Synthese
https://doi.org/10.1007/s11229-017-1662-y

Logical expressivism, logical theory and the critique of inferences

Georg Brun[1]

Received: 11 August 2017 / Accepted: 20 December 2017

Abstract The basic idea of logical expressivism in the Brandomian tradition is that logic makes inferential relations explicit and thereby accessible to critical discussion. But expressivists have not given a convincing explanation of what the point of logical theories is. Peregrin provides a starting point by observing a distinction between making explicit and explication in Carnap's sense of replacing something unclear and vague by something clear and exact. Whereas logical locutions make inferential relations explicit within a language, logical theories use formal languages to explicate inferential roles and meanings of ordinary-language expressions. But Peregrin also holds that the whole point of logical theories is to provide perspicuous models of inferential structures in ordinary language practice. This turns explication into a mere continuation of making explicit by other means, and it leads to a one-sided conception of logic which has no room for evaluating inferential practice in light of logical theories. As a more convincing alternative, I suggest that expressivists rely on the method of reflective equilibrium. This approach is closely related to Carnapian explication, but it has the potential of correcting informal inferential practice without dubious ambitions to replace ordinary languages by logical formalisms.

Keywords Explication · Expressivism · Inferentialism · Reflective equilibrium

✉ Georg Brun
Georg.Brun@philo.unibe.ch

1 Institute for Philosophy, University of Bern, Länggassstrasse 49a, Postfach, 3000 Bern 9, Switzerland

Published online: 03 January 2018

1 Introduction

The idea that logic makes inferential relations explicit lies at the heart of the expressivist approach to logic as it has been defended by Brandom since *Making It Explicit* (1994) and elaborated by Peregrin, especially in his *Inferentialism. Why Rules Matter* (2014a). Expressivists consider this idea to be an important part of a genuinely critical enterprise. Once the inferential relations we endorse in practice have been made explicit, they become accessible to critical discussion; we can start to argue about them, call them into question, or back them up with reasons.

This paper analyses two methodological pillars of logical expressivism which I will distinguish with the help of the labels "explication" and "making explicit". My goal is neither to defend nor to attack the expressivist project, but rather to point out some specific shortcomings of available expressivist accounts and to suggest remedies which I think expressivists should welcome. On the negative side, I show that expressivists have not paid enough attention to the methodological basis of their project and as a consequence end up with an implausibly one-sided view of logic. Expressivism, it seems, holds that the *whole* point of having logical theories is that they make something explicit. But this also means that logical theories as such have no critical potential that could be used in an evaluation of the inferential practice we have made explicit. On the positive side, I argue that this is not an unavoidable consequence. Expressivists who are ready to invest in a more ambitious methodological basis can rely on a substantial contribution of logical theories to the critical project they want to foster. Taking Peregrin's work as a starting point, I argue that the method of reflective equilibrium provides an effective basis for a convincing conception of logic which is in line with the basic tenets of expressivism.

Section 2 draws attention to the problem that Brandom and most authors who make use of his ideas fail to make sufficiently clear what exactly expressivism with respect to logic amounts to. The reason for this lack of clarity is that logical expressivists frame their core claim—logic explicates—in ways that admit different readings, depending on whether "logic" refers to something operative in a given language or to a logical theory. In this respect, Peregrin's work provides a significant improvement since, without discussing the point at length, Peregrin often observes a distinction between making something explicit within a language and explicating something with the help of a logical theory. Section 3 analyses this distinction.

Against this background, it becomes conspicuous that the available expressivist approaches to logic emphasize making explicit to a degree that leads to an unconvincingly one-sided understanding of what the point of logical theories is. As I will argue in Sect. 4, even Peregrin's understanding of explication effectively relegates logical theory to the continuation of making explicit by other means, in a way that does not take seriously the idea that logical theories play a significant role in the critical evaluation of our practice of inferring. Peregrin's conception of explication also shares the limitations of Carnap's classical account, which explains explication as a method that deals with individual concepts, not as a method for developing entire logical theories. Given these shortcomings, I argue in Sect. 5 that adopting the method of reflective equilibrium is a promising move, which fits well with Peregrin's approach to logic. Section 6 finally shows how this methodological basis underpins the view that logical

theories can play a substantial role in the critical evaluation of inferential practices without giving rise to a problematic program of language reform.

2 An unclarity in logical expressivism

The basic idea of the expressivist understanding of logic is that logical expressions serve the function of making inferential relations explicit. Although this gives us an important clue as to how expressivists understand logical theories, it does not give us a complete explanation if only because logical theories do not boil down to mere collections of logical expressions. In fact, it is far from clear what the role of logical theories is supposed to be in the expressivist programme. One reason for this unclarity is that the expressivist literature does not sufficiently explain how exactly key terms such as "logic", "logical vocabulary", "making explicit" and "explication" are to be understood.

To begin with, "explication" is a well established term in the philosophy of logic, usually referring to a method of which Carnap gave the classical description (Carnap 1962:§§2–3). A typical example is the claim that a proof-theoretical definition of "$\vdash$" provides an explication of the concept of valid inference. The expressivist literature, however, uses "explication" not in Carnap's sense but as a terminological variant of "making explicit" and "expressing". This becomes obvious in the many instances in which we find formulations such as "[Philosophy's] task is an *expressive, explicative* one" (Brandom 2001: p. 91) and:[1]

> I understand the task of philosophers to have as a central element the *explication* of concepts—or, put slightly more carefully, the development and application of *expressive* tools with which to *make explicit* what is implicit in the use of concepts. […] *Explication, making explicit,* is not the same as analysis […] (Brandom 2001: p. 77; italics changed)

However, as far as I know, expressivists never explicitly differentiate their use of "explication" from the standard, Carnapian, sense of this term. That they use "explication" and "making explicit" interchangeably may also be overlooked because expressivists employ the unusual form "explicitate" as an additional synonym for "make explicit" and "express" (see, e.g. Brandom 1994: p. 82, 110, 116; 2000: pp. 86–87). This may lead to the misunderstanding that they want to distinguish "explication" from "explicitation". But this cannot be the case given that they equate both "explicate" and "explicitate" with "make explicit".

This unclarity is multiplied by the fact that expressivists often speak about logic and logical vocabulary in a way that leaves open whether they want to deal with certain structures (e.g. inference relations) or a certain vocabulary (e.g. "and", "if … then") in a given ordinary language, or with logical theories (e.g. zero-order or first-order

[1] Occasionally, Brandom seems to use "explication" not merely as an alternative to "making explicit" but in a sense that comes somewhat closer to Carnap's, e.g. when he writes that some specific social practice may be "explicable in wholly behavioristic terms" (Brandom 2008b: p. 210). But such passages provide no systematic answer to the issues I raise in this paper.

logic) or certain expressions in formal languages (e.g. "∧", "→").[2] In many instances, clearly the former is at issue:

> But *logic* can also be thought of in expressive terms […] a central expressive resource […] is provided by basic *logical vocabulary*. In applying the concept *lion* to Leo, I implicitly commit myself to the applicability of the concept *mammal* to him. If my language is expressively rich enough to contain *conditionals,* I can say that *if* Leo is a lion, *then* Leo is a mammal. (Brandom 2000: pp. 19–20; italics changed)

Here, logic is understood as being constituted mainly by logical vocabulary that comprises ordinary language expressions such as the conditional "If … then", which Brandom interprets as expressing inference relations operative in ordinary language. But when Brandom discusses, for example, Frege's early conception of logic (Brandom 1994: pp. 107–116), "logic" also refers to the formal theory Frege develops in the *Begriffsschrift* and "logical vocabulary" to expressions in Frege's formal language, especially conditionals such as

Consequently, it becomes unclear whether "make explicit" and the synonyms "explicate" and "explicitate" invariably refer to something that is done within one language (as expressivists declare; see below), or whether these verbs also cover situations in which the vocabulary of one language is used to express something that is either implicit in the use of some other language or explicitly expressed in that other language. Here is a passage which evokes the latter type of situation:

> Frege's overall project for his Begriffsschrift is to use conditionals to make it possible to say explicitly what the inferential role of ordinary, nonlogical concepts is. Where, as he thinks is often the case in natural language, the content expressed by words is unclear, the project of expressing them explicitly will show where they need or can use clarification. The project is the rectification of concepts: clarifying them by explicitating their contents." (Brandom 1994: p. 109; italics changed)

All this raises the question of how expressivists can give us a convincing explanation of what the point of logical theories is. If making explicit is something that is done within one language, logical theories cannot make explicit something implicit in another language. So we are faced with the question of how logical theories which use formal languages relate to inferential structures and logical vocabulary in ordinary language. As far as I can see, Brandom never addresses this issue. It therefore might seem more promising to go for the other option and claim that logical theories employ formal

[2] In this paper, I use "ordinary language" to refer to the language in which the practice the expressivists wants to study is couched. Typically, this is a natural language such as English or Finnish used for everyday communication, but it can also be a more technical variant of such a language used for some specialized, often scientific, purpose. "Formal" is used to contrast formal theories with informal accounts, and formal languages with the vernacular, but not to refer to (theories of) formal rather than material inferences (see Brun 2004: p. 24, pp. 38–40 for a distinction of various uses of "formal").

languages to make explicit inferential structures in ordinary language use or meanings of ordinary-language logical vocabulary. But as I will show in Sects. 3 and 4, such an account blurs important distinctions and leads to an implausibly one-sided picture of the role logical theories can play in the expressivist project.

One might object to the diagnosis given in this section, that logical theories, in contrast to logical vocabulary, are simply not on the expressivist research agenda, and neither is the relation of formal systems of logic to ordinary language. However, even if this were so, the questions "What is the point of logical theories? And how do they relate to ordinary language?" remain important, not the least for expressivists, as I will argue in what follows.

3 Distinguishing explication from making explicit

Within the expressivist camp, we can find at least one philosopher to whom some of the complaints just raised do not apply. Peregrin concurs with the diagnosis that Brandom is unclear with respect to the relation of logical theories to ordinary languages (2008a: pp. 111–112). In most of his own writings, Peregrin does not use "making explicit" and "explication" interchangeably,[3] and he discusses the transition from logical vocabulary in ordinary language to expressions in a logical system. His version of logical expressivism is therefore a good starting point for sorting out the unclarity pointed out in the preceding section.

For an analysis of Peregrin's use of "explicate" and "make explicit" (and "explicitate"; see 2008a: p. 119n17), it is best to start with paradigmatic cases. The central uses of "making explicit" are to be found in statements which say that logical vocabulary (e.g. "and" and "something") makes inferential relations between ordinary-language sentences explicit, or that sentences using logical vocabulary can make explicit inferential rules that are in force in an ordinary language. The fact that we endorse, for example, the inference from "Tigger is a tiger" to "Tigger is a striped animal" can be made explicit with a conditional as "*If* Tigger is a tiger, *then* Tigger is a striped animal" (Peregrin 2008a: p. 111), and the rule governing such inferences may be expressed explicitly as "*If something* is a tiger, *then it* is a striped animal" (Peregrin 2014a: p. 26).[4]

In paradigmatic instances of explication, on the other hand, a theoretical framework (e.g. formal inference rules or formal semantics) is used to account for inferential roles and meanings of ordinary-language expressions. For example, logical constants such as "$\wedge$" and inference rules in a formal system, for example "$A \wedge B \vdash A$", explicate inferential patterns governing the logical vocabulary of an ordinary language (e.g. "and") and thereby the meaning of these expressions. And notions such as *inferability* and *consequence* are explicated by abstract relations (e.g. "$\vdash$").[5]

[3] As I will point out in the next section, Peregrin does not consistently observe the distinction I analyse in this section.

[4] The claim that conditionals are suitable for this function is contested (see, e.g. MacFarlane 2008), but this does not affect the points I am going to make.

[5] Peregrin sometimes describes such explications as regimentations (e.g. Peregrin 2014a: p. 205, 212; see also Peregrin and Svoboda 2017).

Peregrin links explication also to Carnap's method of explication, which he characterizes as the deliberate "replacement of a pre-formal, fuzzy, and unclear notion by a formal and precise concept" (Peregrin 2014a: p. 251n8; see also p. 60). Somewhat more precisely, Carnapian explication can be characterized as a process of replacing, for some theoretical purpose, a concept (the "explicandum") by a more exact concept (the "explicatum"), which is explicitly introduced into the system of concepts of a target theory. The adequacy of an explication is a matter of whether the explicatum is similar to the explicandum and useful for the theoretical purpose at hand. Specifically, the explicatum is expected to be exact ("rules for its use […] are given in an exact form"), fruitful ("useful for the formulation of many universal statements") and simple (Carnap 1962: p. 7).[6]

If we take these points together, the difference between explication and making explicit can be explained as follows. In the paradigmatic instances mentioned, explication takes us from ordinary language to a theoretical framework, whereas making explicit is made within ordinary language. The general point, underlined by Peregrin (2008b: p. 269) and Brandom (2013: p. 101n14; 2015), is that "making explicit" refers to making inferential relations that hold in a given language explicit *in that very same* language, be that a natural language or, for example, Frege's *Begriffsschrift* (Peregrin 2008b: p. 269).[7] Explication, in contrast, usually involves a transition from one language to another.[8] This difference alone provides sufficient reason to adopt a terminological distinction between "explication" and "making explicit" (as I will do for the rest of this paper), but there are further differences. Firstly, the key role of the explicit is different. Explication *uses explicit rules* for introducing a concept for which an explicit expression is already available (see Carnap 1962: p. 3). Making explicit *introduces an explicit expression* for something that has been implicit so far, but the new expression need not be introduced explicitly. "And", for example, makes explicit and hence explicitly expresses an inferential role without having been introduced explicitly, whereas explicit rules (e.g. rules of inference or semantical rules) for using "$\wedge$" must be given to introduce it as an explicatum for the explicandum "and". Secondly, there is a difference in their immediate purpose. Explication

[6] This is how Carnap is usually interpreted and I assume that Peregrin shares this interpretation. In Brun (2016), I argue that this interpretation is problematic and develop a more pragmatic reading of Carnap.

[7] When Peregrin deals with making explicit within formal languages, he is sometimes easy to misunderstand as saying that ordinary-language inferences are made explicit with the help of a logical theory. One example is his (2014a: ch. 9) discussion of how *inferability* can be made explicit by introducing an operator ("$\triangleright$") into a standard theory of formal logic. Since Peregrin writes about inferability and discusses a formal expression ("$\triangleright$"), this is easily read as if this operator would explicate what "inferability" means in English. But this is not the idea. "$\triangleright$" is meant to make explicit the relation of *inferability in the formal system* (which Peregrin represents by "$\vdash$"). Inferability in an ordinary language is not at issue. This is difficult to see, firstly, because making explicit is performed within a formal system, whereas in the paradigmatic cases making explicit is performed within an ordinary language. Secondly, since Peregrin has to speak about what is implicit (namely inferability in the formal language), he explicitly refers to the implicit (by "$\vdash$"). This may lure the reader into assuming that the implicit is not what the formal expression "$\vdash$" refers to, but something else; inferability in English is then the obvious candidate.

[8] It is less clear whether Peregrin thinks that explication *always* involves a transition from one language to another, which will also depend on what he takes to be criteria for individuating languages. But these issues need not be addressed for present purposes. I only rely on the hopefully uncontroversial assumption that in the paradigmatic instances mentioned, explication involves a transition from one language to another.

is a method of improvement for theoretical purposes; it aims at progressing from the vague, unproductive and complicated to the exact, fruitful and simple. The immediate goal of making explicit is not improvement but coming up with an expression that explicitly captures something implicit (which, as expressivists hold, serves the further purpose of making it possible to argue about what has been made explicit). Consequently, explication and making explicit do not have the same conditions of adequacy. Explication aims at replacing a given concept by a concept that is more exact, fruitful and simple and thereby theoretically more useful. Such improvements have a price in terms of similarity: explicandum and explicatum cannot be used in exactly the same way, typically because they are extensionally different. In contrast, the goal of making explicit is just explicitness: to express explicitly that which has been implicit so far. The similarity, or even identity, of the explicit to the implicit takes therefore overriding priority.[9] Making explicit in itself does not involve correcting or improving upon the implicit. Specifically, improving exactness does not play the important role it typically does in explication.

4 Logical theories and the expressivist agenda

Once explication has been distinguished from making explicit, we can look more closely into the question: what is the point of logical theories, especially formal theories? And more specifically: what is the purpose of logical explications of ordinary languages expressions? Given the explanations in the preceding section, one might think that the answer is straightforward: explications are meant to provide more exact, fruitful and simple 'counterparts' to ordinary language expressions. But the questions remain: what is the point of having such counterparts? What are logical theories good for?

Peregrin's answer is that the point of having logical explications is that they provide perspicuous models of ordinary language; they are not meant to improve on ordinary language:

> Another way [the one Peregrin endorses] to see the logical vocabulary of the languages of logic, and indeed the whole logical languages, is as simplified and idealized models of natural language. Such models disregard many features of natural language and thus reach an idealized form that is, though not more functional, more perspicuous. […]
>
> I feel suspicious about endeavors to really 'improve' on natural language […], I am convinced that language, having been formed by natural selection, is more

[9] Most explanations of making explicit seem to simply take what is explicit to be identical with, not only similar to, what has been implicit so far. At least this is the most natural interpretation of passages like "The expressive role of the conditional is to make *explicit,* in the form of a claim, what before was *implicit* in our practice of distinguishing some inferences as good." (Brandom 2000: p. 81) or "[Logic provides] the expressive tools permitting us to endorse in what we say what before we could endorse only in what we did." (Brandom 1994). One might question whether identity is not too strong a requirement, but this point need not concern us here.

> perfect than we can make it via our engineering. (Peregrin 2014a: p. 205; see also 1995: p. 215; 2008a: p. 101, 112)

It is certainly plausible that (at least some) logical theories provide simplified and idealized ("Galilean") models which give a perspicuous view of some inferential structures operative in natural language.[10] But Peregrin apparently takes this to be their overriding or even exclusive goal. He thereby turns explication into a continuation of making explicit by other means. It is a continuation *of making explicit,* because by presenting logical languages as having the function of making something perspicuous about an ordinary language, Peregrin presents them as contributing to the same ends as making explicit, namely to provide an explicit expression that captures a feature of a practice that is couched in ordinary language. It is a continuation *by other means*, firstly, because logical explications involve a transition from ordinary language to a logical formalism, whereas making explicit is done *within* a language; secondly, because making explicit does not include the idealizations and simplifications which logical languages are expected to realize; and, finally, because Peregrin holds that the wffs of the so-called logical 'languages' are schematic and hence "$p \rightarrow q$", for example, does not have a meaning, at least not in the same sense as "If this is a tiger, then it is a striped animal" does (Peregrin and Svoboda 2017, e.g. ch. 4.1).

Peregrin's tendency to emphasize perspicuous models is unfortunately so strong that he is not consistently loyal to his own official explanation of making explicit as something that is done within a language. There are passages in which he describes the relation between (expressions of) ordinary language and (expressions of) formal languages of logic in terms of "making explicit", thereby blurring the distinction analysed in the preceding section.[11]

Moreover, according to Peregrin, providing perspicuous models is not merely one important function of the step from ordinary language to logical theories. It constitutes the very reason why we may turn to logical theories and delimits what we can legitimately do by means of logical explications:

> The purpose of formalization is to help us see certain aspects of language and its functioning clearer, to achieve what we might call, after Wittgenstein […], an *übersichtliche Darstellung* ('perspicuous representation')[.] It is justified *to the extent, and only to the extent* that it fulfils this function; and it must be constantly evaluated from this point of view. (Peregrin 2014b: p. 209; italics in the second sentence GB)

This dovetails with Peregrin's scepticism about improving on natural language (expressed in the previous quote). But what kind of "improvements" is Peregrin suspicious about? Certainly, he does not hold that, given a particular ordinary-language argument, logic can only make the inference relation perspicuous, but must not evaluate or criticize the inference. Peregrin defends a *normativist* expressivism and this means

[10] On idealizing models in logic, see also Peregrin (2001: p. 189, 245), Hansson (2000) and Sainsbury (2001: ch.6.1). For a critical discussion of the view that formal logical languages are models *of natural language,* see Dutilh Novaes (2012: esp. ch. 3.3.1).

[11] e.g. Peregrin and Svoboda (2013: p. 2916, 2919; 2016: p. 64, 68, 70, 77; 2017: p. 4, 69, 71, 96, 137, 140, pp. 147–148, 160).

that one can be criticized for not following the rules (Peregrin 2014a: p. 206). What Peregrin's expressivist programme (in his 2014a) excludes is criticizing and possibly revising the logical standards embodied in ordinary-language practice of inference. If he is opposed to criticizing particular ordinary-language arguments, then only if and because they are sanctioned by these standards.

However, such a conception of logic is too one-sided, also from an expressivist point of view. According to the expressivist project, making explicit serves the goal of making it possible to give and ask for reasons: "only what is explicit can be assessed, discussed and possibly also modified or rejected" (Peregrin 2014a: p. 203; see also p. 187; 2008b: p. 269). The expressivist sees logic in the service of the aims of enlightenment. She seeks to bring our practice within reach of arguments, so that she can evaluate the practice, potentially criticize it and suggest improvements.[12] But making explicit is only a necessary condition for critical evaluation. By itself, it has no critical potential. This point becomes very clear in Brandom's discussion of the critical potential of his expressivist philosophy (e.g. Brandom 1994: p. 127; 2000: p. 81; 2001: p. 77; see also Lance and Philip 1994: pp. 373–375). And the above analysis has shown that essentially the same is true of logical explication in the sense Peregrin understands it. Providing perspicuous models may be very useful as a basis for critical evaluation, but in itself it has no critical potential.

We therefore face the question of what additional resources we may have for a critical evaluation of our practice of inferring. What should we take as a basis for arguing about inference relations? Traditionally, logic is meant to play a key role. The idea of using logic as a critical tool, known as the tradition of an *ars iudicandi,* has been a driving force of the project of developing logical theories since its inception. If our logical tools deem certain inferences valid and some of our commitments inconsistent, this should give us reason to accept these inferences as valid and to weed out the inconsistencies. In particular, logical theories give us reason to accept patterns of inference by proving their validity, to reject patterns of inference by proving their invalidity, and (in non-paraconsistent contexts) to accept the principle of non-contradiction. With respect to specific inferences, logical theories give us reason to accept inferences proved to be valid, to reconsider inferences instantiating invalid patterns of inference and to give up at least one of a set of inconsistent claims.[13] Such critical functions of logical theories go substantially beyond making explicit and providing perspicuous models, either because they directly challenge patterns of inference that have been made explicit, or because some calculating, not merely making explicit or perspicuous, is necessary to arrive at a verdict of (in)validity or inconsistency.

Nonetheless, expressivist accounts of logic leave critical functions of logical theories completely out of the picture. They are, for example, simply not to be found on Brandom's agenda. But I see no reason why expressivists should cling to the idea that logic's sole job is a continuation of making explicit by other means, instead of granting that logical theories have critical functions in addition to the modelling function Peregrin assigns to them. And in fact, I doubt that Peregrin, and perhaps other

[12] See also Brandom's comments on Sellars's conception of the Socratic method (Brandom 1994: p. 106).

[13] On the asymmetrical role of valid and invalid patterns of inference, see Cheyne (2012).

expressivists as well, intend to reject critical functions of logical theories altogether, despite their emphasis on providing models of a practice couched in natural language as a central goal of logical theories.

Nor do I think that insisting on critical functions of logical theories is ruled out by Peregrin's reservations about improving on natural language. Insisting on logic being a normative theory with a critical function does not amount to advocating a programme of ideal language which aims at modifying natural languages with the help of logical theories, e.g. by introducing "→" into English, by giving "if … then" the meaning of "→" or by resorting to a formal language for purposes for which natural languages are obviously more effective, such as everyday communication. So we do not need to take issue with Peregrin's scepticism about such ideal-language proposals. Taking the critical role of logic seriously first of all means to insist that logical theories can force us to accept verdicts on validity (and related properties of sentences and arguments). This is not the same as forcing us to switch to another language, although we might have reason to do so for some special purposes. But logic can, and should, give us reason to revise the standards of inference in natural language should they prove to be problematic.

5 Reflective equilibrium for logical expressivists

Realizing the critical role of logic requires us to go beyond the expressivist positions which either do not deal with logical theories in their relation to ordinary languages at all (e.g. Brandom) or understand explication in a way that does not take critical functions of logical theories seriously (e.g. Peregrin's position analysed above). So we need to give another account of the relation between ordinary-language practice and logical theory. This is also necessary for another, independent, reason. Explication as explained by Peregrin does not really fit into the Carnapian paradigm. Carnap's account of explication is limited to dealing with individual concepts,[14] but Peregrin's inferentialist programme aims at coming up with a logical theory and consequently should deal with entire systems of rules, inferences and concepts.

To address these issues, I suggest that we adopt the method of reflective equilibrium. There are several reasons for this move. On the one hand, reflective equilibrium is, historically and systematically, a further development of Carnapian explication, which is tailored to developing systems of concepts and theories.[15] On the other hand, reflective equilibrium is not only a well developed and promising account of a methodology for logical theorizing,[16] but it is, as I will argue, also particularly well suited to an expressivist approach to logic. Indeed, Peregrin defends reflective equilibrium as the method for developing logical theories in his recent collaboration on formalization with Svoboda (Peregrin and Svoboda 2013, 2016, 2017). However, Peregrin

[14] This is true of Carnap's explicit accounts of explication, although Carnap, of course, was engaged in developing entire systems of concepts (see Brun 2016, 2017 for more detail).

[15] I substantiate this claim in Brun 2017.

[16] The idea of applying reflective equilibrium to logic is due to Goodman (1983). See also Brun (2012, 2014a), Peregrin and Svoboda (2016, 2017), Resnik (1985, 1996, 1997, 2004) and Shapiro (2000) on reflective equilibrium in general (Baumberger and Georg 2016; Elgin 1996).

and Svoboda do not systematically explore the consequences of turning to reflective equilibrium for logical expressivism. Before we can do that, we need to look more closely into the method of reflective equilibrium.[17]

The application of the method of reflective equilibrium to logical theories can be characterized roughly as follows. We begin with our commitments to the (in)validity of inferences, the (in)consistency of sets of sentences and other logically relevant properties of inferences or (sets of) sentences. Such commitments can be expressed in an explicit judgement or be implicitly endorsed in practice; their content can be general (e.g. "Conjuncts can be inferred from conjunctions.") or it can concern specific inferences or (sets of) sentences (e.g. "'This is copper' is incompatible with 'This is an insulator'."). Commitments also come in degrees, from the faintly adopted to the firmly held. We then try to find a theory that accounts for these commitments (including e.g. a system of inference rules or a model theory, and some definition of validity). Since this typically leads to discrepancies, we start a process of mutually adjusting commitments and theory. For example, a logician may start with the following specimens of valid (1) and invalid (2) inferences:

(1) Tigger is a tiger. If Tigger is a tiger, he is a dangerous animal. Therefore: Tigger is a dangerous animal.
(2) If Tigger is bouncy, Piglet is scared. Piglet is scared. Therefore: Tigger is bouncy.

The logician then resolves to include *modus ponens* ($\varphi \rightarrow \psi;\ \varphi \vdash \psi$) but not affirming the consequent ($\varphi \rightarrow \psi;\ \psi \vdash \varphi$) in her theory, and consequently arrives at a system of inference rules that includes $\varphi \vdash \neg\varphi \rightarrow \psi$ and thereby sanctions inferences she does not want to accept as valid:

(3) Tigger is a tiger. Therefore: If Tigger is a not tiger, he is a striped animal.

To eliminate the conflict, the logician may either adjust her commitments and accept (3) as valid, or revise her formal rules of inference in a way that eliminates $\varphi \vdash \neg\varphi \rightarrow \psi$, perhaps opting for some system of relevant logic.

This process is successful if we reach a state of reflective equilibrium, which is characterized by five requirements. (i) The commitments and the theory need to be in agreement; this is a requirement of internal coherence, which includes the consistency of the commitments and the theory, as well as that the commitments can be inferred from the theory. (ii) The theory needs to be supported by (or at least be compatible with) background theories. In the case of logic, it is debated which background theories are relevant to this requirement of external coherence. Brandom and Peregrin endorse, for example, an inferentialist philosophy of language as well as specific arguments from learnability (Brandom 2008a: pp. 134–136; Peregrin and Svoboda 2017: p. 163) and conservativity (Brandom 1994: p. 125; Peregrin 2014a: ch 2.7., 8.2). (iii) The theory needs to do justice to epistemic goals that promote its systematicity, making sure that we get a systematic account of logical validity and not merely a list of our commitments adjusted for consistency and re-labelled a "theory". The epistemic goals

[17] In what follows, I will rely on the general account of reflective equilibrium from Baumberger and Georg (2016) and Brun (2014b). What Peregrin and Svoboda (2016, 2017) say about reflective in equilibrium is compatible with this more detailed account.

include general "virtues" of theories such as being constituted by a well-organized and simple system of exactly formulated, broadly applicable principles. More specific to logical theories is the goal of a formalism which permits proofs of validity (and other logical properties) that are rigorous, cognitively transparent and, if possible, amenable to decision procedures. (iv) The resulting position needs to respect antecedent commitments adequately so that we can be sure that we have a theory *of validity*. This prevents the process of mutual adjustments from changing commitments so drastically that we really change the subject. (v) The resulting commitments must have some credibility independent of their agreement with the theory. The last two points play a conspicuous role in expressivist approaches to logic. Expressivists stress (iv) when they insist that logical theories must account for the standards of inference we in fact endorse in our practice. And this insistence is based on the fundamental assumption that these standards in principle establish what correct inference is, thereby securing (v).

The description given so far focuses on how reflective equilibrium can be applied to the development of a formal theory of validity (and other logical properties) which is intended to account for the standards of validity that are operative in the ordinary-language practice of inferring (this is what Peregrin would call "an explication of *inferability* by, e.g. the abstract relation '$\vdash$' in a system of natural deduction"). However, this focus is simplified in crucial respects. The relation of commitments to elements of a logical theory involves a transition from sentences and inferences endorsed in ordinary language to logical formulas and the other way around (what Peregrin would call, e.g. "an explication of an if-then sentence by a formula containing '$\rightarrow$'"). To make this manifest, we can reconsider (1)–(3). The conflict described above arises only if our logician assumes that the if-[then] sentences in (1)–(3) can be adequately formalized as instances of $\varphi \rightarrow \psi$; that is, with the help of a material conditional. And this assumption tacitly involves criteria of adequate formalization. Further possibilities for dealing with the described conflict are therefore available. Instead of adjusting her commitments or the formal rules of inference, our logician may take another look at her assumption that the formalizations with material conditionals are adequate according to the criteria of adequate formalization she (probably tacitly) relies on. If not, the formalizations have to be changed; if yes, she may opt for adapting her theory of formalization. All this shows that developing a theory of validity must go hand in hand with developing a theory of formalization and providing formalizations of inferences and sentences.[18]

In what follows, I will argue that moving from Peregrin's explication to reflective equilibrium as a methodological basis for logic has important consequences for our understanding of logical theories. In particular, it leads to a more convincing view of what the point of doing logic is, and especially of the critical functions of logical theories.

[18] See Brun (2014a) for more details on how to apply reflective equilibrium in a broader setting that includes formalizations and theories of formalization. For general discussions of (theories of) logical formalizations of inferences see, e.g. Baumgartner and Lampert (2008), Brun (2004), Peregrin and Svoboda (2017) and Sainsbury (2001). For extensive references to further literature see Brun (2004, 2014a).

6 Underpinning critical functions of logic

In contrast to explication as Peregrin conceives of it, reflective equilibrium underpins critical functions of logical theories as I described them in Sect. 4. Peregrin and Svoboda acknowledge this as the "normative role of logic",[19] but they emphasize more its limitations and do not analyse its consequences for an expressivist understanding of logic. In what follows, I explore how logical theories can offer a critical potential to the expressivist. Specifically, I want to show that expressivists who hold that logical theories are justified by reflective equilibrium can defend the view that logic serves two closely related aims: providing perspicuous models of inferential practice, and devising theories of inference that may give us reason to revise inferential practice. From the perspective of reflective equilibrium, the first point is secured by the epistemic goals which drive the process of theory development in logic. They call (inter alia) for theories which provide explicit and perspicuous representations of inferential patterns. This accords with Peregrin's picture of logical expressivism (in Peregrin 2014a), which, however, does not include the second aim, the critical function of logic, as we saw in Sect. 4. What hast to be shown, therefore, is that turning to reflective equilibrium makes room for an adequate *expressivist* understanding of logical theories which includes their critical role.

The decisive factor for the critical function of logic is that the move from explication to reflective equilibrium replaces the linear structure that leads from explicandum to explicatum with a process of mutual adjustments of commitments and theory. This means that in some cases, logical theories can constitute a reason to modify inferential practice, while in other cases we had better adapt our logical theory. On the one hand, the process of developing a reflective equilibrium starts with trying to account for the validity commitments we actually have and therefore presupposes that our inferential practice has some legitimacy. And although this does not mean that our logical commitments are immune to revisions, they must be respected by our resulting position in reflective equilibrium. On the other hand, even though developing logical theories with the method of reflective equilibrium starts by representing inferential patterns, it does not end there. Developing a logical *theory* means that we have to devise a coherent and systematic account of valid inference, as explained in Sect. 5. In practice, our commitments to the validity of inferences cannot simply be codified into a theory which is consistent and does justice to the epistemic goals that drive the enterprise of theory development in logic. Rather, we have to proceed by mutual adjustments of commitments and theory, striving for a coherent position that strikes a balance between the epistemic goal of a systematic theory and the need to adequately respect antecedent commitments. We do not invariably alter the theory if it does not mirror pre-existing practice. It may well be better to eliminate mismatches by adapting our commitments in light of the logical theory.

The example (1)–(3) can be used to illustrate various ways how considerations of coherence and systematicity may figure in the critical evaluation of our inferential

[19] Peregrin and Svoboda (2017): p. 64, 91, pp. 102–105 (see also 2016: p. 78, 2013: p. 2900, 2921, 2922).

practice and give us reason to revise it.[20] As we have seen, entertaining a commitment to the validity of (1) and the invalidity of (2) and (3) may easily lead one into an inconsistent position because a system of inference rules which includes *modus ponens* may also sanction the so-called "paradoxes of implication". Let us now consider a situation in which our logician has indeed developed a system of inference rules that sanctions $\varphi \vdash \neg \varphi \rightarrow \psi$. The expressivist will rightly point out that making explicit the commitments is the basis for diagnosing the inconsistency and entering a critical discussion of what to do about it. However, it is important to note, firstly, that such a diagnosis will be available only if we go beyond merely giving a perspicuous representation of commitments by means of formulas. A proof is needed to make sure that the system sanctions $\varphi \vdash \neg\varphi \rightarrow \psi$.[21] Secondly, that this rule can be derived constitutes a (defeasible) reason not only to accept (3) as valid, but also to adjust the inferential standards which stand against the validity of inferences, the formalizations of which are deemed valid by $\varphi \vdash \neg\varphi \rightarrow \psi$. It is, of course, possible that this reason "loses" against the commitment that (3) is invalid and that our logician instead resolves to adjust her system of rules rather than her inferential practice. But even in this case, it is still a reason our expressivist logician must take seriously *for expressivist reasons* because it is grounded in her commitments, which represent the inferential practice she follows.

Furthermore, if we proceed to a critical discussion about how to deal with this situation (as it is part of the expressivist agenda), the perspective of the resulting logical theory will play an important role. Here are some paradigmatic considerations. Simply giving up $\varphi \vdash \neg\varphi \rightarrow \psi$ is not an option. Further amendments to the logical theory are necessary to ensure that $\varphi \vdash \neg\varphi \rightarrow \psi$ no longer follows from other rules. One strategy would be to stipulate straightforward exceptions for inferences such as (3). But expressivists should reject such ad hocery since it undermines the very idea of a critical discussion of inferential practice (besides sinning against simplicity). Better strategies are available if we admit that the original incoherence shows a tension in the relevant commitments. This leaves our logician with the task of deciding whether she should adjust the commitments or the theory (or maybe both). If she takes the first option, the logical theory plays an essential role because it constitutes the reason for adjusting the commitments, as discussed before. If she takes the second option, she will have to come up with an alternative system of rules, perhaps resorting to some relevant or purely implicational logic. But in all likeliness, such a move will either lead to new discrepancies and give her reason to change some other commitments, or it will result in a logical theory that accounts only for a fragment of the relevant commitments and the problems are bound to reappear as soon as our logician tries to develop a system with broader scope—which is something the expressivist should insist on because addressing only a tiny fragment of our validity commitments undermines the

[20] Of course, the critical evaluation of inferential practice can draw on additional resources, for example on arguments that refer to background theories from philosophy of language, general accounts of normativity and so on.

[21] This claim rests on the assumption that our logician has not adopted $\varphi \vdash \neg\varphi \rightarrow \psi$ as a basic rule of inference, which, of course, she has no motivation to do given her commitment to reject inferences such as (3).

expressivist programme of a critical evaluation of *our inferential practice,* not just some unproblematic part of it. In all these arguments, we can observe that it is crucial for the critical evaluation of our inferential practice whether we have developed a position in which our commitments are *coherent* with a *systematic* logical theory.

But can we square this view with Peregrin's scepticism about using (certain elements of) logic as a replacement for (certain elements of) natural language? After all, explication aims at replacement and this feature is also present in reflective equilibrium. And what about the adjustment of commitments in light of the theory? It might seem that this hallmark of reflective equilibrium ultimately calls for replacing (aspects of) natural language by components or products of logical theories.

However, defenders of a reflective-equilibrium-based conception of logic need not deny that natural language "is perfect *with respect to the ends it serves*" (Peregrin 2014a: p. 205). After all, saying that "$\wedge$" explicates "and" means only that "$\wedge$" replaces "and" for the theoretical purpose the explication serves, which is not the original purpose of "and". This is in line with Peregrin's view that expressions of the logical formalism replace the natural-language expressions for the specific purpose of providing a model of inferential patterns, not for any old purpose and specifically not for the communicative ends natural languages are standardly used for. And similarly, I fail to see any reason why defenders of reflective equilibrium should not hold that "$\wedge$" is meant to substitute for "and" in proofs of validity. This does not imply that "$\wedge$" could replace "and" for other purposes or in ordinary discourse. After all, the point of reflective equilibrium is not that commitments are simply replaced by theories, but rather that it justifies both commitments and theories by bringing them into agreement.

Moreover, we must distinguish between improvements in light of logical theories and replacements by elements of logical theories. Consider for example the (maybe not very reasonable) suggestion of improving ordinary-language use of "if …then" by bringing it in line with the connective "$\rightarrow$" in classical logic. This is not a proposal to introduce the expression "$\rightarrow$" into a natural language, but rather to adopt different rules for using "if …then". And even if we think that such an (alleged) improvement would best be described as replacing the "if … then" which is subject to the traditional norms by an "if … then" that is governed by some new rules, this is a replacement within an ordinary language. It would be the deliberate result of some logical theorizing, but in other respects quite similar to conscious language change for the sake of, for example, political correctness or greater precision, which are just part of ordinary developments of ordinary language.

In short, if we rely on the method of reflective equilibrium, developing a logical theory does not only aim at providing a model of inferential practice; by the same token, it means developing systematic norms that may give us reason to revise our logical practice. In this sense, reflective equilibrium is a method which justifies not only logical theories but also the resulting commitments to the validity and other logical properties of inferences or (sets of) sentences. Expressivists should welcome this result. It explains why logical theories not only can provide perspicuous representations of logical relations and properties, but also have a critical potential. In line with expressivist principles, this critical potential is rooted in the fact that logical theories are systematic accounts of inferential standards we are actually committed to in our ordinary language practice. This makes it possible for expressivists to draw on logi-

cal theories in assessing and potentially revising inferential practice without thereby engaging in a dubious attempt at replacing ordinary languages by logical formalisms.

Acknowledgements For helpful comments on earlier versions, I am grateful to Dominique Kuenzle, Jaroslav Peregrin and Christine Tiefensee. This paper is part of the project "Reflective Equilibrium—Reconception and Application", supported by the Swiss National Science Foundation (Project 150251).

References

Baumberger, C., & Brun, G. (2016). Dimensions of objectual understanding. In S. R. Grimm, C. Baumberger & S. Ammon (Eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science* (pp. 165–189). New York: Routledge.

Baumgartner, M., & Lampert, T. (2008). Adequate formalization. *Synthese*, *164*, 93–115.

Brandom, R. B. (1994). *Making it explicit. Reasoning, representing and discursive commitment*. Cambridge: Harvard University Press.

Brandom, R. B. (2000). *Articulating reasons. An introduction to inferentialism*. Cambridge: Harvard University Press.

Brandom, R. B. (2001). Reason, expression, and the philosophic enterprise. In C. P. Ragland & S. Heidt (Eds.), *What is philosophy?* (pp. 74–95). Yale: Yale University Press.

Brandom, R. B. (2008a). *Between saying and doing. Towards an analytic pragmatism*. Oxford: Oxford University Press.

Brandom, R. B. (2008b). Responses. In P. Stekeler-Weithofer (Ed.), *The pragmatics of making it explicit* (pp. 209–229). Amsterdam/Philadelphia: Benjamins.

Brandom, R. B. (2013). Global anti-representationalism? In H. Price (Ed.), *Expressivism, pragmatism and representationalism* (pp. 85–111). Cambridge: Cambrige University Press.

Brandom, R. B. (2015). *Between saying and doing. Interview by Richard Marshall*.http://www.3ammagazine.com/3am/between-saying-and-doing/. Accessed 5 Nov 2016.

Brun, G. (2004). *Die richtige Formel. Philosophische Probleme der logischen Formalisierung* (2nd ed.). Berlin: de Gruyter.

Brun, G. (2012). Rival logics, disagreement and reflective equilibrium. In C. Jäger & W. Löffler (Eds.), *Epistemology: Contexts, values, disagreement. Proceedings of the 34th international Ludwig Wittgenstein symposium, Kirchberg am Wechsel, Austria 2011* (pp. 355–368). Frankfurt A.M.: Ontos.

Brun, G. (2014a). Reconstructing arguments. Formalization and reflective equilibrium. *Logical Analysis and History of Philosophy*, *17*, 94–129.

Brun, G. (2014b). Reflective equilibrium without intuitions? *Ethical Theory and Moral Practice*, *17*, 237–252.

Brun, G. (2016). Explication as a method of conceptual re-engineering. *Erkenntnis*, *81*, 1211–1241.

Brun, G. (2017). Conceptual re-engineering: From explication to reflective equilibrium. *Synthese*. https://doi.org/10.1007/s11229-017-1596-4.

Carnap, R. (1962/1950). *Logical foundations of probability* (2nd ed.). Chicago/London: University of Chicago Press/Routledge and Kegan Paul.

Cheyne, C. (2012). The asymmetry of formal logic. In M. Peliš & V. Punčochář (Eds.), *The logica yearbook 2011* (pp. 49–62). London: College Publications.

Dutilh Novaes, C. (2012). *Formal languages in logic: A philosophical and cognitive analysis*. Cambridge: Cambridge University Press.

Elgin, C. Z. (1996). *Considered judgment*. Princeton: Princeton University Press.

Goodman, N. (1983/1954). *Fact, fiction, and forecast* (4th ed.). Cambridge: Harvard University Press.

Hansson, S. O. (2000). Formalization in philosophy. *The Bulletin of Symbolic Logic*, *6*, 162–175.

Lance, M. N., & Kremer, P. (1994). The logical structure of linguistic commitment I. Four systems of nonrelevant commitment entailment. *Journal of Philosophical Logic*, *23*, 369–400.

MacFarlane, J. (2008). Brandom's demarcation of logic. *Philosophical Topics*, *36*, 55–62.

Peregrin, J. (1995). *Doing worlds with words. Formal semantics without formal metaphysics*. Dordrecht/Boston/London: Reidel.

Peregrin, J. (2001). *Meaning and structure*. Aldershot: Ashgate.

Peregrin, J. (2008a). Brandom's incompatibility semantics. *Philosophical Topics*, *36*, 99–122.

Peregrin, J. (2008b). What is the logic of inference? *Studia Logica*, *88*, 263–294.
Peregrin, J. (2014a). *Inferentialism. Why rules matter*. Basingstoke: Palgrave Macmillan.
Peregrin, J. (2014b). Logic and the pursuit of meaning. In P. Stalmaszczyk (Ed.), *Philosophy of language and linguistics. The legacy of Frege, Russell, and Wittgenstein* (pp. 193–212). Berlin: de Gruyter.
Peregrin, J., & Svoboda, V. (2013). Criteria for logical formalization. *Synthese*, *190*, 2897–2924.
Peregrin, J., & Svoboda, V. (2016). Logical formalization and the formation of logic(s). *Logique et Analyse*, *233*, 55–80.
Peregrin, J., & Svoboda, V. (2017). *Reflective equilibrium and the principles of logical analysis. Understanding the laws of logic*. New York: Routledge.
Resnik, M. D. (1985). Logic: Normative or descriptive? The ethics of belief or a branch of psychology? *Philosophy of Science*, *52*, 221–238.
Resnik, M. D. (1996). Ought there to be but one logic? In B. J. Copeland (Ed.), *Logic and reality. Essays on the legacy of Arthur Prior* (pp. 489–517). Oxford: Clarendon Press.
Resnik, M. D. (1997). *Mathematics as a science of patterns*. Oxford: Clarendon Press.
Resnik, M. D. (2004). Revising logic. In G. Priest, J. C. Beall, & A.-G. Bradley (Eds.), *The law of non-contradiction: New philosophical essays* (pp. 178–194). Oxford: Clarendon Press.
Sainsbury, M. (2001). *Logical forms. An introduction to philosophical logic* (2nd ed.). Oxford/Malden: Blackwell.
Shapiro, S. (2000). The status of logic. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 333–366). Oxford: Clarendon Press.