2). The following three steps focus on individual inferences.

2. Formalization takes us from inferences to formulas.[3] More precisely, given an inference $I = \langle A_1, \ldots, A_n \rangle$ in an ordinary language **A,** a formalization of I in a logical system[4] **L** is an ordered pair $\langle F, \kappa \rangle$, in which F is a sequence $\langle \phi_1, \ldots, \phi_n \rangle$ of formulas of **L** and $\kappa = \{\langle \alpha_1, a_1 \rangle, \ldots, \langle \alpha_m, a_m \rangle\}$ is a correspondence scheme that specifies one-to-one for each non-logical symbol α_i $(1 \leq i \leq m)$ occurring in some ϕ_j $(1 \leq j \leq n)$ an expression a_i of **A** (augmented by auxiliary expressions for representing argument-places of predicates). For convenience, 1-place sequences will be identified with their single member in formalizations of sentences; $\langle \phi_i, \kappa \rangle$ is then a formalization of A_i in **L** $(1 \leq i \leq n)$.[5]

3. A proof of $\phi_1, \ldots, \phi_{n-1} \Rightarrow \phi_n$ is attempted, using the formal syntax or semantics of **L**.

4. If the preceding steps have been carried out adequately and successfully, the result can be carried over to the inference and from there to the original argumentation.

   This scheme incorporates two limitations, which I will not discuss extensively. First, there is an asymmetry between showing that an argument is valid and showing that it is invalid (cf. the discussion initiated by Massey 1975; see Cheyne 2012). The procedure outlined in figure 1 is silent on invalidity.

---

[1] In philosophy, logical analysis often deals not with arguments but with individual sentences (or certain types of sentences, such as "if … then …"-sentences or sentences with definite descriptions). As a *logical* analysis, it is an analysis with respect to the sentence's possible occurrence in (in)valid arguments. The following scheme can easily be adapted to the analysis of individual sentences, but I will not discuss the ramifications such a modification might have.

[2] I use "ordinary" in a broad sense covering not only English, Greek etc. as they are used in everyday communication, but also more technical variants thereof, which can be found in, e.g. science and theology; Anselm's Latin is an instance of the latter.

[3] In the literature, "formalization" is often used with other meanings, most notably referring to the development of formal systems (cf. Morscher 2009, 3–40 for a survey).

[4] The notion of a logical system will be further discussed in sect. 5. For the time being, I assume that a logical system comprises at least a logical formalism, which is a formal language with a semantic or proof-theoretic definition of validity and further logical notions.

[5] As a simplification, I assume that logical forms be attributed to sentences, as opposed to, for example, propositions or tokens (but see Brun 2008). Furthermore, the requirement that correspondence schemes specify one-to-one relations is strictly speaking too strong and could be relaxed to functions from non-logical symbols to expressions of **A** (cf. Brun 2004, ch. 6.1).

The reason is that according to the standard definition of logical validity, an inference is valid (relative to **L**) if and only if it has at least one valid logical form (in **L**).[6] Consequently, identifying one valid logical form suffices for showing that an inference is valid, whereas for showing invalidity, all logical forms of the inference in question have to be examined. It is unclear how one could in general argue for having met this requirement since even within one logical system different, for example, more or less specific, formalizations can be given for the same inference.[7]

Secondly, since steps 2 and 3 focus on individual inferences, applying the results of step 3 to the argumentation in the original text is typically not straightforward. Only in the simplest case, does an inference cover the entire original argumentation. More often, it will represent only one element of a more complex structure of arguments. A comprehensive logical reconstruction should therefore also include an explicit representation of argumentative structure; that is, of how the argumentation can be broken down into individual inferences and of the relationships between the individual inferences. But even if a complex argumentation has been analysed into inferences that all have been shown to be valid or inductively strong, there is still the question of what those inferences contribute to the original argumentation. While the literature extensively deals with classifying argumentative structures and diagramming methods, consequences for evaluating the logical validity of a complex argumentation are frequently neglected (but see, e.g. Dorn 2006; Betz 2012).

In this paper, I focus on the question of how we may justify the claim that carrying out a procedure as outlined above can be used to show the validity of a reconstructed argument. Whereas within a logical system **L**, the **L**-validity of a sequence of **L**-formulas can be proved in a strict sense following exact rules, such proofs are in general not available for inferences, let alone for "real" argumentative texts.[8] Their validity can only be shown with the help of informal arguments which show, in addition to the formal proof required in step 3, that the logical system **L** provides an adequate explication of the informal notion of logical validity, that the inference has been adequately formalized (step 2) and that the analysis used to reconstruct inferences from the original text is adequate (step 1). The details of the required informal arguments get uneven attention in the literature. Most neglected are theories of formalization, which I will therefore briefly introduce in section 3. The bulk of this paper will not discuss specific theories of validity and formalization but explore how the method of reflective equilibrium may be used as a framework for justifying the adequacy claims just mentioned. Sections 4 and 5 deal with justifying accounts of validity and theories of formalization; section 6 addresses the justification of the reconstruction of particular arguments.

Relying on the method of reflective equilibrium seems promising for addressing two particular challenges raised by the procedure sketched above. To begin with, the structure of reconstruction is simplified in figure 1 and speaking about "steps" might suggest a linear order of independent tasks. However, this is not intended, even if the four steps can be understood as more exact counterparts to three different informal questions which drive the reconstruction of arguments: "What are the individual arguments?" (step 1), "What are their logical forms?" (step 2) and "Are they valid?" (step 3 and 4). These questions motivate a distinction between text-interpretation in step 1 and logical investigation proper starting with step 2, but argument analysis is intertwined with formalizing and proving validity. On the one hand, determining individual inferences sets up the target for subsequent logical examination. On the other hand, the result of attempted formalizations and logical proofs can have consequences for the question of how exactly we should analyse the original argumentation and formalize the resulting inferences. This results in feedback loops and as a consequence, the justification of the various steps is interdependent.

A second challenge is that argument reconstruction as a whole and the individual steps distinguished above are guided by a variety of goals which are partly antagonistic. Consequently, trade-offs are unavoidable and this calls for a method of justification suitable for dealing with the resulting problems of striking a balance. Furthermore, the diversity of objectives is another reason for feedback effects in reconstructing, as will become clear in section 2, which discusses the various aims of argument reconstruction, analysis and formalization.

---

[6] A more exact definition is provided in sect. 3.3. Strictly speaking, all statements about validity, formalization and ascriptions of logical forms are relative to a logical system **L**. For the sake of simplicity, I often do not make this explicit.

[7] For diverging views on this point see Brun 2004, ch. 13; Blau 2008, 148–50; Baumgartner/Lampert 2008; Lampert/Baumgartner 2010; Brun 2012a.

[8] As Montague's pioneering work (Montague 1974) has shown, logical systems can be developed in which strict proofs of **L**-validity can be given for inferences in a fragment of ordinary language.

2.   Goals of argument reconstruction

The goals of argument reconstruction can be looked at from two perspectives. On the one hand, argument reconstruction may be instrumental in realizing all kinds of aims. Philosophically, a spectrum of goals between exegetical and exploitative interpretation is important (cf. Rescher 2001, 60). One can strive for a meticulous exegesis, trying to understand as accurately as possible whether and why an argument is valid on the author's own terms. Or one can investigate what may be the strongest argument that can be constructed following more or less closely the author's line of reasoning. Philosophically less respectable goals actually play no less important roles, for example, impressing others with one's skills in dissecting passages from St. Anselm. On the other hand, argument reconstruction involves argument analysis and formalization, and therefore it is committed to the goals which are constitutive for these logical techniques. These goals will now be discussed in more detail before I return to the goals of argument reconstruction at the end of this section.

2.1.   Goals of logical argument analysis

A first goal of argument analysis is to provide completely explicit ordinary language arguments; that is, enthymemes must be converted into complete arguments, and every premise and the conclusion must be specified as a complete sentence. This is not a goal of formalizing because we stipulated that formalization operates on inferences, which are the well-ordered, non-enthymematic products of argument analysis.[9]

Equally important are certain problems of ambiguity. The challenge is sometimes characterized as follows: Whereas argument reconstruction does not deal with actually assessing the truth of premises – this is a task of other kinds of research –, it must deal with factors that determine whether the truth of the premises can be assessed. Specifically, it must establish precision or exactness in the sense of eliminating ambiguity, context-dependency and vagueness of terms, up to a point that suffices for assessing the truth of premises and conclusion. Although precision in this sense is often mentioned in the literature (e.g. Morscher 2009, 4–5, 10–1, 17–8), logical validity of inferences and arguments need not be affected by these troublesome phenomena. From a purely logical point of view, argument analysis need not embark on a dubious programme of eliminating all ambiguity, context-dependence and vagueness. It is sufficient, if argument analysis deals with three problems directly affecting logical form. (i) Equivocations must be eliminated. The details of this requirement are difficult to specify, but the general idea is clear enough: within an inference, corresponding tokens of the same type must have the same semantic value (cf. Brun 2008). (ii) Argument analysis must also eliminate syntactical ambiguity, such as different readings of scope ("I saw the man with the telescope"). In practice, this may call for tinkering with ordinary language by introducing brackets or indices to disambiguate syntactical structure. (iii) Ambiguous "logical" expressions must be replaced or completed appropriately. If we suppose that "or" has an inclusive and an exclusive reading, this is a case in point.[10]

All three points (i)–(iii) can pose serious problems for argument analysis and in many cases cannot be carried out completely before a formalization is attempted. And what's more, formalizing may be the most effective and sometimes also the only practicable means of detecting and finding a way of dealing with those problems. And the same is true for making the argument completely explicit. In many cases, there is no better way to find out about missing elements in an argument than formalizing it and trying to prove its validity.

Additionally, argument reconstruction is often considerably facilitated by reformulations which eliminate unclear formulations or stylistic variations, for example, by replacing different expressions with one and the same if they can be considered synonymous in the context at hand. Reformulations, however, are potentially problematic moves. Argument analysis may undermine the goals of argument reconstruction if it deals informally with issues for which a more respectable theoretical treatment is available. Specifically, logical relations such as equivalence must not be dealt with in argument analysis if they can be dealt with by a logical system which is considered to be relevant in the context at hand (for examples, see (2) in sect. 3.1, and (I2) and (I3) in sect. 6.2).

All this means that there will often be reasons to revise a previous argument analysis in the light of attempted formalizations. However, the distinction between argument analysis and formalization must

---

[9] An alternative approach to enthymemes that does not rely on specifying additional premises is discussed in Brun/Rott 2013.

[10] That there are ambiguous logical expressions in natural language is a contested claim. LePore/Cumming 2009, for example, attack many popular examples (see, e.g. ch. 6.4 on inclusive and exclusive "or").

not be blurred unnecessarily. Argument analysis should do what is necessary to provide a starting point for subsequent formalization, but no more. As far as possible, inferences should be framed in the ordinary language of the original argumentation and present sentences of the original text (see the examples in Sect. 6). If rewording or the use of technical devices such as brackets and indices is unavoidable, it should be kept at a minimum. Inferences therefore do not provide a representation of a logical form of an argument in the way a formalization does. Thus a division of labour is established which ensures that formalization's task of identifying logical forms is not tacitly pre-empted by argument analysis.

2.2.　　　Goals of formalization

Quine most succinctly characterized the aims of formalizing as "to put the sentence into a form that admits most efficiently of logical calculation, or shows its implications and conceptual affinities most perspicuously, obviating fallacy and paradox." (1970, 396) For a more detailed account, we may refer back to the scheme from section 1. It first of all expresses the idea that formalizing is instrumental in showing that an argument is valid. A formalization provides the formulas that substitute for the premises and the conclusion of the inference in a proof of logical validity. Thereby, it becomes possible to use the resources of a logical formalism for giving the desired proof. Since logical proofs are designed to show that an inference is valid not just for any reason but in virtue of a logical form of its premises and conclusion, a first goal of formalizing is that it should provide an expression of the logical formalism which represents a logical form and exclusively a logical form of the inference in question.[11] It is the task of a theory of formalization to spell out the requirements formalizations must meet if they are to count as an adequate representation of a logical form of the original inference. Theories of formalization will briefly be dealt with in section 3, but three comments are appropriate here. Firstly, formulas exclusively represent logical forms, but formalizations, strictly speaking, do not because they include a correspondence scheme. This is not a serious worry. The correspondence scheme is only used in assessing the adequacy of the formalization (cf. sect. 3); it cannot play any role in proofs of validity since it is not an expression of the logical formalism.[12] Secondly, formalizing does not aim at representing *the* logical form of an inference, since in general an inference can be ascribed more than one logical form (Brun 2004, ch. 13). For one thing, ascription of logical forms is relative to a logical system. A conclusion with a logical form represented as $\exists x(Fx \wedge Gx)$, for example, cannot have this logical form in a system of zero-order logic. For another, logical forms can be analysed in a more or less specific (or "fine-grained" or "detailed") way (see p. 12). The conclusion just mentioned could also be formalized with $\exists x(Hx)$. Thirdly, the requirement of representing exclusively a logical form must be seen in connection with the Leibnizian ideal of strict, if possible even mechanical, proofs, which are reliable and intersubjectively valid. However, this is not something that can be demanded from individual formalizations. It rather calls for a logical system with rules of proof that are explicit, formal (in the sense of not referring to the meaning of formulas) and effective (it is always possible to decide whether a given manipulation of signs conforms to the rule) (cf. Frege 1879, IV-VI). Additionally, we may favour proofs that are as simple as possible. This ideal has important consequences for the following discussion, since it implies that a formalization provides a representation of a logical form that is exact in the sense just explained – even if it is the formalization of a sentence whose logical forms are not entirely clear. Hence, it is often necessary to square the exactness of formalizations with the requirement that formalizations be defended as adequate representations of logical forms which we can plausibly ascribe to the original inferences. This is a main reason why theories of formalization are difficult to develop.

---

[11] I speak of formulas *representing* logical forms of sentences because I think that the notion of logical form is best analysed as referring to certain features of sentences which are relevant to the validity of inferences (e.g. how sentences can be analysed into logically relevant constituents, or the property of being true or false represented by all formulas; cf. Brun 2004, chs 1.2, 4; 2008). But the discussion in this paper is compatible with views that take formulas to be logical forms (e.g. Sainsbury 2001) or interpret logical forms as separate entities, such as patterns instantiated by sentences (e.g. Lemmon 1987).

[12] I no longer subscribe to the view that correspondence schemes specify how non-logical symbols abbreviate natural language expressions and that hence the formulas in formalizations are expressions of a semi-formal language with a fixed meaning (Brun 2004, chs 6, 10.3; cf. Epstein 2001, 12–4). Rather, I hold that formulas are schematic expressions without any fixed meaning and that correspondence schemes are merely used to specify a relation between non-logical symbols and ordinary language expressions; they do not turn formulas into meaningful expressions.

For philosophers, another goal is at least as important: Formalizations should provide a transparent representation of logical forms, in Wittgenstein's words (1953, § 122), a "surveyable representation" that provides an overview of complex structures. It should be as easy as possible to "read off" logical forms from formalizations (cf. Wittgenstein 1921, 6.122) and any kind of unclarity about logical form should be eliminated. This goal must be understood against the backdrop of the so-called "misleading form thesis", of which Anselm was an early defender (see note 42 below). We cannot read off logical forms directly from the "surface structure" of sentences; that is, the structure they have according to a "naive" grammatical analysis, which is not based on logic but on a conglomeration of judgements about substitutability in certain contexts, subject-predicate distinctions and similar (pre)-theoretic considerations (cf. Brun 2004, 160–5; Sainsbury 2001, 339–47). In a transparent representation, the relationship between syntactical properties of formulas and the logical forms they represent should be as simple as possible to grasp.[13] Again, this is an ideal that calls for choosing an appropriate logical system. Relevant factors include the way syntactic structures are represented, choice of logical symbols[14] and abbreviation[15]. It is, for example, probably a matter of empirical fact that for most people formulas in Polish notation are less transparent than formulas in an infix-notation with brackets. Since transparency is a goal related to cognition, it is relatively independent from the ideal of mechanization mentioned above (cf. Frege 1893, VIII): Polish notation tends to simplify proofs. Furthermore, the transparency of a formalization depends on how specific it is. Simplicity usually enhances transparency and favours less specific formalizations, whereas the goal of proving validity calls for a sufficiently specific formalization. In case of valid inferences, the optimal compromise is usually the least specific formalization that permits a proof.[16] It can often only be found by trial and error. If a validity proof fails, we can try to refine the involved formalizations step-by-step until they are specific enough to permit the desired proof; or a successful proof can provide a clue to how it can be simplified by using less specific formalizations. Thus, success or failure of attempted proofs sometimes can prompt us to go back to a previous stage of reconstruction and revise a previously developed formalization.

2.3.   Two misconceptions about the goals of formalization

In the literature, translation is regularly invoked as a paradigm that should help to understand what logical formalization is (e.g. Barwise/Etchemendy 2008; Guttenplan 1997, 98–9; Hintikka/Bachman 1991, 247). Although convenient and suggestive, speaking of formalizing in terms of translating leads to misconceptions, which strongly tells against promoting this *façon de parler* to a methodological paradigm.[17] Two points are illuminating in the present context.

Firstly, there is the idea that the adequacy of a formalization should be explained in terms of some form of sameness of meaning, since this is what counts in case of translating. This immediately invites the objection that formulas are expressions of a schematic language and hence formalizations do not have a fixed meaning that could be compared to that of a natural-language expression. At best, we get the explanation that, for formalizations, sameness of meaning amounts to sameness of truth conditions (e.g. Barwise/Etchemendy 2008, 84–6). Whereas this is a good start for a criterion of adequate formalization (cf. sect. 3.1), taking the analogy further gets Barwise and Etchemendy into trouble. Their next claim is that equivalent formalizations are analogous to stylistic variants. Having the same truth conditions, they are equally adequate formalizations of the same sentences and may only differ in transparency or being more or less close to ordinary language form (whatever that exactly means). However, equivalence is subject to logical proof and should not be trivialized by simply choosing the same formalization for any two equivalent sentences (cf. Brun 2004, 235–40). Independently, equating adequate formalization with preserving truth conditions also faces the problem that matching truth conditions is not a sacrosanct goal of accepted practice of formalization. Confronted with a sentence whose truth-conditions are not entirely clear, we do not look for a

---

[13] A similar point about transparency can also be made with respect to proofs. Tennant (1997, 61), for example, argues that natural logic in "tree design" shows much more clearly on what premises or assumptions a conclusion depends than the more common line-by-line layouts.

[14] Often, a convenient choice of non-logical symbols will make it easier to remember the correspondence scheme. But this is not a matter of transparency in the sense explained.

[15] Abbreviation merely for the sake of economical expression is an entirely subordinate goal. In contrast to abbreviation as a factor contributing to transparency, it does not serve logical purposes any more than using shorthand.

[16] This is Quine's maxim of shallow analysis (1960, 160).

[17] For a more detailed critique of the paradigm of translation see Brun 2004, ch. 8.1.

formalization with unclear truth-conditions. If we do not reject the sentence as a candidate for formalizing, we rather settle the matter by formalizing. Or we fill in "truth value gaps" by exploiting a grammatical analogy to other sentences; for example, when we formalize "All unicorns are ruminants" as an instance of $\forall x(\phi x \to \psi x)$, just like "All cows are ruminants" (cf. Haack 1978, 33).

The point just made is related to a second misconception. If a translator has to "grasp what the speaker or writer means and to express it in the other language" (Hintikka/Bachman 1991, 4), we may think that formalizing amounts to grasping a logical form and expressing it in a formula. Even if not entirely wrong, this characterization draws an excessively one-sided picture of formalizing. For it is easily read as claiming that logical forms are somehow hidden in sentences, ready for us to find there; perhaps we only have to abstract from all aspects of meaning except for truth conditions to lay bare the "skeleton" of logical form.[18] However, logical forms are not simply given and are not found by sheer abstraction. First of all, determining a logical form amounts to distinguishing between logically relevant and logically irrelevant features of sentences. In formalizing, we draw this distinction against the background of a logical system, and doing so can be reason to deviate from our "normal" understanding of the sentence. We may, for example, detect ambiguities that are irrelevant to the purposes of ordinary communication, even though they can affect the validity of inferences. (E.g. whether "All heads of horses are heads of animals" includes the "analytic" claim "All horses are animals"; cf. Brun 2012a). Or we decide to deliberately deviate from ordinary usage for the sake of unambiguity or systematicity, as mentioned in the preceding paragraph. In sum, formalizing is not merely abstracting but also involves creative and normative aspects of constructing logical forms and resolving matters of logical form, which are alien to paradigmatic forms of translation. This also explains why "reconstruction" aptly characterizes the nature of formalization and argument analysis. It evokes the picture of a construction which is guided by a pre-existing object or situation as, for example, in Wittgenstein's example of using a configuration of toy cars and dolls as a reconstruction of a road accident (Wittgenstein 1979, 7). In contrast to pictures of X-raying and archaeology, such a process of re-constructing leaves room for creative or normative departures for the sake of promoting the goals of formalizing (without, of course, necessitating deviations).[19]

2.4.   Goals at the level of argument reconstruction

As observed at the beginning of this section, logical argument reconstruction serves instrumental goals, most notably exegetical or exploitative interpretation, but also goals which are constitutive for its being a *logical* technique. This first of all includes promoting the aims of argument analysis and formalization just discussed. Furthermore, explicitness and transparency become relevant on the level of complex arguments in two ways. Firstly, an argument reconstruction should show unequivocally and in a way that is easy to grasp how an argumentation was analysed into individual inferences and how these inferences are linked together. For this purpose, there are well-known methods which represent that the conclusion of one inference is a premise of another using diagrams or cross-referencing formulas and sentences involved in inferences. Secondly, formalization should be uniform in the following sense. Within one argument reconstruction, all formalizations should at least use compatible correspondence schemes which respect the requirement of avoiding equivocation.[20] Additionally, it is normally desirable to formalize all occurrences of the same sentence by the same formula. If different formalizations of the same sentence are available, it is often best to use one which is at least as specific as all other formalizations already available. This makes it possible to show the validity of any inference for which this can be done relying on already identified logical forms. However, transparency can also speak in favour of using different formalizations in different inferences because representing unnecessary detail detracts from the logical structure which is actually responsible for validity.

The considerations in this section underline once more that a plausible account of argument reconstruction cannot be forced into the simplified linear scheme outlined in section 1. Figure 1 depicts formalizing exclusively as replacing inferences with sequences of formulas and ignores that formalizations frequently generate further effects and feedback-loops. In this way, the figure draws

[18] Metaphors involving skeletons and X-raying are rather common in the literature, e.g. Haack 1978, 23; Strawson 1952, 49.

[19] My use of "reconstruction" does not imply that ordinary language arguments are deficient and need to be improved upon, nor that formal languages are models of ordinary languages in the sense of "model" familiar from the philosophy of science (cf. Dutilh Novaes 2012, ch. 3.3.1).

[20] Two correspondence schemes are compatible iff the two schemes neither assign different ordinary language expressions to the same non-logical symbol nor different non-logical symbols to the same ordinary language expression.

attention to the question of how formalizations may be justified without acknowledging that analysing an argument as a particular inference may also be justified by appealing to a resulting formalization. Ambiguities of logical form are a good case in point (cf. the example "All heads of horses are heads of animals" above). They often go unnoticed until a formalization is effectively attempted. In such cases, one needs to go back to argument analysis, since evidence that may be used to decide between two possible interpretations must be sought in the original context.[21] This calls for a methodological approach that integrates interactions between formalization and argument analysis. In fact, even a more comprehensive perspective will be needed. Confronted with a formalization that unexpectedly does (not) permit a certain proof of validity, a range of (mutually non-exclusive) moves are possible. We can of course try another formalization or revise our argument analysis, but we can also change our informal judgements concerning the validity of the inference in question, or we can, in exceptional cases, even amend our logical system by changing the explication of logical validity or by modifying our theory of formalization. Selecting the appropriate manoeuvres requires that we also consider interactions between particular formalizations and the development of logical systems. In contrast to unidirectional models of translating and explicating, the framework of reflective equilibrium promises to account for the more complex structure of formalization and argument reconstruction. Before we look into this framework in more detail, I briefly introduce theories of formalization.

3.   A quick guide to theories of formalization

Theories of formalization can pursue at least three different but related approaches: conceptual analysis of *logical form, formalization* and other notions*,* criteria of adequate formalization and procedures of formalizing. In this section, I focus on the question of how accepted practice of formalization can be captured by criteria of adequate formalization.[22] Such criteria are motivated by a certain conception of logical form and at the same time contribute to making it more precise. In what follows, I distinguish three types of criteria. Firstly, criteria of correctness, which ensure that a sentence and its formalization match with respect to their inferential role and truth conditions (sect. 3.1).[23] A second type of criteria requires some syntactical correspondence of sentences and formulas (sect. 3.2). Thirdly, formalizing should be systematic (sect. 3.3).

   Although these criteria more or less implicitly guide accepted practice of formalization, they also raise problems and fail to provide a sufficient condition for adequate formalization. Ultimately, procedures of formalizing are necessary if these difficulties are to be resolved without giving up the basic goals and assumptions of the traditional project of formalization.

3.1.   Correctness: matching inferential role and truth conditions

A necessary condition for formalizations to be adequate is that they must be correct in the sense of not allowing validity proofs for inferences which are definitely invalid by informal standards. Using "*i*-valid" and "*f*-valid" to abbreviate "valid according to informal standards" and "valid according to the standards of a logical formalism",[24] this can be framed more precisely as follows:

   (VC)   A formalization $\langle \phi, \kappa \rangle$ of a sentence S in a logical system **L** is correct iff for every
            inference $I = \langle A_1, \ldots, A_n \rangle$ which includes S (i.e. $S = A_i$ for some $1 \leq i \leq n$): if there is at
            least one formalization $\Psi = \langle \langle \psi_1, \ldots, \psi_n \rangle, \lambda \rangle$ of I in **L** such that (i) Ψ includes ϕ as a
            premise or conclusion (i.e. $\phi = \psi_i$ for some $1 \leq i \leq n$), (ii) λ is compatible with κ, (iii) $\langle \psi_i, \lambda \rangle$

---

[21] The hermeneutic principle of charity cannot be applied by referring exclusively to the reconstructed inference and its formalization. Even if one interpretation results in a valid argument whereas another gives us a most absurd fallacy, the latter *may* be the more charitable interpretation; for example, if the argument in question is the author's straw man.

[22] A much more extensive discussion can be found in Brun 2004; Brun 2012a contains a summary of a few essential points. For further literature on formalization see the references in Brun 2004, as well as Kamitz 1979; Blau 2008, 141–50, 187–9; Baumgartner/Lampert 2008; Lampert/Baumgartner 2010; Peregrin/Svoboda 2013; Peregrin/Svoboda forthcoming.

[23] To simplify, I focus on formalizations of sentences. Inferences are formalized as sequences of formulas using one common correspondence scheme (cf. p. 2).

[24] "Formal" is an ambiguous qualification of validity, meaning either "in virtue of logical form" or "according to the standards of a logical formalism" (i.e. *f*-valid). The first is opposed to "material", the second to "informal". The distinction between material and formal requires explication; cf. (L) in sect. 3.3.

is a correct formalization of A_i for all $\psi_i \neq \phi$ ($1 \leq i \leq n$) and (iv) Ψ is f-valid in **L** (i.e. $\psi_1, \ldots, \psi_{n-1} \Rightarrow \psi_n$), then I is i-valid.

One obvious problem with (VC) is that the criterion appeals to further *correct* formalizations. This unavoidable circularity 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 (for examples and more detailed discussion see Brun 2004, ch. 11.2–3; Peregrin/Svoboda forthcoming).

In the context of semantical formalisms (in contrast to e.g. natural deduction based, purely proof theoretical frameworks), the requirement of correctness can also be framed in terms of matching truth conditions of sentences and their formalizations.[25] Given a sentence and a formalization in a logical system **L**, the basic idea is to construct in the semantics of **L** an interpretation of the non-logical symbols emulating the semantic values of the corresponding ordinary language expressions. Relying on informal reasoning, we can then test whether the formula and the sentence have the same truth conditions. Sameness of truth conditions means not merely having the same truth value, but the same truth value in all possible conditions.[26] (Otherwise, every true sentence could be correctly formalized by any tautology.) This motivates the following criterion for first order logic:

(TC)   A formalization $\langle \phi, \kappa \rangle$ 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, $\mathcal{I}(\phi)$ matches[27] 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 a 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 $1 \leq i \leq n$).

In principle, one may expect the two criteria to be equivalent, but whether in fact they are is a question of whether our informal judgements about truth conditions and inferential role fit together. To simplify, I will focus on (TC), but the discussed problems similarly apply to (VC) as well.

An example shows how (TC) functions:

(1)   Smith ate tomato sorbet and Jones was amused.
(1.1) $p \wedge q$ p: Smith ate tomato sorbet; q: Jones was amused

Suppose that (1) is true. Then its two sub-sentences must also be true. If we interpret p and q correspondingly (i.e. as true) then, according to the semantics of propositional logic, $p \wedge q$ is also true, just as (1). If (1) is false, at least one of its sub-sentences is false. $p \wedge q$ is then also false in a corresponding interpretation and hence has the same truth value as (1).

(TC) presents two problems. First, it treats any formalization equivalent to a correct formalization as correct (cf. Davidson 1980, 145). This admits preposterous formalizations (e.g. $\neg\neg(p \wedge q) \wedge (s \rightarrow (r \rightarrow s))$) in place of $p \wedge q$) and unacceptably trivial proofs for inferences involving equivalent sentences. For example, the formalizations in (2.1) are (TC)-correct and no more than the f-validity of $\phi \Rightarrow \phi$ is needed for "showing" that (2) is valid:

(2)   All Martians are green. Therefore: There are no Martians that are not green.
(2.1) $\forall x(Fx \rightarrow Gx) \Rightarrow \forall x(Fx \rightarrow Gx)$ Fx: x is a Martian; Gx: x is green

Using (2.1) to "show" that (2) is valid is unacceptable since the validity of (2) is a matter of relations between quantifiers and negation. It is the point of first order logic to provide a proof which shows why such relationships give rise to valid inferences. The goal of formalizing is to make such proofs possible, hence it must not pre-empt them. However, the unacceptability of formalizing equivalent sentences with the same formula is relative to the logical systems we consider relevant. If we imagine that for some reason we had only universal quantifiers[28] or zero-order logic, (2.1) would be no worse than formalizing "Jack and Jill went up the hill" and "Jack went up the hill and Jill went up the hill" using the same formula.

---

[25] In essence, this criterion is adopted from Blau (1977, 6–10; 2008, 146).

[26] This point is not unmistakably clear in Blau and Sainsbury (e.g. Sainsbury 2001, 64, but see 161–3, 192, 343).

[27] It is necessary to speak of "matching" rather than "identical" semantic values since, for example, truth may be represented by the number 1 in the semantics of **L**.

[28] Consequently, minimizing logical symbols not only increases parsimony of the logical vocabulary but also reduces the range of non-trivial proofs (pace Borg/LePore 2002, 88).

As Blau (1977, 15–7) has shown, this problem also opens the door to mock proofs for all *i*-valid single-premise inferences by conjunction elimination alone. For example, there is no need to bother with quantification for showing that (3) is valid:

(3)   Messner has climbed every Himalayan peak. Therefore: Every Himalayan peak has been climbed by somebody.

(3.1) $p \wedge q \Rightarrow q$ *p*: Messner has climbed every Himalayan peak
                                          q: Every Himalayan peak has been climbed by somebody.

Assuming that the conclusion informally follows from the premise, (3.1) meets (TC): if the premise is true, *p* and *q* are both true in a corresponding interpretation and hence also $p \wedge q$; if it is false, *p* is false and hence $p \wedge q$ as well. This method of "showing" validity can be extended to all *i*-valid inferences, if we accept that the premises of an inference are formalized as one conjunction.

Secondly, (TC) is not distinctive enough if materially *i*-valid inferences are involved:

(4)   Today is Monday and tomorrow is Tuesday.
(4.1) $p \wedge q$ *p*: Today is Monday
(4.2) $p \vee q$ *q*: Tomorrow is Tuesday

If (4) is true, both its sub-sentences are true and hence (4.1) and (4.2) are true in a corresponding interpretation. If (4) is false, both its sub-sentences are false and so are (4.1) and (4.2) in a corresponding interpretation. Thus, (4.1) and (4.2) are both (TC)-correct, but surely we expect (4.2) to be inadequate. The problem is that there are no possible conditions in which "Today is Monday" and "Tomorrow is Tuesday" have different truth values. It simply cannot be that tomorrow is Tuesday but today is not Monday. Since (TC) relies on informal reasoning about truth conditions, the question whether (4) is true under such absurd conditions does not arise. It would arise explicitly, should we use truth tables to apply (TC). However, such a move must be rejected because it informally makes no sense to assess the truth value of (4) with reference to semantically impossible conditions.

3.2.   Corresponding syntactical surface

In a fairly obvious move, the problems raised for correct formalizations may be addressed by introducing additional criteria which enforce some similarity between sentences *S* and their formalizations $\langle \phi, \kappa \rangle$. (3.1) can be criticized as inadequate because the expression corresponding to *q* does not occur in the premise of (3), and (4.2) because it contains "$\vee$" instead of "$\wedge$". As general rules to be applied to correct formalizations, we could require that all logical symbols in ϕ have a counterpart in *S* and that κ does not include ordinary language expressions not occurring in *S.* To block trivial proofs for the equivalence of a conjunctive sentence with its counterpart in reverse order, we could stipulate that non-logical symbols in ϕ occur in the same order as their counterparts in *S.*

Rules operating on the syntactical surface implicitly guide the common practice of formalizing, 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 virtually endless lists of exceptions. While we can point out, for example, that "there are no" is not an English counterpart of "$\forall$" and hence the conclusion of (2) is not adequately formalized in (2.1), the premise of (2) should not be considered inadequately formalized in (2.1) on the grounds that it contains neither a counterpart of "$\rightarrow$" nor "is a Martian" or "is green". For the rules to be at all manageable, their application should be restricted to comparing correct formalizations with respect to their adequacy. However, this does not eliminate the basic problem, which is captured in the misleading form thesis (see p. 6). What is more, the surface rules do not even draw on naive grammar but just rely on the substring-relation. If they were the essence of formalizing, logical forms would be reduced to sequences of logical and non-logical expressions (for a critique see, e.g. Sainsbury 2001, 352).

On the other hand, the surface rules are well motivated. If there are sentences which are in a non-trivial way equivalent by virtue of their logical forms, this is a matter not only of their truth conditions but also of their syntactical features. Moreover, one may argue that the divergences between the premise of (2) and its formalization in (2.1) are innocuous because they are systematic; that is, standard practice generally formalizes "All … are …"-sentences as instances of $\forall x(\phi x \rightarrow \psi x)$.

The problem with surface rules is that they require formalizations to mimic a naive analysis of ordinary language expressions instead of enforcing a correspondence between (parts of) sentences and (parts of) formalizations that systematically respects logically relevant differences between

sentences.[29] Rather than introducing rules that are the more plausible the more liberal we interpret them, we should try to spell out in detail how ordinary-language expressions can be formalized adequately depending on their syntactical structure. As the misleading form thesis shows, naive grammar is not up to this task. We need a considerably more ambitious analysis. Specifying accurate rules of syntactical correspondence most likely comes close to devising a systematic procedure of formalization.

3.3. Formalizing systematically

A first way of appealing to systematicity is exemplified in another reaction to the problems with (TC): (4) should be formalized as (4.1) in analogy to other simple "… and …"-sentences such as (1). Systematicity in this sense is well entrenched in philosophical practice, which aims at finding ways of formalizing not individual sentences but rather sentences of a certain type. The point of Russell's analysis of "The present King of France is bald" is that his analysis of this sentence is a blueprint for formalizing definite descriptions. Similarly, sentences of the form "All … are …" are standardly formalized as instances of $\forall x(\phi x \rightarrow \psi x)$, even against misgivings we may have because in many examples with subject-terms with null-extension this results in truth conditions we would not informally come up with (e.g. Wittgenstein 1980, 53). The strategy of formalizing analogous sentences analogously relies on a principle of "parity of form" (Russell 1905, 483; cf. Epstein 2001, 171–2, 178, 185–7). Undisputedly adequate formalizations are used as models for formalizing less clear cases. This raises problems closely resembling the difficulties affecting surface rules: how can we appeal to "parity of form" in a way that appreciates the misleading form thesis? Ultimately, the relevant form must be determined by logically significant similarities or differences between sentences (which are relative to the logical systems we consider relevant in the context at hand). Again, we need an exact specification of the classes of sentences which can be formalized as instances of the same scheme.

The technique of formalizing step-by-step is a second element of standard practice calling for systematic formalizations (e.g. Barwise/Etchemendy 2008, chs 11.3–4; Epstein 2001, 182). While the strategy of formalizing analogously is systematic in linking adequate formalizations of various similar sentences, the step-by-step method is systematic in relating different formalizations of the same sentence. Starting, for example, with

(5)   Every head of a horse is a head of an animal.
(5.1) $\forall x(Fx \rightarrow Gx)$ *Fx*: *x* is a head of a horse; *Gx* *x* is a head of an animal

we can formalize the two predicates "*x* is a head of a horse" and "*x* is a head of an animal" as (5.1.1) and (5.1.2) respectively and substitute the results for *Fx* and *Gx* to get (5.2):

(5.1.1) $\exists y(Hy \wedge Ixy)$ *Hx*: *x* is a horse; *Ixy*: *x* is a head of *y*
(5.1.2) $\exists y(Jy \wedge Ixy)$ *Jx*: *x* is an animal

(5.2) $\forall x(\exists y(Hy \wedge Ixy) \rightarrow \exists y(Jy \wedge Ixy))$

In its vague form, the instruction of formalizing step-by-step produces inadequate results (e.g. for Donkey-sentences; see Barwise/Etchemendy 2008, ch. 11.4; more sophisticated accounts are available though, e.g. LePore/Cumming 2009). Moreover, the strategy of formalizing step-by-step presupposes that we explicitly deal with formalizing not only sentences (as we have so far) but also their parts, predicates for example, and the way they make up the original sentence. There is, however, sound motivation for the step-by-step strategy in arguments of compositionality (as emphasized in, e.g. Davidson 1984), which combine the two senses of "systematicity" mentioned. Such arguments call for theories capable of accounting for the productivity of ordinary languages. We expect both theories of syntax and theories of formalization to deal not only with the specific examples which guided their development but also with the unlimited number of sentences and inferences of the language in question (cf. Borg/LePore 2002, 96–8).[30] This leaves us with the challenge of specifying in detail how formalizing step-by-step is supposed to work and how it bears on the adequacy of formalizations.

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. If this

---

[29] As argued in the context of (2.1), what differences are logically relevant depends on what features of sentences we can formalize in the logical systems we consider relevant.

[30] I do not want to imply that this necessitates assuming that ordinary language has an inherent logical structure (as in Borg/LePore 2002, 96–100).

analysis of principles guiding standard practice of formalizing is correct, it reveals an ideal not typically thought to be present: formalizations should be the product of an effective procedure, an algorithm which accounts for the systematic, compositional, nature of the ordinary language at hand. Without such a procedure, the criteria discussed remain problematic and fail to provide a sufficient condition for adequate formalization (cf. Brun 2004, chs 12.4, 14).

As arguments for the misleading-form thesis show, syntactic descriptions of naive grammar are not a suitable basis for specifying satisfactory criteria and formalization procedures. However, it does not follow that rigorous procedures of formalization are impossible, only that they must rely on a far more sophisticated analysis of ordinary language. This in turn calls for empirical investigations of language structures in relation to logical formalisms as pioneered by, for example, Montague (1974), Davidson (1980; 1984) and Lewis (1970). This research and other contributions drawing on Chomsky's theory of language also suggest that formalisms other than standard first-order logic, such as generalized quantifier theory, may be more promising.[31]

Currently available formalization procedures only deal with fragments of ordinary language, but we can also argue about the adequacy of formalizations by pointing out that they could (not) plausibly be the product of a systematic procedure. In this spirit, a criterion can be derived from a postulate of hierarchical structure (PHS) formulated in terms of a relation of *more specific* (or "detailed" or "fine-grained") between formalizations (Brun 2004, ch.13; cf. Castañeda 1975, 68–71):

(MS)   Given two formalizations $\Phi = \langle \phi, \kappa \rangle$ and $\Psi = \langle \psi, \kappa \rangle$ in a system of first-order logic **L,** 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.

(PHS)   If $\Phi = \langle \phi, \kappa \rangle$ and $\Psi = \langle \psi, \kappa \rangle$ are two adequate formalizations of a sentence *S* in **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.[32]

For example, (5.2) is more specific than (5.1) because it results from (5.1) by substituting [*Fx*/∃*y*(*Hy* ∧ *Ixy*), *Gx*/∃*y*(*Jy* ∧ *Ixy*)].

(PHS) is not a result that could be proved using the concepts of adequacy developed so far. It is rather independently motivated as a basic postulate since it is presupposed by what I take to be the standard definition of formal validity (cf. Brun 2012a):[33]

(L)   An inference *I* is formally *i*-valid relative to **L** iff *I* has at least one adequate formalization in **L** that is *f*-valid.

With respect to (L), one could ask why having at least one adequate *f*-valid formalization should suffice. Why not require that all, or perhaps the majority of adequate formalizations be *f*-valid? The answer is that (L) presupposes that the various adequate formalizations of an inference constitute a certain unity. (PHS) guarantees that this is indeed the case. We can then argue that, if there is an adequate *f*-valid formalization Φ of an inference *I*, all other adequate formalizations Ψ do not count against *I*'s *i*-validity: (i) if Ψ is equivalent to Φ, then the equivalence guarantees that Ψ is *f*-valid as well; (ii) if Ψ is less specific than Φ, then if Ψ is not *f*-valid, this simply means that we need a more specific formalization, such as Φ, to show the *i*-validity of *I;* (iii) if Ψ is more specific than Φ, then substitution theorems guarantee that Ψ is *f*-valid as well (Kleene 1952, §34); (iv) if Ψ is neither equivalent to nor more or less specific than Φ, (PHS) guarantees that there is an adequate formalization X that is more specific than Φ and Ψ and *f*-valid (because it is a substitution instance of the *f*-valid Φ), hence we have the same situation as in (ii) with X in place of Φ. Without the (PHS), it could happen in case (iv) that there is no adequate formalization X more specific than Φ and Ψ and then we had no reason why the *f*-invalidity of Ψ should not count against the *i*-validity of *I.*

---

31 It should be kept in mind that Chomsky's LF is a representation designed to capture empirically given syntactical structures, not determined by considerations of logical validity (Chomsky 1986, 205n, cf. 67, 156).

32 Variants of definitions (MS) and (PHS) that cover inferences (with a finite number of premises) and their formalizations can be specified, e.g. using a one-to-one mapping from sequences of formula $\langle \phi_1, \ldots, \phi_n \rangle$ to "inference-conditionals" $\phi_1 \wedge \ldots \wedge \phi_{n-1} \rightarrow \phi_n$.

33 Strictly speaking, (L) is only a blueprint for definitions as long as **L** is not specified.

(PHS) can be used as a criterion of adequate formalization (called "criterion of hierarchical structure", "(HSC)" for short), ruling that 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. By referring to formalizations related by substitutions, (HSC) also incorporates a core aspect of the requirement that formalizing be systematic in the spirit of the step-by-step method and compositionality. All adequate formalizations of a sentence, except the least specific ones (i.e. single sentence letters), can be interpreted as being derived by substitution from the least specific ones. While this criterion specifies a necessary condition of adequacy, it is merely a negative one, it can only be applied if we have two rival formalizations for the same sentence, and it does not determine which one is inadequate.

We are now in a position to discuss how the various moves involved in argument reconstruction may be justified by the method of reflective equilibrium.

4.   Reflective equilibrium and the justification of accounts of validity

Goodman (1983, ch. III.2) introduced the idea of a reflective equilibrium by describing a process of mutual adjustments which aims at establishing an agreement between judgements about the (in)validity of inferences and principles of validity. He defended the view that such an agreement is central to the justification of both the judgements and the principles. The account of reflective equilibrium I suggest extrapolates from this basic idea, but I also draw on Rawls (1999; 1975), Daniels (1996) and especially Elgin (1996; 2014).[34] I first discuss the agreement between judgements and principles, then the process of mutual adjustments and the criteria of justification. Simultaneously, I show how the account can be applied to theories of valid inference.

An equilibrium involves four basic elements: judgements, principles, background theories and a relation of agreement between them. Let us, for the moment, ignore background theories. By speaking of "judgements" and "principles", Goodman draws attention to the contrast between the account of validity given in a logical system and the judgements of validity somebody actually forms or is ready to accept. As examples of systems of principles, we may think of axioms with a deduction-rule, rules of natural deduction or rules defining validity in terms of semantic tableaux. Judgements include those statements about (in)validity and those acts of (not) accepting inferences as valid that we actually employ in our practice of inferring. Unfortunately, the literature is unclear about what distinguishes judgements and principles. The crucial difference is neither that judgements and principles are about something different (both are about the validity of inferences), nor that judgements are about particular inferences while principles are general.[35] Rather, principles simply decide what inferences are valid according to the system they (the principles) constitute, whereas judging an inference as valid means that we are ready to treat the inference as valid, for example to accept its conclusion as true if we accept the premises. Only paradigmatically, a judgement is an explicit statement that some inference is (in)valid. Judgements can also be expressed in behaviour, for example, by putting forward an inference or by treating a given inference as acceptable. In short, any behaviour that commits somebody to the (in)validity of an inference counts as a judgement and is relevant to the justification of principles of validity. To avoid the association of judgements with explicit statements and the oxymoron of "particular principles", it is better to speak of "commitments" (as in Elgin 1996) vs. "systematic elements" instead of "judgements" vs. "principles". It is important to remember that "commitment" and "systematic element" are technical terms. "Systematic" is not used in the sense of "orderly" or "methodological" but of "part of a system". Commitments come in degrees; they need not imply firm acceptance but may also be merely working hypotheses; and they need not be intuitive or spontaneous, but can also be the product of previous theorizing or of background theories.

According to the metaphor of "equilibrium", an agreement between a set of commitments and a theoretical system is crucial for justifying the commitments and the system. This is usually explained in terms of coherence. It requires (at least) consistency of commitments, consistency of systematic

---

[34] For a survey of the various accounts of reflective equilibrium see Hahn 2000 and Stein 1996.

[35] Goodman (1983, 64) explained the contrast between "judgements" and "principles" as particular vs. general, but Rawls (1975, 289) made clear that there are also general judgements. Examples are the judgements that all inferences from conjunctive statements to one of their conjuncts are valid or that from a general statement all instantiating singular statements can validly be inferred. In fact, there are also particular "principles", for example in axiomatic systems with non-schematic formulas as axioms and a rule of substitution (for examples, cf. Church 1956, 158).

elements, and that the commitments be derivable from the system.[36] In the case of logic, one might require that an inference is valid according to the relevant commitments iff it is valid according to the system. In practice, the requirement must be attenuated because, strictly speaking, it does not permit the development of theories which cover only part of a range of commitments. A system of zero-order logic, for example, is expected to sanction only valid inferences as valid, but not to cover all valid inferences. This means that it can only justify some validity-commitments, but not all of them and no invalidity-commitments.

Normally, the desired agreement will only be reached by a process of mutually adjusting commitments and systematic elements. We may revise a systematic element that stands against a clearer commitment or a commitment that stands against a more weighty systematic element; or, if we do not have firm commitments, we may simply let the system decide (Goodman 1983, 64, 66).[37] Neither systematic elements nor commitments have as such a privilege of not being revisable. Furthermore, the process of adjusting can yield different logical systems. Examples include not only systems which are different yet equivalent in the sense of classifying the same sequences of formulas as valid (e.g. zero-order logics with truth-tables and natural deduction), but also systems that sanction different sets of validity-commitments as illustrated by the discussion about the justification of intuitionistic logic (cf. Prawitz 1977; Daniels 1980; Haack 1982). As Goodman insists, this process of justification is not to be confused with a record of how we in fact did arrive at the commitments and systematic elements in question. It rather spells out what is required for a justification. We need to be able to describe how the system could have been developed from antecedent commitments, irrespective of whether we have actually built the system in this manner.

The exposition given so far ignored that a (so-called "wide") equilibrium includes background theories in addition to commitments and systematic elements.[38] Background theories are relevant to the justification of commitments and systematic elements in the foreground but can be treated as independently justified to some degree. Examples relevant to logics are semantical theories or theories of speech acts. Including background theories is a holistic element, and the difference between foreground and background is a question of perspective only. Background theories have three important characteristics. They contribute to the justification of foreground theories. They are a third "player" in the process of mutual adjustments (besides commitments and systematic elements) and can also be adjusted. Thirdly, they need, for their justification, their "own" reflective equilibrium, one in which they are in the foreground. Consequently, a reflective equilibrium includes commitments, systematic elements and background theories, but only the first two are justified by this reflective equilibrium. The distinction between background and foreground theories reflects the fact that inquiry must proceed piecemeal even if justification is holistic.

As a justification of commitments and systematic elements, an agreement between the two is not sufficient. Following Elgin's terminology (1996, 107, 127–8; 2014, 254), I call such an agreement an "equilibrium"; for *reflective* equilibrium, three additional criteria must be met.

The second criterion requires that at least some current commitments have a minimal epistemic standing of "initial credibility" or "initial tenability", independent of their coherence with other commitments and with systematic elements (Goodman 1952, 62–3; Scheffler 1963, 314–25; cf. Elgin 1996, 101–7; 2014, 254).[39] This criterion meets the standard objection that coherence cannot generate justification *ex nihilo* (cf. Brun 2014).

The third criterion requires that antecedent commitments be adequately respected. This should guarantee that the process of developing an equilibrium does not implement revisions so drastic that we end up with a system which does not count as a theory of valid inference any more. Somebody proposing a "logical system" which declares all inferences drawn by certain people to be "valid" would have changed the subject and ended up with a theory of authorship perhaps, but not of logical validity. To avoid such a "changing of the subject", justification requires to respect the commitments made

---

[36] In the context of logical systems, the conditions of consistency and derivability raise the question of whether the logical notions involved in these metatheoretical criteria agree with the logical notions that they are employed to justify (cf. Brun 2012b).

[37] Goodman's description is simplified. In particular, conflicts and agreement are not limited to two elements but relate to an entire system and a set of commitments (cf. Scheffler 1963, 317–8).

[38] Daniels (1979; 1980) introduced background theories to explain Rawls's (1975) notion of a wide reflective equilibrium. Although Goodman did not explicitly speak of background theories, his epistemic holism implicitly introduces them as well (cf. Elgin 1996, 13; 2014).

[39] This use of "initial" is not related to "antecedent" as used in "antecedent commitments". This contrasts with Elgin's discussion (1996, 107, 110), which does not distinguish between my second and third criterion and demands initial credibility of antecedent commitments.

before the process of mutual adjustments was started. Such "antecedent" commitments may be just pre-systematic assumptions, but more often they are based on a previous, now at least partly superseded system. "Respecting" means that if antecedent commitments are discarded, replaced or changed, we must be ready to explain why this has been done. Typically we will then refer to the relative weight of commitments and systematic elements, which in turn partly depends on what they contribute to realizing pragmatic-epistemic goals. It is important to note that, according to the analysis suggested here, commitments are involved in two contrasts. In the preceding paragraphs, they were contrasted with systematic elements as not being part of the system, now we have the requirement that current commitments respect antecedent commitments.

The fourth criterion blocks the "conservative" strategy that avoids revising commitments whenever possible, in effect limiting adjustment of commitments to resolving conflicts within the set of commitments. In Goodman's account this is prevented by a list of almost incidentally introduced "virtues" of systems. "Convenience", "theoretical utility" (Goodman 1983, 66), "economy" and "resultant integration" (Goodman 1983, 47) all may be reasons to adjust otherwise clear commitments. Such desiderata are of special importance because they represent the central pragmatic-epistemic goals which motivate the transition from ordinary commitments to a systematic account in the first place. Doing justice to relevant epistemic desiderata ensures that we get a *system*, not merely a list of our commitments. We expect a logic to be a system of general principles that is well-organized, exact, comprehensive and as simple as possible; more specifically, we also will aspire for logical systems which allow for rigorous proofs of validity, which are sound and complete, and maybe even decidable. However, the pragmatic-epistemic goals may be partly antagonistic (e.g. simplicity vs. scope of application) and can be given different relative weight, reflecting different aims that guide the construction of a system. It seems dubious that we could specify in advance how in general the necessary trade-offs should be made. This is one more reason for expecting and actually welcoming that the method of reflective equilibrium allows for alternative systems to be justified. Standard formalisms of zero-order logic are easy to read, but if simple proofs are at a premium, Polish notation may be preferable and if economy in number of axioms and undefined connectives is the main concern, a single-axiom system (cf., e.g. Church 1956, 159) may be the one to choose.

Condensed into a short formula, commitments and systematic elements are in reflective equilibrium if they are in agreement (with one another and with background theories), have some independent credibility, do justice to pragmatic-epistemic goals and respect antecedent commitments. Whether a reflective equilibrium constitutes a justification obviously depends on what one expects a justification to accomplish. Demands for some kind of ultimate reasons will not be satisfied, neither in the sense of reduction to principles that hopefully cannot be denied, nor in the sense of giving reasons whose validity does not depend on any previous commitments. Showing that commitments and systematic elements are in reflective equilibrium also does not amount to ensuring their truth. Desiderata such as simplicity, precision and wide scope of application are not truth-conducive, but cognitive goals in themselves (cf. Hempel 1988).

5.   Justifying logical systems and theories of formalization

In the preceding picture of the justification by reflective equilibrium, some crucial points are missing. The relation of agreement between commitments and systematic elements is not as simple as standard accounts of reflective equilibrium suggest. A claim such as "commitments must conform to systematic elements" cannot be interpreted as straightforwardly requiring that commitments about the validity of inferences be logical consequences of the logical formalism. In fact, the commitments and the logical formalism generally are phrased in different languages and therefore cannot stand directly in a logical relationship. Talking about an agreement between commitments and systematic elements presupposes that our logical theories include more than just a logical formalism which specifies a formal language and principles of valid inference.

As a first additional element, we need a theory about the relationship of formalization, which provides criteria for associating logical formulas and expressions of some ordinary language.[40] Without such a theory we are not in a position to claim that a logical system says anything at all about the validity of inferences framed in some other language than the logical formalism itself. If, for example, we want to determine whether an inference with the conclusion "Some critics admire only one another" is valid according to Quine's *Methods of Logic* (cf. 1982, 293)*,* we cannot do so without deciding on the question of how that sentence can be formalized adequately. Will $\exists x(Fx \land \forall y(Gxy \supset Gyx))$ do the job? Formalizing arguments and explicating the informal notion of validity in a logical

---

[40] A theory of formalization also suffices to account for the converse relation of "verbalization", which holds between a formalization Φ and an inference I just in case Φ is an adequate formalization of I.

formalism are two intertwined tasks. Having a solution to only one of these problems would be entirely pointless when it comes to analysing the validity of arguments. Moreover, the adequacy of an explication of validity in some logical system and the adequacy of formalizations in this system cannot be assessed independently.

Secondly, an informal interpretation of the logical formalism is needed for developing a systematic account of validity or a theory of formalization. Consider once more the procedure described in section 1. The formalism of a system of zero-order logic, for example, provides a formal language and a notion of proof that can be used for formulating an expression such as $p \wedge q \Rightarrow q$ and for giving a proof for it. But more is needed for establishing a commitment about the validity of an ordinary language argument. We need to correlate a particular inference (e.g. "7 is odd and 5 is even. Therefore 5 is even.") with a sequence of formulas and a correspondence scheme (e.g. $\langle p \wedge q, q \rangle$ and $\{\langle p, 7 \text{ is odd}\rangle,$ $\langle q, 5 \text{ is even}\rangle\}$), assess this formalization as adequate, prove within the formalism an appropriate expression containing "$\Rightarrow$" (e.g. $p \wedge q \Rightarrow q$) and on this basis conclude that the original argument is valid. In doing so, we rely on certain assumptions about what the ordinary language "counterparts" of the elements of the formalism are. We presuppose, that "$\Rightarrow$" corresponds to "is valid" and not to, say, "is logically independent of". More exactly, we presuppose that "$\phi_1, \ldots, \phi_{n-1} \Rightarrow \phi_n$" corresponds to "the argument with the premises $A_1, \ldots, A_{n-1}$ and the conclusion A_n is valid" if $\langle \phi_1, \ldots, \phi_n \rangle$ together with a correspondence scheme is a formalization of $\langle A_1, \ldots, A_n \rangle$. Such rules of correspondence are what I call an "informal interpretation" of the logical formalism. They must be distinguished from interpretations in the sense of formal semantics. If the formal semantics include, for example, a function from the set of sentence-letters to the set {t, f}, then this interpretation-function is part of the formalism, whereas the correlation of "t" with "true" and "f" with "false" is a matter of informal interpretation (the letters are suggestive, but we could introduce a logical system in which "f" or any other symbol whatsoever is correlated with "true").

This leaves us with two consequences for applying the method of reflective equilibrium to the justification of logical systems. First, giving an explication of validity must go hand in hand with developing a theory of formalization. Secondly, we must acknowledge that any such justification presupposes an informal interpretation of the logical formalism at hand and, if needed, we should be ready to give an explicit account of it. Strictly speaking, the method of reflective equilibrium cannot be applied to an account that comprises nothing but a formalism; i.e. rules for using certain signs. Rather, the primary objects of justification are informally interpreted logical systems which include a logical formalism as well as a theory of formalization (cf. Brun 2004, 51–2; cf. Resnik 1985, 225).

The most promising way of integrating formalization into the process of justification makes use of the distinction between background and foreground theories. If we focus on the theory of validity, as in section 4, a theory of formalization is assumed as a background theory – although in fact, this commonly boils down to relying not on a theory but merely on informal and largely implicit considerations. As we have elaborate and well-studied theories of deductive validity but no sufficiently developed theories of formalization, the more interesting scenario has reversed positions: relying on background theories that include a theory of validity and further relevant theories, such as semantics, we can try to develop a theory of formalization making use of the method of reflective equilibrium. This calls for balancing principles of adequate formalization with judgements of, for example, the form "Inference *I* can be adequately formalized with the sequence of formulas $\langle \phi_1, \ldots, \phi_n \rangle$" (see also Peregrin/Svoboda 2013; forthcoming). Moreover, theories of formalization must promote the goal that logical formulas can be used in place of inferences in proofs of validity. They should ensure that applying the logical formalism leads to results that sufficiently correspond to commitments about the validity of inferences and about properties of their logical forms. The question of what exactly is needed here leads directly into the heart of a theory of formalization. It calls for specifying criteria of adequate formalization. Most likely, more or less extensive revisions of commitments are needed if an equilibrium should be reached, as exemplified by the standard treatment of the traditionally so-called "a-sentences" as invariably having a form $\forall x(\phi x \rightarrow \psi x)$.

6. Justifying particular reconstructions: a case study

So far, we have seen how reflective equilibrium is applied to the theories of validity and formalization that underwrite the reconstruction of specific arguments. This section explores whether and how the methodology of reflective equilibrium may be applied to the process of reconstruction itself. This approach seems promising since, in practice, we can usually observe a going back and forth between text, inferences and formalizations, revising both the formalizations and the inferences we formulate as the analysis of the argumentative text. Reflective equilibrium promises to be capable of dealing with such feedback as well as with the interaction between formalizing and developing logical systems. However, it is not just self-evident how we should apply the method of reflective equilibrium to argument reconstruction. If, for example, we are reconstructing an argument against animal

experimentation, we formalize the author's statements about the moral acceptability of treating animals in certain ways. But our reconstruction does not aim at developing principles of animal ethics, as in standard scenarios of using the method of reflective equilibrium for developing moral theories. It rather aims at specifying a formalization that can be justified as adequate and be used to show that the argument is valid (if it is). Hence, we can employ the method of reflective equilibrium in justifying reconstructions of arguments if we interpret argument analysis and formalization as a process that replaces commitments to the structure of the argumentative text, the arguments and their validity with corresponding systematic elements.

In what follows, I explore this approach, using an excerpt from Anselm's *De casu diaboli* for exemplary reconstructions. I will first investigate a short argument in considerable detail and then discuss a slightly more extensive passage more briefly to point out some additional aspects. In developing the reconstructions, I will proceed in the usual informal manner, which is characteristic for the "daily business" of argument reconstruction by philosophers. My focus will be on how the various manoeuvres of argument analysis and formalization play together and how they may be justified by the method of reflective equilibrium. Investigating the application of defined procedures of formalization or of argument reconstruction lies outside the scope of this paper. Reflective equilibrium is compatible with the application of such procedures, but it does not require their use.[41]

Here is the opening passage of *De casu diaboli* (DCD, 233) with my word-for-word translation:[42]

(A)   DISCIPULUS. Illud apostoli: "quid habes quod non accepisti": dicitur hominibus tantum, an et angelis?
MAGISTER. Nulla creatura habet aliquid a se. Quod enim seipsum a se non habet: quomodo a se habet aliquid? Denique si non est aliquid nisi unus qui fecit et quæ facta sunt ab uno: clarum est quia nullatenus potest haberi aliquid nisi qui fecit aut quod fecit.
D. Vere clarum.
M. Sed neque ipse factor neque quod factum est potest haberi nisi ab ipso factore.
D. Nec hoc minus clarum.
M. Ille igitur solus a se habet quidquid habet, et omnia alia non nisi ab illo habent aliquid.

Student. Is the apostle's [word] "What do you have that you have not received?" said of men only, or also of angels?
Teacher. No creature has anything from itself. After all, what does not have itself from itself: how does it have anything from itself? Moreover, if there is not anything except the one who made and whatever is made by the one: [then] it is clear that by no means can anything be had except [if it is] what made or what [this one] has made.
S. Truly clear.
T. But neither this maker nor what is made can be had if not from this maker.
S. That is no less clear.
T. Only that [maker] therefore has from himself whatever he has, and everything else does not have anything except from that [maker].

Applying argument analysis to this passage presupposes that certain preliminary questions have been settled at least provisionally. Specifically, we assume, that this passage was put forward as containing arguments, that it is a well-chosen excerpt without significant omissions, that the citation is accurate and the translation faithful. These are important points of text-analysis, but I will not investigate them in detail here.

6.1.   Reconstruction of the first argument

The argument analysis now starts out with various commitments. Some of them are about how the text may be structured into individual arguments. Others relate to an individual argument, specifically to what its premises and conclusions are, what their logical forms are and also whether the argument is

---

[41] As a procedure of argument reconstruction, Czermak et al. 1982 is of special interest in the present context because they explicitly address the interplay between argument analysis, formalization, selection of a formalism and deciding between various readings of a text.

[42] I use Anselm's text to illustrate strategies and problems of argument reconstruction. Its philosophical or theological interpretation is not my concern (see, e.g. Sweeney 2012, 211–5). So far, *De casu diaboli* has been of interest to logicians and philosophers of language predominantly for Anselm's early version of the "misleading form thesis" (e.g. Henry 1967) and his discussion of *nothing*, which starts right after the quoted passage (e.g. Henry 1974, 337–9; Horwich 1975; King 2003).

valid. Formulating inferences is a way of recording some of these commitments. To begin with, we may identify a first argument in (A):[43]

> (A1) No creature has anything from itself. After all, what does not have itself from itself: how does it have anything from itself?

and start with the following fragment of an inference:

> (I1.1) [1] What does not have itself from itself does not have anything from itself.
>
> …
>
> [2] No creature has anything from itself.

(I1.1) reveals that in our reading of (A1), we identify one premise and a conclusion, that they can be phrased as stated, and that we leave open whether some additional premises may need to be added. There are additional commitments, not explicitly represented in (I1.1). Some of them have already been mentioned as answers to preliminary questions of argument reconstruction. Another example is that the question in (A1) is merely rhetorical and expresses an assertion. The following commitments are worth pointing out:

> (C1) Segment (A1) picks out one argument from (A).

> (C2) Argument (A1) has one premise.

> (C3) Argument (A1) is valid.

> (C4) The conclusion of (A1) is a universal negative statement.

Quite plausibly, reading (A1) in its context gives rise to further commitments. For example:

> (C5) Anselm holds that being a creature implies not having oneself from oneself.

> (C6) Anselm uses "is a creature" and "has been made" synonymously.

These commitments clearly vary in weight. Whereas (C4) may be quite strongly endorsed, (C2) is hardly more than a working hypothesis which will be readily abandoned. As regards content, commitments (C1) and (C2) relate to argument analysis, (C3) to the evaluation of the argument at hand, (C4) to one of its logical forms, and (C5) and (C6) to relevant implicational and semantical relations respectively. Commitments may also rely heavily on our interpretation of other parts of the dialogue or draw on our knowledge about Anselm and medieval theology. (C6) is an example. That Anselm probably uses *facere* and *creare* as synonyms in (A) is inspired by the fact that such a language use is also frequently found in the Bible.[44]

Background theories can be a source of commitments as well. An important example is a hermeneutic principle of charity that speaks in favour of understanding Anselm as putting forward a valid and sound inference.[45] This does not only back up (C3) but also calls for interpreting Anselm's use of "make", "create" and "have" in a technical theological sense which diverges from the vernacular in a way that leaves room for his claims to be true or at least not blatantly false. Principles of charity are applicable to the validity of arguments since they call for striving for reconstructing arguments as rational, and validity is an aspect of rationality (Quine 1960, 59; Davidson 1973). The principles discussed in the literature differ with respect to how they are applicable to individual arguments. What I have in mind here is first of all a "presumptive" principle; one that calls for a defeasible assumption of validity for individual arguments (cf. Scholz 2001, Pt. II). In contrast, "holistic" principles require that our interpretation of an author be favourable on the whole (e.g. Davidson 1973). Such principles cannot be applied directly to individual arguments, but they nonetheless generate a tendency in favour of reconstructing individual arguments as valid. "Tie-breaker" principles apply if there are alternative interpretations compatible with the text. They require, other things being equal, to select the most favourable interpretation (e.g. Vorobej 2006, 29–30).

---

[43] To make cross-reference easier, I use the following system of labelling: "A" tags argumentative texts, "I" inferences, "C" commitments and "F" formalizations. "(A1)" labels the first argument extracted from (A), "(I1.1)" the first recasting of (A1) as an inference etc. Numbers in square brackets will be used to keep track of the individual premises and conclusions and to link them to their formalizations.

[44] See Petrus Lombardus, lib. II, dist. I, c. II. In other contexts, Anselm uses *facere* in a very general sense similar to the English *do* (cf. Uckelman 2009).

[45] See also the contributions by Löffler and Reinmuth in this volume.

All our commitments must be thought of as associated with some relative weight. We are, for example, much more committed to the two claims that (A1) has the conclusion mentioned in (I1.1) and that this conclusion has a logical form in line with (C4) than to the validity of (A1) and even less to any exact wording of the premises which may be added to (I1.1). At this stage of argument reconstruction our commitments may include also rather dubious claims. (C1), for example, is certainly provisional as long as we have not started to deal with the rest of (A); and leaving room for additional premises in (I1.1) is clearly at odds with (C2).

All this is not to say that a complete list of relevant antecedent commitments could be compiled at the beginning of an argument reconstruction, nor that we have to come up with such a list sooner or later. For one thing, I am sceptical about what such a completeness-requirement would actually amount to. And the validity of an argument can normally be assessed on the basis of relatively few commitments about its logical forms. In our example we can hope that a relatively coarse-grained analysis in first-order logic will suffice and that we can consequently ignore more detailed logical structures. Moreover, identifying and explicitly recording relevant commitments is not only a prerequisite but also a product of the process of argument reconstruction. Although the process of reconstruction starts with recording some commitments, it usually proceeds by discarding some of them and introducing others. Consequently, the resulting set of current commitments will typically intersect with the set of antecedent commitments, but neither set will include the other.

To proceed with our reconstruction, we have to introduce systematic elements. This is done by specifying a formalization that promises to capture our commitments as well as to promote the goals of argument reconstruction as specified in section 2. Based on (I1.1) we may suggest:

(F1.1)	[1]	$\forall x(\neg Sx \rightarrow \neg Ax)$	Sx: x has itself from itself
		…	Ax: x has anything from itself
	[2]	$\neg\exists x(Cx \wedge Ax)$	Cx: x is a creature

At this point, several conflicts call for resolution. To begin with, we noted a conflict among the commitments. (I1.1) is hardly compatible with (C2). Additionally, there are conflicts between commitments and the proposed formalization. Since (F1.1), like (I1.1), indicates a missing premise, it is also incompatible with (C2). We now have several options. Firstly, we could insist that (A) starts with a one-premise argument, stick to (C1) and (C2), and remove the ellipsis from (I1.1) and (F1.1). The resulting formalization is not valid and it is quite implausible that the two sentences in (I1.1) permit a formalization that can be proved valid. So we would come under pressure to give up (C3) as well. Consequently, charity speaks against this option, especially if there is another option which squares better with our commitments overall. Secondly, we could dismiss the assumption (C2) that the argument at hand has just one premise and complement (I1.1) and (F1.1) with an additional premise that preserves (C3). One possibility for doing so would be to revise the argument analysis, give up (C1) and look in the subsequent text for an additional premise. However, there is no obvious candidate for such a premise and this impression will be confirmed by the reconstruction of the rest of (A), which I present below. So we may look more closely at the option of supplying a premise not explicitly mentioned by Anselm.

There are several strategies available for coming up with an appropriate premise. In the present case we may notice that (I1.1) resembles an incomplete syllogism. If we start with (F1.1) and note that its premise is equivalent to the contrapositive $\forall x(Ax \rightarrow Sx)$, it is easy to identify a corresponding fragment of a syllogism of the second figure:

(Camestres)	[1]	P a M
	[3]	S e M
	[2]	S e P

Relying on *Camestres,* we can suggest a formalization that turns (F1.1) into a valid inference:

(F1.2)	[1]	$\forall x(\neg Sx \rightarrow \neg Ax)$	correspondence scheme as in (F1.1)
	[3]	$\neg\exists x(Cx \wedge Sx)$	
	[2]	$\neg\exists x(Cx \wedge Ax)$	

Based on (F1.2), we can now amend (I1.1) by adding [3] as a premise:

(I1.2)	[1]	What does not have itself from itself does not have anything from itself.
	[3]	No creature has itself from itself.
	[2]	No creature has anything from itself.

Selecting [3] as an additional premise because it is suggested by (F1.2) is an instance of the strategy of letting the logical system decide in cases where our commitments are inconclusive. However, this move is buttressed by (C5) which motivates [3]. Most probably it is also compatible with

our other (mostly tacit) commitments about Anselm's argument. This would have been different for many premises that turn (F1.1) into a valid inference (say, our favourite contradiction), although other options, e.g. $\forall x\neg(Sx \wedge Cx)$, would yield equally defensible alternatives to (F1.2).

The conflicts and strategies for resolution just discussed are only a part of a wider spectrum. Probably, the most significant forms of conflict between commitments regarding an argument and proposed formalizations are:[46] Firstly, we may find that the formalization is inadequate; that is, there is a discrepancy between the logical form as represented in the formalization and our commitments relevant to the logical forms of the argument. The preferred way to find out about such conflicts is to apply a theory of formalization, which consequently must be included in our background theories. To rectify the situation, we can either amend the formalization or we can decide that the argument may be analysed as a different inference which in fact can be adequately formalized as suggested. Secondly, the result of a formal analysis of validity can be at odds with a commitment to the validity or invalidity of the argument at hand. To show this to be the case, we need a background theory in which we can prove validity. Again, an inadequate formalization may be blamed for diverging judgements of validity. But other strategies for adjustments are possible as well. If an adequate formalization proves to be valid, we may conclude that, contrary to our commitment, the inference is valid after all and that commitment has to go. On the other hand, if an adequate formalization unexpectedly turns out to be invalid, we should not jump to the conclusion that our argument is invalid. For there are other possibilities to check out first. Maybe we just need a more specific formalization to prove its validity (see the example below), or we are dealing with an enthymeme and have to fill in some premises, or we may decide on reconstructing some elements of the argument as weaker or stronger claims (cf. Jacquette 1996). In some cases we may also consider changing our background theories. We may, for example, chose another logical system or, in exceptional cases, consider amending a logical formalism or a theory of formalization.

6.2.   Reconstruction of the remaining arguments

Let us now turn to the rest of (A). I will discuss this part of Anselm's text in less detail and focus on aspects which have been less prominent so far. First of all, the argument analysis of (A) is not straightforward at all. So far, we have assumed that (A) starts with (A1) as one argument for [2]. However, [2] is prominently placed at the beginning of (A) in a way that suggests that the entire rest of the quoted passage serves to support it, perhaps by one multi-step argument which leads to [2] via several intermediate conclusions. Moreover, it is plausible to break down (A) into several arguments because there are two expressions which can be used to tentatively identify conclusions. If we interpret "it is clear that by no means can ..." as expressing an entailment and "therefore" as signalling a conclusion, we can identify two one-premise arguments:

(A2)   Moreover, if there is not anything except the one who made and whatever is made by the one: [then] it is clear that by no means can anything be had except [if it is] what made or what [this one] has made.

(A3)   But neither this maker nor what is made can be had if not from this maker. Only that [maker] therefore has from himself whatever he has, and everything else does not have anything except from that [maker].

(I2)   [4] There is not anything except the maker and whatever is made by the maker.
      [5] Nothing is had except the maker and whatever is made by the maker.

(I3)   [6] Neither this maker nor what is made can be had if not from this maker.
      [7] Only that maker has from himself whatever he has, and everything else does not have anything except from that maker.

The inferences specified so far illustrate the four basic operations of argument analysis. While the transition to (I1.2) involved adding (premise [3]) and reordering ([1] and [2] from (A1)), reaching (I2) and (I3) requires deleting (the student's remarks) and reformulating (e.g. "the one who made" as "the maker"). This last move expresses a commitment that the two descriptions "the one who made" and "the maker" are used interchangeably.

The question of how the various arguments in (A) fit together is now best tackled with the help of formalizations. To simplify the resulting formulas, I will formalize the definite description "the maker"

---

46 See also Löffler 2006, 127–8 for a classification.

with an individual constant and an explicit claim of uniqueness ([8]). This implies that the interlocutors presuppose rather than assert the existence of an individual fitting the description.[47]

[8] $\forall x \forall y(Fxy \rightarrow x=a)$ *Fxy*: *x* made *y*

(F2.1)   [4] $\neg \exists x \neg (x=a \lor Fax)$ *a:*   the maker

           [5] $\overline{\neg \exists x(Ix \land \neg (x=a \lor Fax))}$ *Ix*:   *x* is had

(F3.1)   [6] $\neg \exists x(Jax \land x \neq a) \land \neg \exists x \exists y(Jyx \land Fay \land x \neq a)$ *Jxy*: *x* is had from *y*

           [7] $\overline{\forall x \forall y(Hxyx \rightarrow x=a) \land \forall x \forall y \forall z(x \neq a \land Hxyz \rightarrow z=a)}$ *Hxyz*: *x* has *y* from *z*

Now (F2.1) is valid, but (F3.1) is invalid. The latter is obviously also a consequence of the fact that, besides identity, [6] and [7] do not share any predicates. This motivates a revision of the formalizations. Using the step-by-step strategy, we can develop a more specific formalization by applying the substitutions [*Iy*/$\exists x \exists z Hxyz$] and [*Jyz*/$\exists x Hxyz$]:

(F2.2)   [4] $\neg \exists x \neg (x=a \lor Fax)$

           [5] $\overline{\neg \exists y(\exists x \exists z Hxyz \land \neg (y=a \lor Fay))}$

(F3.2)   [6] $\neg \exists z(\exists x Hxaz \land z \neq a) \land \neg \exists z \exists y(\exists x Hxyz \land Fay \land z \neq a)$

           [7] $\overline{\forall x \forall y(Hxyx \rightarrow x=a) \land \forall x \forall y \forall z(x \neq a \land Hxyz \rightarrow z=a)}$

(F3.2) is still not valid, but becomes so if [5] from (F2.2) is added as a premise. However, we still need to address the overall structure of (A), specifically, the question of whether and how the arguments (A1)–(A3) relate to [2]. Firstly, a valid inference can be constructed which essentially derives [2] from [7]. This requires, on the one hand, to add the premises [8], [9] and [10], which formalize the claims that there is at most one maker, that the maker did not make itself and that being a creature is equivalent to having been made (cf. commitment C6):

[9] $\neg Faa$

[10] $\forall x(Cx \leftrightarrow \exists y Fyx)$

On the other hand, the formalization of [2] needs to be replaced by a more specific one by applying the substitution [*Ax*/$\exists y Hxyx$].

[2] $\neg \exists x(Cx \land \exists y Hxyx)$

Secondly, combining premise [1] with the formalizations [4]–[10], does not open up promising possibilities of constructing new inferences with conclusion [2]. This speaks in favour of analysing (A) as containing two independent arguments for the same conclusion, a simple one and a multi-step argument:

(Argument from the creature)   [1]   [3]
                                      $\overline{\quad\quad[2]\quad\quad}$

(Argument from the creator)   [4]
                              $\overline{[5]}$ [6]
                                  $\overline{\quad[7]\quad}$ [8]   [9]   [10]
                                        $\overline{\quad\quad[2]\quad\quad}$

The reconstruction of these two arguments provides a typical example of how argument analysis is intertwined with formalizing and validity proofs. Identifying the individual inferences and how they fit together was largely based on investigating relations of valid inference between prospective formalizations and candidates for additional premises. It seems exceedingly difficult to come up with the suggested reconstruction of (A) if one should strictly follow the linear procedure of figure 1 and try to complete the argument analysis without any help of formalizations and validity proofs.

6.3.   Justifying the reconstruction

So far, the reconstruction of the excerpt of *De casu diaboli* has revealed a surprising variety of commitments and conflicts. But the process of reconstruction I described also illustrates that further

---

[47] It also implies that claims such as "There is exactly one thing which is identical with the maker", formalized as $\exists !x(x=a)$, are logically rather than factually true.

commitments may be introduced which more easily go unnoticed. Formalizing the *Argument from the creator* using *Hxyz* throughout presupposes, among other things, that the occurrences of the verb "have" do not give rise to a fallacy of equivocation. What Anselm exactly means by *habere* is irrelevant to the validity of the *Argument from the creator*. But the proposed formalizations assume that it is the same in [5]–[7] and [2]. Consequently, we have to accept this as an additional commitment. Another implicit commitment is generated by using first-order formalizations. In the *Argument from the creature*, for instance, formalizing (I1.2) by (F1.2) avoids the existential commitment of the alternative syllogistic formalization by *Camestres*. Although this is irrelevant to the validity of (I1.2), the two formalizations differ in their truth conditions and hence cannot both be correct.

As these examples show, reaching an equilibrium involves more than accepting formalizations and a proof of validity. Rather, it includes a number of commitments, many of which are seldom made explicit. In fact, if we propose to reconstruct an argument as a particular inference and formalize it in a certain way, we commit ourselves to everything represented in the inference and the formalization as well as to the result of a formal analysis of validity. This means that the argument at hand can be framed in exactly these words; that it has a logical form as represented in the formula, which in turn includes, for example, that we accept any added premises or that there are no equivocations; and that it is valid if there is a proof of validity for the formalization. Only if we accept all these points as current commitments, will we have reached the equilibrium necessary for justifying the argument reconstruction at hand.

Besides such an agreement, justification by *reflective* equilibrium also requires that the formalization we propose does justice to the goals of formalizing mentioned in section 2. As we have seen, this is above all a matter of working with an appropriate logical formalism and less a question of choosing a particular formalization. Nonetheless, one may wonder whether in our example the proposed formalizations could not be replaced by more transparent ones. This, however, seems not possible without resorting to formalizations which do not permit showing the validity of Anselm's arguments (e.g. F3.1) or to less adequate formalizations (e.g. (F1.3) below). As a further requirement for reflective equilibrium, we must also respect our antecedent commitments. This means that the commitments currently resulting from reconstructing an argument as an inference with a certain formalization should be defensible in light of the commitments we had when we started our argument analysis. In our example, we should be in a position to defend the current commitments that are in agreement with the proposed formalizations of [1]–[10] and with the inferences involved in the two arguments (*from the creature* and *the creator*) against the antecedent commitments expressed in the initially suggested inferences (I1.1), (I2) and (I3), in (C1)–(C6), and in the various comments I made in the discussion of the example. In fact, this seems to be rather unproblematic, as we only have to give up (C2), which says that Anselm's argument (A1) has one premise only, and the hypothesis (p. 20) that (A) advances one complex argument.

The resulting justification of the argument reconstruction has the following characteristics. Firstly, it is a justification of *both* the formalizations we arrive at and the current commitments that are in agreement with it, specifically, commitments about what the premises and the conclusions of the arguments are, about their logical forms and about the arguments' validity. It is important to note that we may also have antecedent commitments that cannot be justified by argument reconstruction alone and lie outside the scope of the discussed reflective equilibrium. An example is the presumption that Anselm's arguments are not only valid but sound. Secondly, the resulting justification is relative to antecedent commitments (including (I1.1), (I2), (I3), (C1)–(C6) and those specified in the discussion above) as well as to pragmatic-epistemic goals as discussed in section 2. If the initial commitments are mistaken, so may be the resulting reconstruction. For example, one could possibly argue that "by no means can" in (A2) is part of the conclusion, not of a conclusion indicator. This would be a reason to criticize (I2) and the formalizations of [5] for lacking an expression of modality. Thirdly, justification is pluralistic. There is in general no such thing as *the one right* argument reconstruction.[48] In dealing with (A1) we could, for example, give more weight to a transparent representation of truth conditions and go for a formalization which is equivalent to but less adequate (with respect to surface rules) than (F1.2):

(F1.3)   [1] $\forall x(Ax \rightarrow Sx)$ correspondence scheme as in (F1.1)
           [3] $\forall x(Sx \rightarrow \neg Cx)$
           [2] $\overline{\forall x(Ax \rightarrow \neg Cx)}$

---

48 Of course, there may be cases in which we have good reason to let our reconstruction be guided by the assumption that there is only a very limited range of acceptable reconstructions. We know, for example, that law makers typically strive for formulations that do not admit of plausible alternative readings, albeit with limited success.

Or we could insist that using first-order logic in a reconstruction of Anselmian arguments is anachronistic, opt for syllogistics and a formalization with *Camestres*. The alternative reconstructions reflect different weights assigned to the relevant commitments and to the different goals guiding the argument reconstruction. (F1.3) certainly fares better in an exploitative than in a strictly exegetical reconstruction. A similar point can be made about the formalization of definite descriptions in [4]-[9]. Using an individual constant together with the uniqueness condition [8] may be defended in an exegetical setting (if we ignore the problem mentioned in note 47) since it is plausible that in (A) Anselm presupposes rather than asserts the existence of what he describes as "the maker". In an exploitative reconstruction, however, one may want to avoid "creationist" presuppositions and consequently opt for a Russellian formalization.[49]

Finally, a justification by reflective equilibrium will also help to enforce the hermeneutic principle of text fidelity. It does so in two ways corresponding to two readings of the principle. For one thing, the requirement of respecting antecedent commitments rules out that we completely ignore our initial interpretation of the text, which is not yet the product of our argument reconstruction. For another, an equilibrium will only be reached if we are in fact ready to accept the commitments that result from our current reconstruction and most of these commitments are not merely about inferences or the author's views but claims about how to interpret *the text at hand.*

7.   Concluding remarks

To conclude, I briefly comment on some advantages that the method of reflective equilibrium promises as a framework for reconstructing arguments. First of all, the method addresses in a uniform way the justification of the various steps and theories involved in a reconstruction. This includes the argument analysis and formalization of the argument as well as the theories of validity and formalization which constitute the background against which arguments are reconstructed.

Secondly, in its application to the reconstruction of a specific argumentation, the method provides a framework which integrates the various aspects driving the reconstruction. It encompasses a wide range of commitments and does not just deal with an inference that is to be replaced by a formalization. It also acknowledges the variety of goals guiding a reconstruction. The requirement of representing a logical form of the inference in question is covered by a theory of formalization; the remaining goals – allowing a proof of validity, providing a transparent representation of a logical form, promoting exegetical insights and so on – are accounted for by choosing a logical system and a suitable formalization, and by adjusting commitments and formalizations.

Furthermore, an argument reconstruction as presented in the Anselm case-study includes not just one but possibly many formalizations of successively developed inferences. The method of reflective equilibrium places these formalizations and inferences in a setting in which the justification of the resulting formalization is not only a matter of its adequacy with respect to the corresponding inference, but also depends on its relation to all commitments and goals which guide the reconstruction.

Finally, the method of reflective equilibrium provides the means to account for the feedback-loops that are characteristic of the logical reconstruction of arguments. It is therefore time to give up the picture from section 1 in favour of a more complex account which adds feedback-loops to the various steps in figure 1. Formalization is a crucial element of logical scrutiny of arguments, but replacing an inference by a formalization is only one move in the reconstruction of an argument. Logical reconstruction rather requires us to mutually adjust formalizations and our commitments about the argumentation in the text at hand.

---

[49] Further options for alternative reconstruction may be inspired by the fact that the second conjunct of [7] in (F3.2) implies the first conjunct but is not needed for the proposed derivation of [2] from [7]–[10].

References

DCD Anselm of Canterbury. *De casu diaboli.* In *Opera Omnia,* Schmitt (ed.) 1968. Stuttgart-Bad Cannstatt: Frommann/Holzboog. Vol. 1, 231–76.

Barwise, J./ Etchemendy, J. 2008. *Language, Proof and Logic.* Stanford: Center for the Study of Language and Information.

Baumgartner, M. / Lampert, T. 2008. Adequate Formalization. *Synthese 164,* 93–115.

Betz, G. 2012. On Degrees of Justification. *Erkenntnis 77,* 237–72.

Blau, U. 1977. *Die dreiwertige Logik der Sprache. Ihre Syntax, Semantik und Anwendung in der Sprachanalyse.* Berlin: de Gruyter.

Blau, U. 2008. *Die Logik der Unbestimmtheiten und Paradoxien.* Heidelberg: Synchron.

Borg, E. / LePore, E. 2002. Symbolic Logic and Natural Language. In: Jacquette, D. (ed.). *A Companion to Philosophical Logic.* Oxford: Blackwell. 86–101.

Brun, G. 2004. *Die richtige Formel. Philosophische Probleme der logischen Formalisierung,* 2nd ed. Frankfurt a. M.: ontos.

Brun, G. 2008. Formalization and the Objects of Logic. *Erkenntnis 69,* 1–30.

Brun, G. 2012a. Adequate Formalization and De Morgan's Argument. *Grazer Philosophische Studien 85,* 325–35.

Brun, G. 2012b. Rival Logics, Disagreement and Reflective Equilibrium. In: Jäger, C. / Löffler, W. (eds). *Epistemology: Contexts, Values, Disagreement, Proceedings of the 34th International Ludwig Wittgenstein Symposium.* Frankfurt a.M.: Ontos. 355–68.

Brun, G. 2014. "Reflective Equilibrium without Intuitions?" *Ethical Theory and Moral Practice 17,* 237–52.

Brun, G / Hirsch Hadorn, G. 2009. *Textanalyse in den Wissenschaften. Inhalte und Argumente analysieren und verstehen.* Zürich: vdf.

Brun, G. / Rott, H. 2013. Interpreting Enthymematic Arguments Using Belief Revision. *Synthese 190,* 4041–63.

Castañeda, HN. 1975. *Thinking and Doing. The Philosophical Foundations of Institutions.* Dordrecht: Reidel.

Cheyne, C. 2012. The Asymmetry of Formal Logic. In: Peliš, M / Punčochář, V. (eds). *The Logica Yearbook 2011.* London: College Publications. 49–62.

Chomsky, N. 1986. *Knowledge of Language. Its Nature, Origin, and Use,* Westport: Praeger.

Church, A. 1956. *Introduction to Mathematical Logic. Vol. 1.* Princeton: Princeton UP.

Czermak, J. / Dorn, G. / Kaliba, P. / Nieznanski, E. / Pühringer, Ch. / Zwickl-Bernhard, Ch.. 1982. Über die ersten sechs Sätze der Monadologie. *Conceptus 16,* 89–96.

Daniels, N. [1979]. Wide Reflective Equilibrium and Theory Acceptance in Ethics. In: Daniels 1996, 21–46.

Daniels, N. [1980]. On Some Methods of Ethics and Linguistics. In: Daniels 1996, 66–80.

Daniels, N. 1996. *Justice and Justification. Reflective Equilibrium in Theory and Practice.* Cambridge: Cambridge UP.

Davidson, D. [1973] 1984. Radical Interpretation. In: Davidson 1984, 125–39.

Davidson, D. 1980. The Logical Form of Action Sentences. Criticism, Comment, and Defence. In: *Essays on Actions and Events.* Oxford: Oxford UP. 105–48.

Davidson, D. 1984. *Inquiries into Truth and Interpretation.* Oxford: Oxford UP.

Dorn, G.J.W. 2006. Deskriptive Argumente und Argumenthierarchien. In: Kreuzbauer/Dorn 2006, 167–201.

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

Elgin, C.Z. 1996. *Considered Judgment.* Princeton: Princeton UP.

Elgin, C.Z. 2014. Non-Foundationalist Epistemology. Holism, Coherence, and Tenability. In: Steup, M. / Turri, J. / Sosa, E. (eds). *Contemporary Debates in Epistemology,* 2nd ed. Malden: Blackwell. 244–55.

Epstein, R.L. 2001. *Predicate Logic. The Semantic Foundations of Logic.* Belmont: Wadsworth.

Frege, G. [1879] 1988. *Begriffsschrift und andere Aufsätze,* Darmstadt: Wissenschaftliche Buchgesellschaft. Translated in van Heijenoort, J. (ed.). 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* Cambridge, Mass.: Harvard UP. 1–82. (References are to the original pagination).

Frege, G. [1893/1903] 1998. *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet,* Hildesheim: Olms. Translated in 1964. *The Basic Laws of Arithmetic. Exposition of the System.* Berkeley/Los Angeles: University of California Press.(References are to the original pagination).

Goodman, N. [1952] 1972. Sense and certainty. In: *Problems and Projects.* Indianapolis: Bobbs-Merrill. 60–8.

Goodman, N. 1983. *Fact, Fiction, and Forecast,* 4th ed. Cambridge, Mass.: Harvard UP. [1st ed. 1952]

Guttenplan, S. 1997. *The Languages of Logic. An Introduction to Formal Logic,* 2nd ed. Oxford: Blackwell.

Haack, S. 1978. *Philosophy of Logics.* Cambridge: Cambridge UP.

Haack, S. 1982. Dummett's Justification of Deduction. *Mind 91,* 216–39.

Hahn, S. 2000. *Überlegungsgleichgewicht(e). Prüfung einer Rechtfertigungsmetapher.* Freiburg: Alber.

Hempel, C.G.. [1988] 2000. On the Cognitive Status and the Rationale of Scientific Methodology. In: *Selected Philosophical Essays,* Jeffrey (ed.). Cambridge: Cambridge UP. 199–228.

Henry, D.P. 1967. *The Logic of Saint Anselm.* Oxford: Clarendon Press.

Henry, D.P. 1974. *Commentary on* De Grammatico*. The Historical-Logical Dimensions of a Dialogue of St. Anselm's.* Dordrecht: Reidel.

Hintikka, J. / Bachman, J. 1991. *What If ...? Toward Excellence in Reasoning.* Mountain View: Mayfield.

Horwich, P. 1975. A Formalization of 'Nothing'. *Notre Dame Journal of Formal Logic 16,* 363–8.

Jacquette, D. 1996. Charity and the Reiteration Problem for Enthymemes. *Informal Logic 18,* 1–15.

Kamitz, R. 1979. Was kann die Anwendung der formalen Logik auf eine Wissenschaft zum Fortschritt der Erkenntnis beitragen? In: Kamitz, R. (ed.). *Logik und Wirtschaftswissenschaft.* Berlin: Duncker und Humblot. 21–128.

King, P. 2003. Anselm's Philosophy of Language. In: Davies, B. / Leftow, B. (eds). *The Cambridge Companion to Anselm.* Cambridge: Cambridge UP. 84–110.

Kleene, S.C. 1952. *Introduction to Metamathematics.* Amsterdam: North-Holland.

Kreuzbauer, G. / Dorn, G.J.W. (eds). 2006. *Argumentation in Theorie und Praxis. Philosophie und Didaktik des Argumentierens.* Wien: Lit.

Lampert, T. / Baumgartner, M. 2010. The Problem of Validity Proofs. *Grazer Philosophische Studien* 80, 79–109.

Lemmon, E.J. 1987. *Beginning Logic,* 2nd ed. London: Chapman and Hall.

LePore, E. / Cumming, S. 2009. *Meaning and Argument. An Introduction to Logic through Language,* 2nd ed. Chichester: Wiley-Blackwell.

Lewis, D. [1970] 1983. General Semantics [with Postscripts]. In: *Philosophical Papers,* Vol. I. New York: Oxford UP. 189–232.

Löffler, W. 2006. Spielt die rhetorische Qualität von Argumenten eine Rolle bei deren logischer Analyse? In: Kreuzbauer/Dorn 2006, 115–30.

Massey, G.J. 1975. Are there Any Good Arguments that Bad Arguments Are Bad? *Philosophy in Context 4,* 61–77.

Montague, R. 1974. *Formal Philosophy. Selected Papers.* New Haven: Yale UP.

Morscher, E. 2009. *Kann denn Logik Sünde sein? Die Bedeutung der modernen Logik für Theorie und Praxis des Rechts.* Wien: Lit.

Peregrin, J. / Svoboda, V. 2013. Criteria for Logical Formalization. *Synthese* 190, 2897–924.

Peregrin, J. / Svoboda, V. Forthcoming. Logical Formalization and the Formation of Logic(s).

Petrus Lombardus. *Sententiae in IV libris distinctae,* Brady (ed.) 1971. Grottaferrata: Collegi S. Bonaventurae ad Claras Aquas.

Prawitz, D. 1977. Meaning and Proofs. On the Conflict Between Classical and Intuitionistic Logic. *Theoria* 43, 2–40.

Quine, W.V.O. 1960. *Word and Object.* Cambridge, Mass.: MIT Press.

Quine, W.V.O. 1970. Methodological Reflections on Current Linguistic Theory. *Synthese 21,* 386–98.

Quine, W.V.O. 1982. *Methods of logic,* 4th ed. Cambridge, Mass.: Harvard UP.

Rawls, J. 1999. *A Theory of Justice,* revised ed. Cambridge, Mass.: Belknap Press.

Rawls, J. [1975] 1999. The Independence of Moral Theory. In: *Collected Papers.* Cambridge, Mass.: Harvard UP. 286–302.

Rescher, N. 2001. *Philosophical reasoning.* Malden: Blackwell.

Resnik, M.D. 1985. Logic: Normative or Descriptive? The Ethics of Belief or a Branch of Psychology? *Philosophy of Science 52,* 221–38.

Rosenberg, J.F. 1994. *Beyond Formalism. Naming and Necessity for Human Beings.* Philadelphia: Temple UP.

Russell, B. 1905. On Denoting. *Mind 14,* 479–93.

Sainsbury, M. 2001. *Logical Forms. An Introduction to Philosophical Logic,* 2nd ed. Oxford: Blackwell.

Scheffler, I. 1963. *The Anatomy Of Inquiry. Philosophical Studies in the Theory of Science.* New York: Knopf.

Scholz, O.R. 2001. *Verstehen und Rationalität. Untersuchungen zu den Grundlagen von Hermeneutik und Sprachphilosophie,* 2nd ed. Frankfurt a. M.: Klostermann.

Stein, E. 1996. *Without Good Reason. The Rationality Debate in Philosophy and Cognitive Science.* Oxford: Clarendon Press.

Strawson, P.F. 1952. *Introduction to Logical Theory.* London: Methuen.

Sweeney, E.C. 2012. *Anselm of Canterbury and the Desire for the Word.* Washington D.C.: The Catholic University of America Press.

Tennant, N.W. 1997. *The Taming of the True.* Oxford: Clarendon Press.

Uckelman, S.L. 2009. Anselm's Logic of Agency. *Logical Analysis and History of Philosophy 12,* 248–68.

Vorobej, M. 2006. *A Theory of Argument.* New York: Cambridge UP.

Wittgenstein, L. [1921] 1989. *Logisch-philosophische Abhandlung. Tractatus logico-philosophicus. Kritische Edition.* Frankfurt a. M.: Suhrkamp. Translated as 1922. *Tractatus Logico-Philosophicus.* London: Routledge and Kegan Paul.

Wittgenstein, L. [1953] 2003. *Philosophische Untersuchungen,* Frankfurt a. M.: Suhrkamp. Translated as 2009. *Philosophical Investigations,* revised 4th ed. Malden: Blackwell.

Wittgenstein, L. 1979. *Notebooks 1914-1916,* 2nd ed. Oxford: Blackwell.

Wittgenstein, L. 1980. *Lectures, Cambridge 1930–1932.* Totowa, NJ: Rowman and Littlefield.