Non-citable preprint. Forthcoming in *Synthese*.

Reflective Equilibrium in Logic

Ben Martin (University of Padova)

Abstract. Among the areas of knowledge that the method of reflective equilibrium (RE) has been applied to is that of logical validity. According to RE in logic, we come to be justified in believing a (deductive) logical theory in virtue of establishing some state of equilibrium between our initial judgements over the validity of specific (natural language) arguments and the logical principles which constitute our logical theory. Unfortunately, however, while relatively popular, RE with regards to logical theorizing is underspecified. In particular, it's unclear what constitute: (i) the relevant logical "data", (ii) logical theories, so that they can be suitably tested, and (iii) the mechanisms under which such theories are tested. Considering the various options for how to interpret the position, we argue that in order to be workable the advocate of RE about logic must embrace what we call an *operationalised*, *wide* and *communal* version of RE with a *rich* understanding of logical theories. Fortunately for the advocate of RE, there is an available account of logic's epistemology which possesses just these properties, *logical predictivism*. However, equally unfortunately, logical predictivism commits the advocate of RE to certain further claims that they have historically been weary of holding. Consequently, it is unclear whether RE about logic itself is a viable proposal, or rather a similar but distinct epistemology of logic.

Keywords: Reflective Equilibrium· Epistemology of Logic · Methodology of Logic · Logical Evidence · Logical Predictivism

1. Introduction

Although the term "reflective equilibrium" was only later coin by Rawls (1971) when discussing theories of justice, as is well known the process of reflective equilibrium (RE) was first proposed by Nelson Goodman (1983) as a method to justify our deductive (and, subsequently, inductive) theory of logic. Broadly speaking, Goodman's proposal is that we come to be justified in believing certain (deductive) logical principles in virtue of establishing a state of equilibrium between our initial judgements over the validity of particular (natural language) arguments and the logical principles which constitute our logical theories.[1]

Since Goodman's initial presentation of the method, RE has been appropriated by other sub-fields of philosophy and proposed as applying to a whole host of other areas of knowledge, including ethics (DePaul 1993), science (Cummings 1998), and philosophy (Lewis 1983). Thus, according to the widening of the proposal, just as logical theories are justified by establishing some suitable state of equilibrium between our initial judgements over specific cases and the theory, so theories in other areas of knowledge are also justified by establishing a state of reflective equilibrium between the theory and the relevant data.

At this point, there may be some raised eyebrows. After all, without significant evidence to the contrary, it would be surprising if there were *one methodology* which was equally suitable for such a range of research areas, with their distinct subject matters. Despite its advocates (Williamson 2007 & 2017), methodological anti-exceptionalism is a contentious position (Martin & Hjortland 2022). Further, even if we found that a particular methodological proposal was *consistent with* the epistemological standards across this wide range of research areas, we might be sceptical that it achieves this consistency at the cost of a lack of specificity and informativeness. It is all well and good to talk about what (epistemologically or otherwise) a set of research areas share, but we should be equally interested in the specific

[1] How should we conceive of these logical *principles*? Are they rules of inference, generalisations over argument schemata, or another kind of law-like generalisations? We come onto this point in Section 3.2.

epistemological features of a field that differentiate it from others, and it is unclear that simply talking in terms of the *mutual adjustment* of some data and target theory will do the job.

A concern about RE similar to this has previously been voiced by Foley (1993: 128):

> [RE] tells you essentially this: take into account all the data that you think to be relevant and then reflect on the data, solving conflicts in the way that you judge best. On the other hand, it does not tell you what kinds of data are relevant, nor does it tell you what is the best way to resolve conflicts among the data. It leaves you to muck about on these questions as best as you can.

So, if RE is to be informative on the epistemology of a research area, we require further specifics on what constitute: (i) the *data*, (ii) the *components of the theory*, (iii) a *conflict* between the two, and (iv) how to *adequately resolve* such conflicts.

Our particular concern here is with the putative use of RE within the epistemology of *logic*, which following Goodman's initial proposal has been advocated more recently as the most plausible account of logic's epistemology (Brun 2014a; Peregrin & Svoboda 2017; Resnik 1999). In virtue of proposing that we do not have some direct unmediated access to the correct principles of logic, RE is what is known as a *non-foundationalist* epistemology of logic (Martin & Hjortland 2022). This contrasts RE with other competing epistemologies of logic, such as *logical rationalism* (BonJour 1998) and *logical semanticism* (Ayer 1936) which propose that we come to know the correct principles of logic *directly*, whether through a quasi-perceptual form of intellectual seeming or epistemic analyticity, respectively. Thus, for instance, in the case of logical rationalism, we just *non-perceptually see* that the disjunctive syllogism is valid or that the law of excluded middle is true.[2] In contrast, non-foundationalist epistemologies propose that logicians do not have *unmediated access* to the correct logical laws. Rather, in order to discover the correct logical laws, inferences must be made from some relevant data in order to inform and subsequently test logical theories.[3]

Foundationalist epistemologies of logic are known to suffer from significant weaknesses, both in terms of the sources of knowledge they posit and the extent to which they distort actual disagreements between logicians.[4] This failure of foundationalist epistemologies of logic lends some initial credence, at least, to the reflective equilibrium proposal. RE, however, is not the only non-foundationalist epistemology of logic on offer. Other candidates include Quinean (1951) evidential holism, abductivism (Priest 2016; Russell 2015; Williamson 2017), and predictivism (Martin & Hjortland 2021). Thus, the failure of foundationalist epistemologies of logic cannot suffice in and of itself to vindicate reflective equilibrium, any more than it can these other proposals.

---

[2] For a discussion of the various foundationalist epistemologies of logic and their historical motivations, see Martin & Hjortland (2022; forthcoming).

[3] RE is sometimes known as a *weakly* foundationalist account of justification (Baumberger & Brun 2021; Brun 2014b). By this it is meant that, in some cases, we have reason to trust the reliability of certain *data* independently of our justification for the theory that the data are being used to inform. This serves to distinguish RE from a purely coherentist account of epistemology, where data would possess no credibility independent of its cohering with the theory. The operative notion of "foundationalism" here, however, is not the same as in the distinction between (non-)foundationalist epistemologies in the philosophy of logic. In the latter distinction, the question is whether one has unmediated reliable access to *logical principles* (which constitute the theory), and not whether we have reliable access to certain logical data (distinct from the theory). We'll come to the important distinction between data, theory and phenomena in Section 3.1.

[4] For the sake of brevity, we won't rehearse here the various problems with foundationalist epistemologies of logic. For discussions of these various problems, see Martin (2021b), Martin & Hjortland (forthcoming), and Williamson (2007).

This raises two questions: Firstly, how exactly does RE about logic differ from these other non-foundationalist proposals? Secondly, in what sense is it superior to them? Unfortunately, at present, insufficient attention has been given to these questions by advocates of RE.[5]

Answering both of these questions is complicated by the fact there is no one privileged version of RE in the literature. While it is well-recognised that there is a distinction to be made between *narrow* and *wide* versions of RE (Daniels 1979), with latter versions also allowing for independently well-evidenced background beliefs to inform our (logical) theorizing, there are further differences among versions of RE over (i) the *nature of the data*, (ii) the *nature of the informed theories*, and (iii) *how this reflective equilibrium is reached*, which need to be taken into account. Thus, before we can begin to evaluate RE's plausibility compared with other non-foundationalist epistemologies, we must first determine the most reasonable version of RE available, at least with regards to logic's epistemology.

Our primary goal in this paper, then, is to assess the suitability of these distinct variations of RE for logic's epistemology. What arises out of this examination is that, in order for a version of RE about logic to be workable, it must: (i) embrace an *operationalised* understanding of logical data, (ii) admit a *wide* scope of sources of evidence for logic, (iii) use a *rich* account of a (logical) theory, and (iv) admit that the process of logical theorizing operates on a communal (rather than individual) basis. We call these the criteria for an *optimal* version of RE.

The question then arises what makes the resulting *optimal* version of RE different from, and more plausible than, other non-foundationalist epistemologies of logic. Ideally, answering this question would involve comparing the *optimal* version of RE to each of the other proposals in turn. However, this would be beyond the scope of any one paper. Thus, we address the question more directly by focusing upon one non-foundationalist proposal, *logical predictivism*, which it has been argued elsewhere is preferable to both Quinean holism and abductivism (Martin & Hjortland 2021). According to predictivism, logical theories are justified, and ultimately chosen, on the basis of their predictive success, explanatory power, and compatibility with other well-evidenced commitments. While predictivism differs from abductivism in virtue of explaining the mechanism by which logics generate testable predictions, allowing rival logics to be comparatively tested against novel data, unlike Quine's evidential holism predictivism does not require that logics are assessed as part of a whole web of beliefs. Logical theories can be tested on their own terms, against their own subject-relevant data. Further, while some non-logical theoretical commitments are relevant to the evaluation of logics, predictivism does not commit us to *all* non-logical theoretical commitments being relevant, unlike Quine's holism. While certain theoretical commitments in mathematics and linguistics may be relevant, for instance, our best theories of astrophysics and microbiology need not be.

We go onto show that while the predictivist model fulfils each of the four criteria specified for an *optimal* version of RE, it also brings to the fore three elements of logic's epistemology that RE (at least, traditionally construed) has

[5] Indeed, the only author we're aware of who makes an explicit distinction between RE and another non-foundationalist epistemology of logic is Woods (2019a), who distinguishes RE from logical abductivism. While RE is characterised as a piece-meal approach to logical theorizing, where logical principles are individually assessed relative to some data, abductivism is conceived of as the evaluation of logical theories as a whole (relative to competitors). Interestingly, and plausibly, Woods himself admits that it's likely RE and abductivism would constitute two different stages within the epistemology of logic, RE at the *embryonic stage* of a logic's research program, and abductivism when there are multiple competing logics to choose between. Unfortunately, this meaningful and interesting distinction is somewhat undermined by the fact that recent advocates of RE propose that we can also assess sets of principles on the basis of their holist properties (such as *fruitfulness* and *scope*; see Brun 2020). This seems to further erode the distinction between RE and other non-foundationalist epistemologies of logic. We come back to this point at the end of the paper.

resisted admitting. The paper then ends with a challenge. Either the advocate of RE must: (i) show that predictivism is mistaken in identifying these features of logic's epistemology and provide another *optimal* version of RE to take its place, or (ii) embrace these elements of logic's epistemology, thereby saving RE at the cost of potentially amalgamating the proposal with predictivism and significantly widening the notion of RE beyond its initial articulations.

The rest of the paper runs as follows. Section 2 provides some important clarification on the nature of the projects RE and predictivism are engaged in (as they could easily be misconstrued), and how we should go about evaluating the proposals. Section 3 then moves onto discussing RE, delineating and assessing its available variants, while section 4 presents logical predictivism. Finally, section 5 highlights how predictivism fulfils the criteria we identified for an *optimal* version of RE, but at the cost (for advocates of RE) of contradicting some of RE's traditional assumptions.

2. Important Preliminaries

2.1 The Epistemology of Logic?

Talk of "logic" is itself rather ambiguous, and so talk of the *epistemology* or *methodology of* logic is bound to derivatively contain ambiguities. Generally, the term "logic" can be referring to: (i) the *research area* itself, which is practiced across philosophy, mathematics and computer science departments; (ii) the *objects*, in this case mathematical systems and tools, which logicians produce; or (iii) the *subject matter* of the research area.

Sometimes the *subject matter* of logic (*qua* the research area) just happens to be the formal objects it produces. Logical systems are often developed and studied for their own sake, as objects of mathematical interest. In such cases, researchers may be concerned to develop a system with a particularly interesting or desirable computational property, such as *decidability*, or to compare the available systems on the market in terms of their *deductive strength*, without any interest in choosing one system over another. This research area is sometimes known, in analogy with mathematics, as *pure* logic (Priest 2006a; Martin 2022). While it is an interesting question what the underlying methodological norms at play here are, and how to explain its epistemology, this is not the activity that RE and predictivism are concerned with.

Instead, there is a second type of activity which logicians (and, more generally, those using logics) are engaged in, which is *applying* the formal systems of pure logic for some purpose: whether this be modelling some particular phenomenon, or achieving some specific goal (such as finding the most effective procedure for AIs to process ethical decisions; see Benzmüller et al. 2020). In fact, there are *multiple* phenomena that logics have been used to model: belief revision norms (Grove 1988), grammaticality (Dalrymple 2001), and implication itself (Martin & Hjortland 2021; Woods 2019b). Again, keeping with the mathematical analogy, this broad research area is called *applied logic* (Priest 2006a; Martin 2022). It is in these cases, when we attempt to apply our formal systems to a given phenomenon, and there are multiple such logics available, that we have rivalry between them, and the question of *theory choice* arises.

Strictly speaking, philosophers of logic could be interested in how we come to develop and evidence the best theories for each of these existent applications of logic. For instance, which methodology we use to develop and test logics of belief revision, and which we use when constructing logical models of grammaticality. We could therefore, in principle, assess both RE and predictivism relative to any of these particular applications of logic. In this paper, however, we will be focusing on one particular application of logic: understanding *logical implication* (sometimes also known as *validity*).

This is for two reasons. Firstly, RE was initially developed with the epistemology of logics of deductive validity in mind, and it is this application that more recent RE proposals have focused upon. It would, therefore, be unfair to (at least

initially) assess the proposal on the basis of logical theory choice when it comes to, say, accounts of grammaticality or belief revision norms. After all, there is no assurance that the mechanisms by which theory choice operates across all of these applications are the same. Secondly, because of the (philosophical) importance of this particular application of logics, attention in the epistemology of logic has tended to focus on validity (Cook 2010; Priest 2016; Williamson 2017). It makes sense, therefore, in keeping with the literature, to concentrate on this application here.[6]

2.2 Evaluation Criteria

Now that we are clearer on what both RE and logical predictivism are accounts *of*, it is important to say a few words about how we should go about evaluating the proposals. Firstly, it is quite common now for discussions of theory-choice within logic to be conducted against a background presumption either for or against *methodological anti-exceptionalism about logic*, the proposal that the mechanisms of theory-choice in logic are the same (in important regards) to those used within the recognized sciences (Martin & Hjortland 2022). Consequently, if one were a committed anti-exceptionalist about logic, one may be tempted to either support or criticise the following accounts of theory-choice in logic on the basis that the mechanisms of theory choice proposed are (not) used within the sciences.

We will resist the temptation here to evaluate the accounts on this basis, for two reasons. Firstly, the proposal that the mechanisms of theory-choice between (all of) the sciences and logic is a particularly strong and implausible interpretation of anti-exceptionalism, which it would be unwise to be unduly influenced by (Rossberg & Shapiro 2021). There are more modest and plausible interpretations of anti-exceptionalism which, rather than emphasizing an equivalence between the methodologies of (all of) the sciences and logic, emphasize a rejection of some of logic's traditional properties, such as its analyticity and foundational status in epistemology (Martin & Hjortland 2022). While this latter version of anti-exceptionalism may lend support to non-foundationalist accounts of logical theory-choice, this rejection of the traditional properties of logic is unlikely to be fine-grained enough to determine or favour a particular non-foundationalist account of logic's epistemology.

Secondly, the question of how theory choice in logic functions is orthogonal to the question of whether the methodology of logic is the same as that of the sciences (Martin 2021b; Russell 2019). One could, for instance, consistently embrace RE about logic whilst proposing that the sciences use a predictivist model of theory-choice (or vice versa). Thus, it would be a mistake to entangle the two questions unnecessarily. Further, if we approach the quest for an understanding of logic's methodology through the prism of anti-exceptionalism, we are in danger of getting the cart before the horse. The conclusion that the methodology of logic is akin to that of the sciences should be a *result* of an independent enquiry into the nature of the methodology of logic, not a desideratum of our account of logic's methodology.

---

[6] What do we take *validity* to be, exactly? For the sake of this paper, we can provide a rather deflationary answer. We take validity to be the property of arguments logicians are interested in when assessing whether *some claims follow (necessarily) from others*, including in mathematics. Of course, one can then raise the further metaphysical question of what *grounds* facts about validity—what is validity, *really*?—and there are multiple available answers to this question, including: (i) *conventionalism*, according to which validity is grounded in linguistic facts (Warren 2020), (ii) *psychologism*, according to which validity is grounded in facts about human psychology (Pelletier et al. 2008), and (iii) *universalism*, according to which validity is grounded in the most general structural features of the world (Williamson 2017). However, we don't need to take a stand on the particular metaphysics of logic here. This a paper about the *epistemology* of logic, not its *metaphysics*, and there is good reason to believe that we can understand how logics are evidenced and the basis on which they are chosen without muddying the waters by presupposing a substantive view on the metaphysics of validity (Martin & Hjortland 2022: Sect. 4). We'll have an opportunity to come back to this point below.

This brings us onto the second point regarding our desiderata for an adequate account of logic's epistemology. Whenever we are considering the epistemology of any research field, there is a possible distinction to be made between how the epistemology of that field *does* proceed, and how it *should* proceed. The same is true for theory choice in logic. It is possible that we could supply different answers to the questions of how theory-choice in logic actually *does* proceed, and how it *should* do so.

In what follows we will be primarily interested in the former question, for two reasons. Firstly, both advocates of RE and predictivism are often explicit that what they are offering is primarily an account of how theory-choice in logic *actually operates* (Peregrin & Svoboda 2017; Resnik 1999; Martin & Hjortland 2021). Thus, treating the descriptive project as primary is tantamount to respecting these authors' own projects.

Secondly, it is recognised that while we would ultimately like to be able to answer the second question, the means through which we go about doing so is by initially answering the first. We have no direct access to answers about how the methodology of a field *should* proceed. The only route we have to formulate a reasonable answer to this question is by firstly providing an answer to the descriptive question. After all, in order to advocate for how the methodology of a research field *ought* to proceed, we need a prior understanding of its research goals, given that different research aims require different methods (Woody 2015). Yet, of course, we do not have direct access to a field's research aims. They can only be understood through a constructive process, built upon the activities of the field's practitioners. These considerations show that using the current methodology of the field is the best guide we have to how it should proceed.

With this in mind, this paper uses a broadly *practice-based approach* to the epistemology of logic, whereby we test available epistemologies of logic on the basis of their compatibility with the ways that logics are actually evidenced in the field (Martin 2022). If a particular account cannot make sense of a whole host of the reasons that logicians actually give in support of logics, and the types of theoretical disagreements that occur between logics and logicians, then this must ultimately count against the proposal.

Lastly, understandably much of the attention when discussing non-foundationalist epistemologies of logic has focused on what is diversely known as the *centrality* (Wright 1985), *background logic* (Martin & Hjortland 2021; Woods 2019b), and *metalogic* problem (Shapiro 2000). According to this problem, all non-foundationalist epistemologies of logic are bound to be incomplete given that in order to evaluate a logical theory relative to some data, one must assess the extent to which the data "coheres with", "is consistent with", or "fits" the theory, all of which *presume* the validity of certain rules of inference. Consequently, we are owed a story regarding which logical principles (or, inferences) a logician is allowed to presume in this process, and how this can be achieved without non-foundationalist epistemologies collapsing into a foundationalist epistemology.[7]

Despite the importance and interest of this problem, it will not constitute part of our evaluation of RE here, for in virtue of being non-foundationalist epistemologies of logic, both RE and predictivism suffer from versions of the same problem. Therefore, neither is at a relative advantage over the other in this regard. Inversely, if a solution to the problem is found for one of the proposals, a solution should be quickly forthcoming for the other too. Consequently, given that we are concerned here with evaluating the plausibility of RE *relative to other non-foundationalist epistemologies*, the problem provides us with no reason to prefer one such non-foundationalist epistemology over another.

[7] Unsurprisingly, this problem facing non-foundationalist epistemologies of logic is one of the main motivations for foundationalist epistemologies of logic (Martin & Hjortland forthcoming).

Now that we are clearer on the projects that RE and predictivism are engaged in, and the criteria we shall use to assess RE, it is time to consider the proposals themselves.

3. Reflective Equilibrium about Logic

All non-foundationalist proposals differ from foundational accounts of logic's epistemology by recognising that logicians do not have *unmediated access* to the correct logical laws. Rather, in order to discover the correct laws, inferences must be made from some relevant data in order to inform and subsequently test logical theories. The various proposals, however, are free to disagree over: (i) what counts as this *suitable data* to inform logical theories, (ii) what constitutes these *logical theories*, and (iii) the *mechanisms* by which these data inform the available theories.

RE itself is no different, for it is not *one* proposal. It comes in a host of varieties, some of which are more plausible than others. While, in broad terms, each proposes that logicians come to develop and justify their logics via a process of *mutual adjustment* between some *given data* and the *principles constituting the (logical) theory*, there is disagreement over: (i) what constitutes the data, (ii) what constitutes the theory, and (iii) how the mutual adjustment proceeds.

To see how versions of RE can differ, it will be instructive to start with Goodman's original articulation of the proposal. Goodman (1983: 63-4) tells us that, contrary to what foundationalist accounts of logic's epistemology propose:

> Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields inacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences… The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.

Here we find the root of the idea behind RE. Rather than having direct access to the truth of the components of the correct logic of validity, the theory must be inferred from our *judgements* over the (un)acceptability of *particular inferences*.

Thus, beginning with *instances of* inferences that we find either acceptable (i_a) or unacceptable (i_u), we then infer *rules of* inference r which (combined) constitute our logic of validity $\mathscr{L}$. These rules are then further checked against other instances of inference. If the new instances of inference are in "agreement" with the existent rules constituting $\mathscr{L}$, then all is well. In contrast, if a particular rule r is found to either sanction an *un*acceptable inference i_u, or prohibit an acceptable inference i_a, then some adjustment must be made. Either we must amend r (and thus $\mathscr{L}$), or we must re-evaluate the status of i_a or i_u. Ultimately, the aim is to continue this process of mutual adjustment until a state of *reflective equilibrium* is reached between the logic $\mathscr{L}$ and the set of judgements over particular inferences I, such that no $r \in \mathscr{L}$ is in tension with some $i_a \in I$ or $i_u \in I$.[8]

[8] Note, there is no requirement that the set I is the union of those sets of inferences we find acceptable and unacceptable. It is perfectly possible for I to also include inferences we suspend judgement over. Further, in talking generally about one's "judgements" over particular inferences, we do not require that one must *explicitly avowal* or *reject* the inference. While one's endorsement or rejection of an inference can take this form, such attitudes can also be displayed implicitly by one's willingness to (not) make an inference, or by one's criticism of others for making an inference. In what follows, unless stated otherwise, when we speak of "judgements" regarding some inference, we allow such judgements to refer to both *explicit* avowals or rejections of an inference, and *practical decisions* to accept or reject an inference. Thanks to an anonymous referee for pushing us on this latter point.

Some features of Goodman's articulation of RE are particularly informative. Firstly, Goodman is clear that the process of justifying our logical theory is *exhausted by* the mutual adjustment between the rules constituting the theory and the set of acceptable inferences. We are told that "in the agreement achieved lies the *only justification* needed for either" (Goodman 1983: 64; emphasis added). Goodman's own version of RE is, thus, what we now call an instance of *narrow* RE. According to versions of *narrow* RE, the *only* suitable data we have to evaluate our theory are our set of judgements over particular cases.[9] Thus, in the specific case of assessing logics of validity, according to *narrow* RE our set of judgements over particular inferences *I exhausts* our evidence for $\mathscr{L}$. Narrow RE is to be contrasted with *wide* RE, in which further background commitments (as of yet unspecified) can also provide evidential support for $\mathscr{L}$ or require its revision. Given the differences in the sources of evidence *narrow* and *wide* versions of RE permit, they need to be assessed separately.

Secondly, Goodman is clear that what constitutes the theory component $\mathscr{L}$ within the mutual adjustment are *rules of inference* (1983: 63). However, this need not be the case. When it comes to thinking of a logical theory $\mathscr{L}$, we have (at least) three options. We can think of $\mathscr{L}$ as comprising:

(i) A set of valid schematized argument forms *F* (plus, in all likelihood, some translation manual between members of *F* and natural-language inferences);

(ii) A set of schematized rules of inference *R* (plus, again, some suitable translation manual between members of *R* and natural-language inferences), which when combined produce a set of valid argument forms; or,

(iii) A more encompassing framework, including sets of propositions expressing laws, definitions, and rules providing a syntax and semantics, which produce the schematized argument forms *F*.

As we shall see, how advocates of RE conceive of the *theory* component also impacts the plausibility of their proposal, and thus proposals which differ in this aspect will need to be assessed separately.

Thirdly, Goodman does not specify who the *subject* of this reflective equilibrium is, whether it is an *individual* or a *community*. If it is the former, this opens up the possibility of there being multiple states of acceptable reflective equilibria, each relativised to an individual. The occurrence of the various acceptable individual-relative reflective equilibria could then be explained as the product of: (i) the differences among the individuals' initial judgements regarding the (un)acceptability of specific inferences, and (ii) different *threshold levels* amongst individuals when it came to revising one's theory in the face of recalcitrant data (and vice versa).

In comparison, in the community case, the process of arriving at a state of reflective equilibrium is a *communal and collaborative enterprise*, where ultimately agreement is sought by the relevant participants on: (i) what constitutes the relevant data; (ii) the (in)compatibility of this data with a particular theory; and (iii) how to revise the theory in light of this data (and vice versa). On such a communal account, while at one time there may be several candidate theories, the expectation is that with an increasingly set of suitable data, and assessment of the various candidate theories relative to this data, there will be a convergence towards an agreed privileged reflective equilibrium. In other words, the community will converge on a (generally) agreed theory.

[9] Some in the RE literature think these "particular cases" can also be constituted of general claims (Brun 2014a). We'll come onto this point in Section 3.1.

Finally, Goodman is unclear about the *content* of the judgements regarding particular inferences that inform our logical theories. When first outlining the process of RE, Goodman talks in terms of finding the inferences *(un)acceptable*, which then informs our theory about which arguments are *(in)valid*. In this case, the content of the data informing our theory is different from that of the theory itself. Thus, the content of the judgements constituting the data cannot directly contradict the content of the theory (even putting aside concerns raised by the background logic problem). Rather, the content (or, consequences) of the theory must be *operationalised* so that it can be suitably tested against the data.

This process of operationalisation, and the distinction between phenomenon and data, is well known from the philosophy of science (Bogen & Woodward 1988). Rarely do we see the phenomenon that our target theory is about. Instead, appealing to certain background assumptions, we must operationalise the consequences of a theory in order to test it against observational or measurable data. For instance, we do not directly observe alpha and beta particles in a cloud chamber. Rather, we observe condensation trails, which are deemed a *reliable indicator* of the occurrence (and direction of travel) of the charged particles. In such cases, maintaining the phenomenon/data distinction is paramount to understanding how the relevant theories are tested.[10]

At other times, however, Goodman (1983: 66-7) implies that the content of the theory and the supporting data are the same, thereby breaking down any phenomena/data distinction:

> The task of formulating rules that define the difference between valid and invalid inductive inferences is much like the task of defining any term with an established usage. If we set out to define the term 'tree', we try to compose out of already understood words an expression that will apply to the familiar objects that standard usage calls trees, and that will not apply to objects that standard usage refuses to call trees… [T]he interplay we observed between rules of induction and particular inductive inferences is simply an instance of this characteristic dual adjustment between definition and usage, whereby the usage informs the definition, which in turn guides extension of the usage.[11]

If the justification of logical theories using RE is similar to providing a definition of "tree", using standard applications of the term as a guide, as the quote suggests, then this would this be tantamount to proposing that we use judgements regarding the *deductive (in)validity* of particular inferences as data for our theories of validity. Yet, proposing that the content of these judgements itself is *about* the inferences' *logical (in)validity* is very different from proposing that we use judgements regarding an inference's *(un)acceptability* as reliable data to evidence one's theory of validity. Just as it is one matter to propose that observing post on the doormat is a *reliable indicator* that the postie has been, and another to suggest one has *actually observed* the postie. Again, as we shall see, how we understand the nature of the data informing logical theories will impact the resulting version of RE.

It is important not to amalgamate this epistemological question over the identity of the relevant logical data—whether they are comprised of judgements over the *(in)validity* of specific inferences, or rather judgements over the inferences' *(un)acceptability*—and the metaphysical question of whether facts about validity *simply are* facts about what we find to be acceptable or not. While the former is a question regarding the content of the data, the latter is a question about what metaphysically grounds the truth of logical claims about validity.

---

[10] Of course, if there is no phenomenon beyond the data—that is, the phenomenon is merely *constructed out of the data*—then no such distinction is required. We come back to this point shortly.

[11] Goodman is talking here about induction, not deduction, but presumably he would say the same about the case of RE with regards to theories of deduction.

Goodman at times uses language which fails to sufficiently differentiate these two questions. For instance, we are told that: "Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity *depends upon* accordance with the particular deductive inferences we actually make and sanction," (1955: 63; emphasis added). Yet, it's unclear whether Goodman merely means that the *evidence* for our logical theories solely depends upon a theory's accordance with the sanctioned inferences, or rather that our sanctioning these specific inferences itself somehow metaphysically grounds facts about validity. While the former interpretation merely commits one to the epistemological thesis that the only pertinent evidence for logical theories regarding validity are judgements over specific inferences, the latter interpretation commits one to the metaphysical thesis that there is nothing to validity beyond these judgements over specific inferences. In other words, facts about validity are *constructed out* of these judgements.[12]

It is currently an open question whether validity is constructed out of these judgements, or rather that such judgements merely serve as a *reliable guide to* facts about validity. The latter view, however, is by far the more prominent position in the field. Metaphysical theories of logic as varied as conventionalism, psychologism, and structuralism, all admit that there is a distinction to be had between facts about validity (the phenomenon) and the evidence we use to discover these facts (the data). For instance, for the conventionalist, the laws of logical consequence supervene upon the linguistic rules of the community, and thus in virtue of being competent language users, native speakers are reliable (though fallible) judges of when a given claim *linguistically follows* from another (Warren 2020). This linguistic data is then thought to serve as a reliable guide to the correct logical laws. Similarly, according to the structuralist, the laws of validity supervene upon the most structural features of the world (Maddy 2007; Williamson 2017). Accordingly, as (certain) individuals are sensitive to these structural features, their judgements over relevant inferences are a reliable guide to the laws. In each of these cases, there is a fact of the matter as to whether an argument is valid or not, independent of the judgements regarding the acceptability of particular inferences, although the latter are treated as a reliable guide to the correct laws of validity. Despite this, there are some existent positions in the metaphysics of logic which do identify the putative phenomenon with the data—validity *just is* our attitudes towards specific inferences. Logical expressivists (Resnik 1985), for instance, propose that the logical laws are nothing but the codification of an individual's own attitude towards which inferences can or should be made. This is, of course, an extreme version of logical anti-realism.

In general, little of what we have to say in what follows should depend upon a particular metaphysical interpretation of logic or have an impact upon the reasonableness of one's particular metaphysics views on logic. This is because it has been shown elsewhere that non-foundationalist epistemologies of logic are consistent with a wide range of metaphysical pictures of logic, including conventionalism, psychologism, and structuralism (Martin & Hjortland 2022: Sect. 4). All of these parties agree that inferential data reliably informs our theory of validity. They simply disagree in their interpretation of *why* these judgements over inferences are reliable evidence for our best theories of validity. RE as an epistemology of logic, therefore, does not privilege a particular metaphysics of logic. In general, then, we can stay relatively non-committal regarding the precise metaphysics of logic when determining its epistemological standards.[13]

---

[12] An anonymous referee suggests to us that they have always interpreted Goodman as endorsing this latter metaphysical view. He may well have, although we think there is insufficient textual evidence to determine the matter. Anyhow, given that what we will go onto argue is independent of a particular metaphysical interpretation of Goodman's views, we need not take a stand on the matter. The important point for us here is that the epistemological and metaphysical questions are not amalgamated or confused for one another.

[13] The possible exception here are those metaphysical accounts of logic, such as expressivism (Resnik 1985), which identify the phenomenon of validity with the inferential data. In such cases, inferential judgements are understood not as defeasible

As this brief reflection on Goodman's initial expression of RE shows, there is significant scope for divergence over how to understand RE when it comes to logic. In particular, there is possible disagreement when it comes to understanding: (i) the *content of* the inferential data informing the theory; (ii) the *constituent parts* of logical theories themselves; (iii) the *scope* of the evidence for logical theories; and (iv) whether the process of mutual adjustment is *individualistic* or *communal*. In the rest of this section, we take each of these possible disagreements in turn, with the aim of establishing those options most suitable for RE about logic, and thus what would constitute an optimal version of RE about logic. We begin with the question of how we should understand the inferential data informing logical theories.

3.1 What is the Data?

When it comes to understanding the content of the inferential data informing our logic, there are three broad options:

> *Naïve* RE: The content of the judgements concern the *logical (in)validity* of the inferences. Thus, the contents of the judgements are ultimately the same as the subject matter of the target theory (namely, validity), thereby collapsing the phenomenon/data distinction.
>
> *Operationalized* RE: The judgements concern not directly the *(in)validity* of the inferences, but another property which is treated as a reliable guide to logical (in)validity, such as the (un)acceptability of the inferences.
>
> *Generalized* RE: As well as judgements concerning the *logical (in)validity* of specific inferences, judgements about *general rules of inference* or *logical principles* (however conceived) should also be included as suitable data.[14]

Let us begin with the comparative plausibility of *naïve* and *operationalised* RE. Both have (implicit) advocates in the literature. For instance, Brun (2014a: 111) tells us that it is a *requirement* of RE within logical theorizing that both the "principles" constituting the theory and the "judgements" constituting the data are *about validity*. Peregrin & Svoboda (2017), in contrast, speak more often of judgments regarding the "correctness" of an inference.

There are several reasons for thinking that operationalised versions of RE are superior to naïve versions. Firstly, the concept of *validity* as used within logics is a wholly technical term. Thus, while it is plausible that prior to logical theorizing individuals have a working concept of a *good*, *correct*, or *acceptable* inference, it is dubious at best that those individuals whose judgements we are treating as a *reliable source of data* for our logical theory possess an understanding of "logical validity" (Glanzberg 2015; Martin & Hjortland 2021). In this sense, the term "validity" is no different to other technical terms in the sciences, such as "force" and "genes", which are introduced to refer to theoretically posited phenomena for explanatory purposes (Hjortland 2019).[15]

---

but reliable *evidence for* an argument's (in)validity, but rather as constituting the phenomenon itself. In this unusual case there may be epistemologically motivated evidence against the metaphysical proposal. We'll come onto these considerations in sections 3.4 and 5. Thanks to an anonymous referee for pushing us on these questions regarding the metaphysics of logic.

[14] *Generalized* RE, therefore, should be considered a supplemented version of *naïve* RE, given that all advocates of generalized RE also admit that judgements over the validity of specific cases can be used as suitable data for informing a logical theory.

[15] Of course, it is an interesting question *why* logicians should treat the attitudes of certain reasoners towards inferences as reliable evidence for an inference's (in)validity at all (dos Santos 2021; Martin & Hjortland 2021). As we noted above, the answer one gives to this question will in all likelihood depend on one's metaphysical commitments regarding logic (Martin & Hjortland 2022: Sect. 4). However, in all such cases, there is no requirement that the reasoners possess the concept of validity in order for these judgements about specific inferences to be treated as a reliable guide towards validity.

Further, advocates of RE often rightly admit that our evidence that a particular inference is endorsed, and thus should be sanctioned by our logic, takes the form *not* of an explicit avowal of the inference but is rather expressed through behaviour (Brun 2014a: 111). For instance, that individuals commonly make an inference of this type, or criticise others for not adhering to the requirements of the inference. Yet, it is quite a leap to infer from this behaviour that the individuals involved believe that the inferences are *logically (in)valid*, which again is a technical term and theoretically loaded, rather than simply *(un)acceptable*. Analogously, descriptive linguistics does not treat native-speakers' speech acts as implicit claims (or beliefs) about *grammaticality*. Rather, in virtue of being competent speakers, these actions are treated as a reliable guide to the grammatical rules of a language.

Finally, operationalized versions of RE provide a better explanation of why theoretical work on the part of the logician is required to distinguish not only deductively from *non*-deductively good inferences, but further *logically* valid inferences from those implications which are purely mathematical or lexical in nature. Even if we admit that there is some qualitative or cognitive difference between the attitudes of reasoners towards deductive and inductive inferences (Goel et al. 1997), there is no established detectable difference between those attitudes reasoners have towards inferences that logicians deem logical, such as instances of *modus ponens* or the *disjunctive syllogism*, and those inferences which depend upon the content of mathematical concepts or lexical entailments. Whether this be an inference in graph theory whose validity relies upon the content of defined mathematical notions such as an (un)directed graph, vertices and edges, or the lexical inference from "We loaded the truck with hay" to "We loaded hay on the truck" (Anderson 1971), neither of which are usually thought to be *logical*.

The question of where to draw this line between logical and non-logical implication is an important (and open) topic for logicians (Sher 1991). Yet, if we were to suppose that the contents of reasoners' judgements are the same as the subject matter of the theory which the data informs, much of this challenging theorizing about the distinction would be unnecessary. Logicians could make the distinction purely on the basis of those inferences reasoners judged to be *logically* (in)valid, purely mathematically acceptable, or rather a lexical entailment. All of these considerations lend support to an operationalised version of RE over naïve versions.

Some in the literature also suggest that, in addition to using judgements regarding specific inferences as data, we may appeal directly to judgements regarding *logical rules* or *principles* to inform a logic (Brun 2014a: 111; Woods 2019a: 324). For instance, that some "initial judgement" over the truth of the principle that *conjunctions imply their conjuncts*, or that the quantifier-negation commutation rules are valid, can serve as suitable data to inform our theory. Yet, admitting such generalized judgements into the set of suitable data would be a mistake for advocates of RE.

Firstly, it brings RE dangerously close to becoming a foundationalist epistemology of logic. After all, what differentiates foundationalist from non-foundationalist epistemologies of logic is that the former suggest we have direct access to the validity of certain logical rules, or the truth of certain logical principles, while the latter do not.[16] Nor can the advocate of RE hope to differentiate their position from foundationalist positions by admitting that these judgements regarding logical rules or principles are a *fallible* guide towards the correct logic, for many foundationalist epistemologies of logic also grant that logical judgements are fallible (BonJour 1998).

[16] As mentioned above, some advocates of RE call the position "weakly foundationalist", meaning that the initial commitments in the form of judgements over specific inferences have some credibility independently of their cohering with a logical theory. However, this is a very different matter to claiming to possess reliable beliefs over the truth of *general claims concerning logical principles* prior to theorizing.

Secondly, in comparison to the regular appeals to the (un)acceptability of particular inferences which can be found in the logical literature, it is rare for logicians to directly appeal to the validity of a *rule* when engaged in theory choice. Take, for example, Williamson's (1994) examination of multi-valued and supervaluationist solutions to the sorites. Rather than directly appealing to the correctness of the relevant classically valid rules of inference which the non-classical logics fail to sanction in order to undermine the latter logics, Williamson produces examples of *specific inferences* that we deem acceptable but which the target logic deems invalid. For instance, Williamson (1994: 106) criticises Halldén's (1949) three-valued gappy logic for entailing that the inference from "Jack is not a philosopher" to "Jack is not a bald philosopher" is invalid when Jack is a borderline case of bald, although it is a perfectly acceptable inference. Indeed, in those rare cases that appeals *are made* to general principles, such as when Slater (1995) infamously attempted to disprove dialetheism by appealing to the principle that, by definition, contradictories could never be jointly true, the appeal gained no purchase in the literature. Debate on the virtues (and vices) of glutty paraconsistent logics carried on regardless.[17]

Actually, there is good reason to think that when appeals to rules of inference or logical principles *are made* within logical debates, they are not made within the context of providing new evidence for a logical theory. Rather, they are *post hoc* appeals, summarising aspects of our presently accepted logical theory which have already been independently supported. In other words, such appeals serve as reminders not to blindly revise that particular aspect of the (logical) theory given that it has significant independent support (on the basis of, say, conforming to specific inferential judgements). The appeals, then, do not themselves have any evidential weight. An example of this phenomenon comes again from Williamson's (1994: §4.2) discussion of vagueness, where he criticises some non-classical semantics for failing to respect truth-functionality, which would lead to many of the advantages gained by classical logic to be lost.[18]

Thus, there are good reasons to think that the optimal version of RE, at least when it comes to logical theorizing, should embrace an *operationalized* account of data.

3.2 What Constitutes a Logical Theory?

Now that we have determined the nature of the *data* that the optimal version of RE should use, we have the further question of how RE should conceive of the *theories* informed by this data. There are three broad options here as well:

Extensional Conception: A logical theory is simply a set of individualised valid argument forms, theorems (assuming a *deduction theorem*), or both, in the object-language of the logic. This is a long-standing conception of what constitutes a logic and was particularly popular before the proliferation of non-classical logics (Barrio et al. 2020; Gabbay 1994).

Privileged Set Conception: A logical theory is constituted of a set of *privileged inference* or *semantic rules* for the logical connectives in the suitable object language which, when combined, deliver the set of valid argument forms (and theorems), as understood in the *extensional* account. One might think this conception is implicit within certain versions of inferentialism, according to which the meaning of a logical connective (in a logic) is fixed by a certain privileged subset of rules (Peacocke 1976). Combined, these rules then deliver the consequence relation of the logic.

[17] As part of their motivation for including such generalized judgements as data, Brun (2014a) appeals to Rawls' own conception of the data informing theories of justice, which *did* include judgements about general principles. However, of course, the fact that these generalized judgements are used as evidence within moral and political philosophy does not mean they are in logic. This is a good example of the dangers of extrapolating from the epistemology of one field to another.

[18] For more on the role of such *post hoc* appeals in logical theory-choice, see Martin & Hjortland (2021).

Richer Conception: Similarly to scientific theories, logical theories are a set of definitions, (meta-)rules, and laws which (combined) produce something like a set of valid argument forms. This conception has been more recently advocated in the anti-exceptionalist literature, in Hjortland (2019), Martin & Hjortland (2021) and Blake-Turner & Russell (2021).

While Goodman himself speaks in terms of logical "principles", he does not explicitly advocate any of the options above. Further, unfortunately, the options have not been sufficiently distinguished in the RE literature. Brun (2014a: 111), for instance, tells us that "[a]s examples of systems of principles [e.g. logical theories], we may think of axioms with a deduction-rule, rules of natural deduction or rules defining validity in terms of semantic tableaux." However, it's unclear whether we should understand these rules as *solely determining* the logic, consistent with the *privileged set* conception, or as part of a *richer* logical theory. Yet, as we shall now see, distinguishing these proposals matters, for not all of them are able to adequately individuate logics and explain how logics would be suitably tested consistent with the RE proposal.

The first point to note is that unsupplemented, both the *extensional* and *privileged set* conceptions are inadequate. If these argument forms and rules are expressed within the object-language of the logic (which they must be), and this object-language is different from the language in which the data is expressed (which it is, being usually natural-language inferences), then in order for these components of the theory to be suitably tested there must be a *translation* (or, *application*) *manual* between the object-language of the theory and the data. The need for the inclusion of such a manual is well recognised from the literature on model-building in the philosophy of science (Weisberg 2013) and is now suitably appreciated by advocates of RE (Brun 2014a; Peregrin & Svoboda 2017).

Given that this translation manual would take the form of supplementary rules within the theory, suitably applying the theory to testable data (and vice versa), we can already see that this recognition for such rules pushes the advocate of RE towards a richer conception of logical theory. After all, there is no requirement for two theories to agree on how this translation should proceed. Thus, if we wish to apply a formal logic to a particular phenomenon (e.g. deductive validity), our theory cannot be constituted solely of valid argument forms (or, a set of privileged rules) in the object language.

However, even if we supplement the *extensional* and *privileged set* conceptions with a suitable translation manual, there are still good reasons to doubt the suitability of either. As far as the extensional conception is concerned, in virtue of a logic being nothing more than a set of valid argument forms (and/or theorems), each of these argument forms must be *individually* evidenced by the available data. After all, so conceived, for any possible non-empty logic $\mathscr{L}_1$ constituted of the set of valid argument forms $\Gamma \cup \varphi$, there is another possible logic $\mathscr{L}_2$ constituted solely of Γ. Thus, in all cases, the argument form φ must itself be directly evidenced by the available data so as to differentiate the evidence for $\mathscr{L}_1$ and $\mathscr{L}_2$.

Yet, this is implausible. Not only are there too many valid argument forms for us to "check" individually against actual inferences, but there are multiple argument forms deemed valid (for instance, by classical propositional logic) that either we do not have judgements over instances of, or it is difficult to even find instances of, due to their complexity. Such tricky complex examples include theorems containing embedded conditionals, such as Peirce's law and the conditional distribution laws (Martin & Hjortland 2021).

Given this, the *extensional* account has no plausible story for how these particular valid argument forms are evidenced, and thus why they are included within the accepted logic. This shows that a plausible account of logical theories must recognise that even if the validity of some argument forms can be directly evidenced by the data (in the form of acceptability judgements over instances), others must be *derivatively evidenced* somehow, perhaps in virtue of some privileged rules being directly evidenced, and these former rules then (combined) necessitating a larger set of validities. This, of course, gives some initial plausibility to the *privileged set* conception.

This brings us onto a second concern with the extensional conception, which is that the conception is unable to explain the occurrence of *spandrels* in logical theorizing. Within the context of logic, a spandrel is an unforeseen consequence of a logic, either in the shape of a surprisingly valid argument form, or an initially plausible argument form which fails to be valid. Such spandrels occur with some regularity in logical theorizing. The most famous probably being explosion and other "paradoxes of implication" in the case of classical logic.[19] Further, the occurrence of these spandrels is sometimes enough to motivate competing theories, such as with relevant logics in the case of the paradoxes of implication. However, if logics were treated merely as sets of argument forms which were individually evidenced, as conceived by the *extensional* account, then spandrels should never arise, nor should they motivate a revision of the prior evidenced argument forms. After all, given that each argument form would be *independently evidenced*, there would never be any unforeseen consequences for these argument forms arising from the other commitments of the logic.

This shows that a plausible conception of logical theories must not only recognise that certain argument forms can be derivatively evidenced somehow, but that explicit (putatively evidenced) commitments on the part of a logical theory can entail further unforeseen commitments through which the theory can be subsequently evaluated.

Clearly, the *privileged set* conception of logical theories improves on both of these counts. Firstly, according to the conception we only need initial direct evidence for a certain privileged set of rules of inference, from which all of the valid argument forms of the logic can then be derived. There is no requirement, therefore, for each of the argument forms a logic validates to be directly evidenced through data. Secondly, it shouldn't be surprisingly that when combined, these directly evidenced rules of inference produce certain unexpected consequences, whether this be an unforeseen valid form or the omittance of a desired argument form. Indeed, this is just how Lewis & Langford (1932: 253) went about showing that explosion was valid in classical logic, through appealing to the rules of addition and disjunctive syllogism.

The *privileged rules* conception, therefore, has the further desirable property that it is able to explain how logicians come to recognise the underlying causes of the (putatively) problematic features of a logic, and thus how to go about suitably revising the theory to avoid these outcomes. Notably, in the case of explosion, one has the option of either invalidating addition, the disjunctive syllogism, or both (Priest 2006b).

As it stands, however, the *privileged set* conception has its own faults, even when supplemented with a suitable translation manual. Firstly, as is well known, there are notable cases of distinct logics which share the same set of rules of inference. For instance, classical logic and the gappy three-valued K3. In this case, mention would also need to be made of the logics' distinct set of theorems, given that the latter K3 has none. Consequently, we cannot simply equate logics with sets of rules of inference. The same problem also holds for other pairs of logics, such as classical and supervaluationist logic, which differ only at the level of meta-inferences. Even including the relevant meta-inferences within one's definition of a logic would not suffice, however. For it has been shown that in order to distinguish classical logic from the substructural family of logics arising from ST, no set of (meta-)inferences within a finite hierarchy of consequence relations will suffice. Instead, one can only distinguish classical logic from each member of this family by appealing to an *infinite* sequence of consequence relations (Barrio et al. 2020).

---

[19] Though there are other examples, such as the case of the *mingle axiom* in relevant logic:

(**M**) $\varphi \rightarrow (\varphi \rightarrow \varphi)$

While thought to be plausible in itself, adding **(M)** to the relevant logic *R* led to the resulting logic *RM* no longer having the desirable (for relevant logicians) variable sharing property (Meyer 1971).

Secondly, even putting the concerns above aside, in order to determine the validities deriving from a set of rules of inference, one needs structural rules determining how such inference rules may be strung together, as well as a definition of *validity* (i.e. logical consequence). Nor should the suitable definition of validity for a logic be thought so obvious that it can be treated as implicit within the logic. After all, there are some logics, such as supervaluationist logics, that include multiple consequence relations in order to make finer-grained distinctions between arguments (Varzi 2007). Further, it is sometimes far from easy to provide a suitable definition of logical consequence for a logic, as has been shown in the debate over how to provide a workable *proof-theoretic* definition of validity (Schroeder-Heister 2006).

This further goes to show that our identity criteria for a logical theory must include not only the semantics for the logic's connectives (whether understood proof-theoretically, model-theoretically, or somehow else), but details on the logic's (potentially infinite hierarchy of) consequence relations, structural rules, and translation manual. Combined, these considerations highlight the superiority of a *richer* conception of logical theories, including all of these elements. Unfortunately, despite the fact that the weaknesses of the *extensional* and *privileged set* conceptions of logic may be (implicitly) recognised in the literature, the details of what should take their place are less forthcoming. As far as we are aware, no advocate of RE has yet supplied a suitable account of what needs to be included within a logical theory so that it can be suitably evidenced, tested by, and revised in light of the available data.[20]

To be workable, an account of what comprises a logical theory needs to not only inform us of: (i) the components that constitute a logical theory, so as to explain the live disagreements among philosophers over the correct logic, but further (ii) how the components of a theory *combine* to generate testable claims, which we need to make sense of the logic "cohering" or "fitting" the data, and (iii) how these components are re-evaluated in light of the data, given the fact that the components cannot be tested individually.

Somewhat ironically, the *extensional* and *privileged set* conceptions have less problems answering these latter two questions, which formed part of their intuitive appeal. It is clear how each argument form, or rule of inference, would be assessed. We simply find the natural-language surrogates of each argument form/rule of inference and see whether there are robust judgments over their (un)acceptability. Further, if there are robust judgements over their unacceptability (without competing judgements to the contrary), they can be simply dropped from the logic to avoid these unsavoury consequences. Thus, the added complexity of the *richer* conception is both a theoretical strength and potential weakness.

This is not to say that such a suitable account cannot be supplied. As we shall see below, within the context of another anti-foundationalist epistemology of logic, *logical predictivism*, such an attempt has been made. However, in order for RE to provide us with a plausible account of logic's epistemology, we are owed a suitable understanding of what constitutes a logical *theory*.

[20] There may be a good reason for this. As an anonymous referee suggests, it's possible some advocates of RE understand the proposal as not supplying us with a fully-fledged account of how logical *theories* are evidenced, but rather simply how certain initial logical *principles* are justified, which serve as the starting point for the development of a fully-fledged logical theory. So understood, RE about logic doesn't attempt to provide us with a comprehensive account of logic's epistemology but rather its embryonic stages. This is a reasonable proposal, though we think ultimately mistaken. We come back to it in Section 5, when comparing RE with logical predictivism.

3.3 Narrow or Wide Reflective Equilibrium?

So far then we have shown that the optimal version of RE must both use an *operationalised* account of data to inform the theory and have a *richer* notion of logical theories than either the extensional or privileged set conceptions, in order to make appropriate sense of logical disagreements.

This now brings us onto the question of the *scope* of the evidence for logical theories. Will it suffice to allow for the theory to be tested solely against judgements regarding the (un)acceptability of particular inferences, as admitted by *narrow* RE? Or, are logics tested against a wider range of sources of evidence, including other independently well-evidenced background commitments, as admitted by *wide* RE?

Here, at least, there is general agreement that a narrow conception of sources of logical evidence won't suffice (Brun 2014a; Woods 2019a). After all, important research programmes within philosophical logic have been motivated by a whole host of further commitments, including: (i) our best theory of truth; (ii) our best mathematical theories; and (iii) general linguistic facts, such as the (putative) fact that our natural languages contain vague predicates.

For example, the two most prevalent motivations behind dialetheism, the thesis that some contradictions are true, which requires embracing a glutty paraconsistent logic, are the liar and Russell paradoxes (Priest 2006b). However, underlying the rationale for a dialetheic solution to these paradoxes, rather than a solution compatible with classical logic, are commitments to (putatively) independently well-evidenced background theories (Martin 2021b). In the case of the liar paradox, these include commitments to an unrestricted truth-predicate and the semantic closure of our formal language, so as not to impose *ad hoc* expressive limitations on our natural languages. In contrast, with the Russell paradox, the rationale for a dialetheic solution is built upon the commitment that the naïve (inconsistent) concept of a set is needed to carry out important work within mathematics, including performing certain operations in category theory and to build the notion of a *hierarchy* (of sets) within set theory itself.[21]

The important question for the advocate of RE is not, therefore, whether such (putatively well-evidenced) background commitments should be included as sources of logical evidence. They must be, to have anything approaching a comprehensive account of logic's epistemology. Rather, the important questions are rather: (i) *Which* background commitments constitute suitable evidence for logics (assuming not all do)? and (ii) How these background commitments *interact with* logical theories to lead to revisions of the latter, assuming that a *richer* notion of logical theories is required?

While an exhaustive answer to the first of these questions is in all likelihood too much to ask for any epistemology of a research field, as we can never foresee what will turn out to be relevant evidence to a research area in the future, we should at least expect the most prominent background commitments to be identified and for a general sketch of how these commitments interact with logical theories to be provided. However, while there has been explicit recognition in the RE literature of some of the types of background commitments that play an important role within logical theory choice (Brun 2014a; Woods 2019a), there has been no discussion we are aware of on how these background commitments interact with logical theories so as to bring about a revision of components of the latter.

[21] A similar story could be told for the motivations behind gappy (Field 2008) and substructural logics (Zardini 2011), based upon the liar paradox, and supervaluationist logics (Keefe 2008), motivated by the (putative) existence of vague predicates in our natural languages.

3.4 Individual or Communal RE?

This brings us onto the final important distinction between varieties of RE, which is significantly underappreciated in the literature. Given the coherentist connotations of RE, one might wonder whether each of us is free to arrive at *our own* state of reflective equilibrium, having suitably balanced the competing pushes and pulls of the available logical data and our working logical theories. On this conception, RE within logic is an *individualistic enterprise* in which we each have different starting points in terms of the inferences we find (un)acceptable, and how we are willing to adapt our logical theories in light of these judgements.

Some seem to embrace this individualistic interpretation of RE when it comes to logical theorizing, drawing an analogy between logic and ethics:

> [U]nlike the scientific case, we cannot contrive for logicians to concur concerning the 'data' unless we do some fancy brainwashing. For it is not just a matter of seeing that they have similar experiences; rather it is a matter of making them come to the same evaluations… I have been discussing reflective equilibrium in the hopes of showing that the epistemology of logical necessity and possibility—in so far as there is one—is intuition based, and, unlike an observation based epistemology, nothing outside us promotes the convergence of conflicting intuitions concerning logical necessity and possibility. In this *logical intuitions are like moral ones*, and disagreements about matters of logic can be as intractable as those over morals. (Resnik 1999: 189).[22]

In contrast, a communal interpretation of RE proposes that the process of arriving at a state of reflective equilibrium is a *communal and collaborative enterprise*, where ultimately agreement is sought by the participants on: (i) what constitutes the relevant data; (ii) the (in)compatibility of this data with a particular theory; and (iii) how to revise the theory in light of this data (and *vice versa*).

In comparison to the individualistic interpretation of RE, where we expect different individuals to come to rest with differing states of reflective equilibria without any *apparent tension* with other individuals' own states of equilibria, the communal interpretation expects the relevant participants to seek consensus. Even if there may be several candidate (logical) theories on the market at any one time, and different members of the community may advocate for different candidates, there is an agreed understanding that these theories are competing with one another, and that participants in the debate are not free to go their own way without defending their position against the challenges of their peers.[23]

While one can find explicit advocates of the *individualistic* interpretation of RE (such as Resnik), identifying explicit advocates of the *communal* interpretation is less easy. Of course, this does not mean there are no advocates for the interpretation. Advocates may just think the position too obvious to argue for. Regardless of this, however, it is important that we explicitly recognise the strengths of the communal interpretation when it comes to RE in logic. It will suffice to focus on two features of logic's methodology here which speak strongly in favour of the interpretation.

Firstly, there are pieces of logical data that all concerned within the enterprise of logical theorizing agree need to be suitably accommodated by a successful logical theory. This is not the same as saying that this data *can be currently*

---

[22] Note, Resnik's description of RE within logic also seems to commit him to a *narrow* form of RE which, as we've pointed out above, is insufficient.

[23] One of the few detailed discussions of the distinction between individualistic and communal versions of RE, within ethics, mentions two questions as relevant to the distinction: (i) Whose judgments and principles count as legitimate inputs into RE? and, (ii) Who is in charge of the process of revising and reconciling these elements? (Baderin 2017: 1). Thus, a further distinction could be made between mixed accounts, which allow for individualised/communal *data* but subsequent individualised/communal *theory revisions*. We won't labour on this distinction here however, given that the case for the communal interpretation of RE in both instances when it comes to logic is clear cut.

accommodated. Rather, the requirement to accommodate the data is deemed to be a desideratum of theories in the field, and until such a theory can successfully do so the research question must remain live. For instance, it is a shared desideratum of logics of validity that they facilitate mathematical research, by sanctioning the most general inferences mathematicians use within their informal proofs. This is most clearly shown in the case of non-classical logics which invalidate classical rules of inference which (putatively) play an important role within mathematicians' proofs. Not only do these non-classical logicians recognise the importance of the challenge of sanctioning these mathematical inferences, but they subsequently attempt to adapt their theories in order to re-accommodate this important data.

For instance, both glutty paraconsistent and relevant logics have been criticised for invalidating the disjunctive syllogism (amongst other rules), which is needed to sanction certain informal proofs (Burgess 1983; Tennant 2004). Recognizing the force of these challenges, advocates of glutty and relevant logics have attempted to answer these concerns by either recapturing the classically valid inference or providing suitable non-classical analogues of the classically valid rules, which are equally able to serve the required purpose. For example, relevant logicians have suggested that whilst disjunctive syllogism is indeed invalid when using an extensional disjunction, a relevant analogue of the rule using the intensional disjunction "fission" *is valid*, and it is this latter rule which explains the validity of the pertinent inferences made by mathematicians (Anderson & Belnap 1975). Glutty logicians, by contrast, have attempted to recapture the validity of these important rules of inference, including disjunctive syllogism and modus ponens, with a *consistency operator* which specifies that these classically valid rules of inference are indeed valid when we are assured the relevant propositions behave consistently, as we are in mathematics (Beall 2013). In both cases then, members of the logical community agree over the data that a successful theory must accommodate.[24]

The second feature of logic's methodology that lends support to the communal interpretation comes naturally from the first. This is the fact that advocates of the various research programmes in logic bother to challenge the claims of one another, and request reasons for endorsing the competing position rather than their own. In this respect, logical theorizing is a *competitive* (and collaborative) enterprise, as in the sciences.

If logical theorizing were an individualistic matter, where what was acceptable for one party to conclude regarding the correct logic was independent of what was acceptable for another party, we should not expect so much time and energy to be spent publishing papers on the relative strengths and weaknesses of the competing research programmes. Of course, this would not stop logicians in their own time, on the basis of their own personal preferences, assessing the available evidence and coming to a conclusion over the "best" logical theory. However, the whole process of peer-review presupposes that we have (generally) shared epistemological and methodological values, and that we are engaged in a collaborative enterprise. One would need a significant error theory to explain why: (i) logicians bother engaging in identifying flaws in others' reflective equilibria, if each had the right to their own equilibrium and there was no inherent tension in each participant having distinct equilibria; and (ii) work on logical theory choice goes through peer review, if there is no assumed background of shared epistemological and methodology norms that determined how we should assess and revise our theory in light of the available data. Yet, this is what the individualistic interpretation of RE suggests.

Indeed, what is surprising about logical disagreements is how robust the agreement over the relevant sources of logical evidence is, as well as how logics should be assessed in virtue of the available data, and the comparative strengths

[24] Note, this "shared data" need not only take the form of inferences that must be accommodated. It can also take the form of established test puzzles, such as the logico-semantic paradoxes, that must be appropriately solved.

and weaknesses of the competing logics (Martin 2021c). After all, this is why logicians bother engaging in debates over the correct logic and attempt to fix their own theories when weaknesses are identified by peers. For this reason, the communal interpretation is by far the most plausible interpretation of RE within logic.[25]

In this section, we have argued that not all versions of RE about logic are equally plausible. Rather, an optimal version of RE for logic should embrace: (i) an *operationalized* understanding of logical data, with (ii) a *rich* conception of logical theories, (iii) a *wide* conception of sources of logical evidence, and (iv) a *collaborative* conception of the logical enterprise. In the following sections, we argue that the good news for advocates of RE is that there is an existent non-foundationalist epistemology of logic which fulfils these four criteria, *logical predictivism*. The bad news, however, is that predictivism also highlights three features of logic's methodology which are inconsistent with RE, at least traditionally understood.

4. Logical Predictivism

According to logical predictivism, logical theories are justified and ultimately chosen on the basis of their predictive success, explanatory power, and compatibility with other well-evidenced commitments.[26] While, according to predictivism, logics can be theories of many different phenomena, such as belief revision and grammatical structures, our concern here is with logics serving as theories of validity, understood as a property of arguments.

In order to be capable of producing both predictions to be tested against suitable data, and fruitful explanations of the target phenomenon, these logical theories are not conceived of as simply sets of valid rules of inference or theorems, but rather are a cluster of definitions, laws and representation rules that provide the underlying semantics and syntax of the theory, as well as specifying how the theory connects to the phenomenon. For illustrative purposes, here is a toy example of classical propositional logic under such an account:

Theory A

Definition 1: Let $\neg\phi$ be Boolean negation.

Definition 2: Let $\phi \rightarrow \psi$ be Boolean material implication.

Representation Rule 1: ⌜ not ϕ ⌝ = ⌜ $\neg\phi$ ⌝.

Representation Rule 2: ⌜ if ϕ then ψ ⌝ = ⌜ $\phi \rightarrow \psi$ ⌝.

Law 1: For every valuation, all sentences are either true or false, and not both.

Law 2: An argument is valid iff, for every valuation *v*, if every premise is true in *v*, the conclusion is true in *v*.[27]

---

[25] This strong evidence for the communal interpretation of the epistemology of logic is one of the few cases in which our understanding of logic's epistemology may have repercussions for specific metaphysical pictures of logic. For instance, *if* expressivist accounts of logic require one to endorse an individualistic model of logic's epistemology (as Resnik's quote suggests), then this will ultimately count against such expressivist accounts. Whether expressivists are indeed ultimately committed to such an individualistic model is another matter however, which we won't take a stand on here.

[26] The brief presentation of predictivism in this section draws upon previous work in Martin & Hjortland (2021 & 2022) and Martin (2021a).

[27] Due to space restrictions here, we are using a particularly simplistic example. The eventual theory would need to be more complex, including an account of the logic's syntax, meta-inferences, and a notion of satisfaction. For more details of what would be included within a theory, and how it would be built and developed over time, see Martin & Hjortland (2021).

According to predictivism, such theories are initially motivated by examples of arguments judged to be *acceptable*. These can either take the form of informal mathematical proofs judged to be acceptable by mathematicians, or natural-language arguments judged to be acceptable by certain "reliable reasoners".[28] For instance, the logician might initially be motivated by the following informal proofs, considered acceptable by mathematicians:

Theorem 1. *Assume* $x \in \mathbb{Z}$. *If* $x^2 - 4x + 7$ *is even, then x is odd.*

Proof. We prove our result indirectly. Suppose x is even, and let $x = 2k$ for some $k \in \mathbb{Z}$, so $x^2 - 4x + 7 = (2k)^2 - 4(2k) + 7$. Then, $(2k)^2 - 4(2k) + 7 = 4k^2 - 8k + 7 = 2(2k^2 - 4k + 3) - 1$, and so $x^2 - 4x + 7$ is odd. Thus, assuming x is even, $x^2 - 4x + 7$ is odd. □

Theorem 2. *For all* $n \in \mathbb{Z}$. *If* $3n + 2$ *is odd, then n is odd.*

Proof. We prove our result indirectly. Suppose n is even, and so $n = 2k$ for some $k \in \mathbb{Z}$. Consequently, $3n + 2 = 3(2k) + 2 = 6k + 2 = 2(3k + 1)$. But, then $3n + 2$ is even, as $2(3k + 1) = 2j$ for some $j \in \mathbb{Z}$, where $j = 3k + 1$. So, if n is even, then $3n + 2$ is even. □

Having assumed that mathematicians' judgements are a reliable (though fallible) guide as to which putative informal proofs are valid and which invalid, the logician is then concerned to provide an explanation of *why* these two proofs are valid. To do so, she first forms a general hypothesis that inferences found across multiple proofs may be valid for the same reasons, namely because they *share some underlying form*. This is the starting point (or, working assumption) for the whole enterprise of *formal* logic. Secondly, she then proposes a concrete hypothesis about the validity of the argument form which she believes the two proofs above exemplify:

Hypothesis 1

All arguments of the form

If not ψ then not φ
If φ then ψ

are valid.

This hypothesis itself, however, does *not* constitute an explanation of *why* the proofs are valid. All it offers is a generalisation which can be subsequently falsified. After all, in a similar regard, the generalization that "All swans are white" does not explain *why* swans are white. For that we need an explanatory model. In order to *explain* why arguments

[28] In recognising that logicians differentiate between *reliable* and *unreliable* reasoners, and use only inferential data from members of the former category to inform their logics, the predictivist model in effect respects the putative *normativity* of logic: that the laws of logic are not simply rules of how individuals *do* reason, but rather are rules related to *successful* reasoning in some (to be determined) sense (Steinberger 2020). Indeed, in virtue of all non-foundationalist epistemologies of logic needing to sort *reliable* from *unreliable* inferential data, all respect the normativity of logic to one extent or another. It would thus be a surprise if logic's normativity in and of itself gave us a reason to prefer one non-foundationalist epistemology of logic over another. That isn't to say, though, that one of the proposals may not gain an advantage over its competitors by being able to explain more thoroughly how logicians are able to *identify* reliable from unreliable data. In general, it is an interesting question what impact logic's putative normativity has on our current best understanding of logic's epistemology (and vice versa). However, given that the putative normativity of logic is not itself a singular position but rather a family of positions (Russell 2020), a detailed evaluation of these connections will need to wait until elsewhere. We thank an anonymous referee for pushing us on these points.

of this form are valid (if they do in fact turn out to be valid, that is), she must propose a theory such as *Theory A* above, which provides a set of rules determining both the characteristic properties of the component parts of the argument as well as the consequence relation. It is these postulates within the theory which allow us to not only differentiate the argument form from those which are invalid, but specify why this form is valid in virtue of its structure, the properties of its constituent parts (which we call the *logical constants*), and the nature of logical consequence as truth-preservation.

For example, in this particular given case, the postulates within *Theory A* provide a possible explanation of why Hypothesis 1 is true, and thus why instances of contraposition are valid, by: (i) showing how the underlying form of these arguments ensures that whenever the premises are true so is the conclusion, through a combination the theory's two definitions, representation rules and Law 1, and then subsequently (ii) using these results to show how the arguments of this form are valid, in virtue of Law 2.[29]

Now, importantly, while *Theory A* offers one possible explanation of the truth of Hypothesis 1, it is not the *only* theory that does so. There are infinitely many other theories that could. So far, all we have done is "fit" the theory to the data. Thus, *Theory A's* advocates need to find further reasons to prefer the theory over competitors. One of the main routes through which they do so is by making successful predictions about the validity of further forms of argument on the basis of the theory's postulates.

The possibility of further supporting her theory on the basis of such successful predictions is facilitated by two facts. Firstly, the postulates within her theory which putatively explain why the generalisation within Hypothesis 1 is true also ensure that other arguments are valid. In principle then, the theory can be tested against whether these further arguments are indeed valid. Secondly, given that in motivating her theory the logician assumes that mathematicians' judgements over the (un)acceptability of putative proofs are a reliable guide to their (in)validity, she can subsequently use the judgements of mathematicians to test these predictions resulting from her theory. In particular, if her theory's predictions are correct, she ought to be able to find instances of these forms of argument within informal proofs.

Testing a theory has three stages. Firstly, one draws out the consequences of the theory's postulates. In the case of *Theory A*, this would include consequences such as:

Consequence 1

All arguments of the form

$$\begin{array}{l}\varphi \\ \underline{\varphi \rightarrow \psi} \\ \psi\end{array}$$

are valid.

Consequence 2

All arguments of the form

$$\begin{array}{l}\varphi \rightarrow \psi \\ \underline{\varphi \rightarrow \neg \psi} \\ \neg \varphi\end{array}$$

are valid.

[29] For a more detailed discussion of how logics provide explanations, akin to scientific models, see Martin (2021a).

Consequence 3

Not all arguments of the form

$$\varphi$$
$$\underline{\psi \rightarrow \varphi}$$
$$\psi$$

are valid.

Notice, these consequences are expressed within the *object-language* of the theory. Secondly, therefore, in order to be tested, these consequences are then *operationalised* into testable concrete predictions. Namely, whether mathematicians find steps within informal proofs of the pertinent form acceptable or not. This requires using the theory's representation rules, just as a scientific model requires representation rules in order for the model to be tested in relation to the external target system. For instance, Consequence 1 would be operationalised as:

Prediction 1

Steps within informal proofs of the form

$$\varphi$$
$$\underline{\text{If } \varphi \text{ then } \psi}$$
$$\psi$$

are found acceptable by mathematicians.

While Consequence 3 would be operationalised as:

Prediction 2

Steps within informal proofs of the form

$$\varphi$$
$$\underline{\text{If } \psi \text{ then } \varphi}$$
$$\psi$$

are *not* found acceptable by mathematicians.

The final stage of the process is then to test these predictions against further informal proofs, not yet used to motivate the theory. Thus, the logician must at this point be engaged in considering various informal proofs, looking for instances of the forms of arguments contained within her predictions.[30] Further, given that some of her predictions cover what mathematicians do *not* find acceptable, she must also look at instances of "pseudo-proofs", where mathematicians judge inferential mistakes to have been made. Good examples of these will often be found in introductory textbooks. Ultimately, if the logician finds that mathematicians' judgements fit her theory's predictions, then the theory is further supported. Inversely, if the judgements consistently contradict its predictions, the theory faces problems. The extent to which the theory is evidenced is dependent upon its success *relative to competitors*—namely, whether the theory is more predictively successful than alternative available theories.

[30] Of course, one of the complications arising here is that the logician can be mistaken about whether an inference within an informal proof is of this relevant form. Such is the reality of interpreting data.

So far then, it's clear that predictivism fulfils our first two criteria for an optimal version of RE for logic, in virtue of having an *operationalised* account of logical data and a *richer* notion of logical theories. It also specifies how logics are first formulated and subsequently tested against the relevant data, through predictions. Admittedly, predictivism does not specify how, exactly, in the face of persistent anomalies a theory should be revised.[31] However, this should not be surprising. After all, in the face of anomalies in the sciences, there are multiple ways in which a theory can be revised to "save the data", the comparative reasonableness of which will depend on the particular theory (or, research programme's) state of development. Such moderate anarchy when it comes to theory revision is palatable however, because ultimately the resultant competing revised theories must all face the tribunal of being tested via their predictions. From this actively competitive enterprise, where different logicians (and research groups) propose solutions to such anomalies and these proposals are subsequently evaluated by the (lack of) success of their predictions, it's clear that predictivism also views logic as a *communal* endeavour. Logicians are not free to rest in any reflective equilibrium they see fit. They are committed to test their proposals in accordance with the communally recognised data and evidential standards.

Predictive success is not the whole story when it comes to logical evidence, however. In addition to this direct evidence for logics in the form of judgements regarding the (un)acceptability of specific inferences, predictivism admits three further forms of *indirect* evidence which can motivate revisions to an existent logical theory (though, notably, not provide evidence *for* a particular resulting theory).

Firstly, there are instances of what are known as *bad company*. These occur when logicians do not have direct evidence against the validity of an argument form *F*, but rather reject it because admitting the validity of its instances would require admitting the validity of instances of another form of argument *F′* which they *do* have direct evidence against, in the form of unacceptability judgements regarding its operationalised surrogates.

A famous example of *bad company* is the relevant logician's rejection of the disjunctive syllogism, which is not rejected because the rule has obvious operationalised instances which are judged to be unacceptable. After all, the relevant logician admits that we need a relevantly valid analogue to replace the seemingly reasonable, but ultimately mistaken, classical rule in order to sanction these acceptable inferences found within informal proofs (Burgess 1983). Instead, the disjunctive syllogism is rejected because, in combination with the rule of addition and a suitable definition of validity, the rule entails the validity of explosion, which *does* have operationalised instances which we judge to be unacceptable (at least, according to the relevant logician).

Thus, *bad company* ensures that in virtue of having direct evidence against the validity of argument form *F′*, the logician can subsequently have good reasons to reject (the conjunction of) those argument forms which require us to accept the validity of *F′*. In such cases, the logician is required to make some adjustment within her theory to ensure the troublesome form *F′* is invalidated.[32]

Now, of course, there will be numerous theoretical adjustments the logician can make to block these unsavoury consequences, just as in cases of *direct evidence* against an argument form. Thus, bad company arguments do not themselves directly lend support to a new theory. The arguments only serve to remove certain candidates from the table—

[31] Though it is able to detail the various options open to logicians in the face of recalcitrant data, via alterations to component parts of the theory; see Martin & Hjortland (2021).

[32] The logician may also have the option to reinstate *F′*, and thereby be forced to reject the evidence against it. Whether this is a viable option will depend upon how *robust* the evidence against *F′* is (see Section 5 below).

namely, those that commit the logician to the validity of *F′* via *F*. In this sense, *bad company* arguments in logic serve a similar function to *internal consistency constraints* on empirical theories (McMullin 2008). In order to find discriminating support for the remaining candidates, new consequences must be drawn from each, predictions tested, and their relative successes compared.

The same is true of the second type of indirect evidence against a theory, called *post hoc rejections*. These rejections arise when a particular argument form clashes with fundamental elements of the theory, and thus suitable adjustments must be made to ensure the form isn't sanctioned by the theory. Thus, in this case the invalidity of *F* isn't determined on the basis of unacceptability judgements regarding putative instances of the form, but rather because *F* fails to meet certain requirements dictated by the theory's laws.

A nice example of *post hoc rejection* comes again from relevant logic, where the axiom,

(A) $\neg(\varphi \rightarrow \varphi) \rightarrow (\psi \rightarrow \psi)$

is omitted from relevant logics not because direct evidence can be produced against it (in the form of unacceptability judgements regarding putative instances), but because it contravenes a law of relevant logics, that an argument is only valid if it adheres to *variable sharing* (Belnap 1960).

The rationale for such rejections is that if an argument form can be shown to be incompatible with a fundamental component of the standing theory, which itself has been previously shown to be well-supported (as part of the theory) through predictive success, then one has good evidence against the incompatible argument form. In the particular case of the "*variable sharing* law", this restriction on the consequence relation was introduced in order to *explain why* other relevantly invalid argument forms, such as explosion, which we do putatively have direct evidence against, are indeed invalid. Consequently, if the relevant logician were to admit **(A)** into the logic they would in effect be undermining their own (fruitful) explanation for the (in)validity of other argument forms.

As with instances of bad company, *post hoc* rejections do not provide any additional positive evidence for a particular theory. After all, they are motivated solely by the postulates and laws already included within the theory. The justification for excluding the argument form will then be wholly dependent upon the existent justification one had for those laws precluding it. Such rejections, therefore, are simply another means to ensure *internal consistency*. Ultimately, the justification one has for a *post hoc* rejection will depend upon the justification one has for the (putatively incompatible) existent postulates within the theory, which can only be gained through predictive success.

The final form of indirect evidence, *clashes with other theoretical commitments*, rather than being akin to internal consistency constraints in the sciences, is related to what Kuhn (1977: 321-3) called "external consistency". Such clashes occur when we combine our logical theory with independently evidenced commitments, and it's shown that the conjunction of our theory with these commitments cannot be true.

Probably the most famous of these clashes arise when we combine our logic with a theory of truth and the incompatibility between the two is brought to our attention via a paradox, whether this be the liar or the Curry. In the face of this putative incompatibility, we must then either revise our logic, revise the relevant independently evidenced commitment, or explain away the apparent incompatibility.

For instance, assume that so far we've found good reason to accept classical logic due to its predictive success. Further, that we also have good independent reasons to accept both the transparency of the truth predicate and the

semantic closure of our natural languages. The former, perhaps, on the basis that it allows us to make blind belief ascriptions to others (Kripke 1975), and the latter because this is what linguists tell us about these languages based upon empirical evidence. For a period of time, we may be content that our three commitments—classical logic, a transparent truth predicate, and the semantic closure of natural languages—are compatible with one another. All is well. But then, a clever associate (Curry 1942) points out that the putative semantic closure of our language allows us to form problematic self-referential sentences such as,

(*C*) If *C* is true, then 0=1,

which, given our further commitments to classical logic and the transparent truth predicate, allows us to infer 0=1.

Given that we have excellent reasons to reject 0=1, and further recognise that variations of (*C*) can be used to commit us to *any claim* we don't wish to be, we come to the conclusion that one of our three prior commitments must go. In the case that we think the evidence in favour of a transparent truth predicate and semantic closure are just too strong, then it is classical logic which must be revised to block these unsavoury consequences.[33]

Note again, however, that many such alterations to our logic will do the job. All that is strictly required to ensure *external consistency* is to make the necessary adaptations to block the unsavoury consequences above. There are a whole host of options for achieving this, including paraconsistent, paracomplete, and substructural proposals. Thus, being able to "provide a solution" to the paradox and re-establish external consistency is not enough. Ultimately, the proposed theory must be tested against competitors via the comparative success of its predictions. This last form of indirect evidence shows how the predictivist model meets our final criterion for an optimal version of RE, by admitting a *wider source* of logical evidence in the form of further well-evidenced theoretical commitments, while possessing the advantage of being able to detail how the success of any resulting revised theory is subsequently tested.

5. Deviations from Reflective Equilibrium

The picture painted by the predictivist model may seem to be totally in keeping with the RE proposal. After all, we have a (logical) theory which is motivated and revised on the basis of some data, which takes the form of judgements regarding the (un)acceptability of specific inferences. Further, as we've noted, predictivism fulfils all four of the criteria we specified for an optimal version of RE about logic.[34] This seems good news for the advocate of RE about logic, as we may have found a model of logic's epistemology that vindicates their underlying proposal.

Admittedly, advocates of RE may feel somewhat uneasy at predictivism's talk of a logic *making predictions*, and further that the only evidence it admits in support of a logic (rather than against it) comes from its *predictive success* and *explanatory power*, relative to competitors.[35] After all, predictive success is not the same as achieving a state of coherence through mutual adjustment, as initial articulations of RE suggested the position was concerned with (Goodman 1983).

[33] This, of course, is the kernel of Priest's (2006b) argument for dialetheism, though based upon the liar sentences.

[34] Indeed, predictivism is the only existent detailed account of logic's epistemology we are aware of that fulfils all four of our criteria for an optimal version of RE about logic.

[35] The term "prediction" has slipped into some presentations of RE, notably Woods (2019a). However, some advocates of RE in logic have explicitly *precluded* the possibility of logical theories making predictions (Peregrin & Svoboda 2021: 20).

Let us assume for the sake of argument, however, that advocates of RE are able to explain away this apparent difference between RE and predictivism, perhaps interpreting the latter's appeal to predictive success in terms of the virtue of "fruitfulness" admitted by RE (Brun 2020).[36] Even in this case, there are three important features of logic's methodology which predictivism brings to the fore, all of which seem contrary to what RE proposes. These serve not only to show that predictivism is distinct from RE, but that RE seems in tension with prominent elements of logic's epistemology.

5.1 Three Prominent Features of Logic's Methodology

Firstly, *certain logical data is robust*. While RE allows for the data informing our theories to have prior credibility, RE requires that in order for our belief in this data to be justified it must ultimately be shown to cohere with our resulting theory (Brun 2020). Thus, in the particular case of logic, while our judgements regarding the (un)acceptability of specific inferences may have some prior credibility, allowing them to inform our theory, in order to justify our belief that these specific inferences actually are (un)acceptable, these judgements have to be shown to cohere with our resulting logical theory. Further, this requirement ensures that it *must* always be possible that a specific judgement is altered if it cannot be brought into line with the theory *and* the theory itself cannot be suitably altered. This is just what is required by the "equilibrium" in reflective equilibrium, and ultimately makes RE "weakly foundationalist".

However, as predictivism highlights, there are certain cases of *robust data* in logic which are not up for revision. No logical theory, for instance, would be deemed acceptable unless it could accommodate important inferences made within informal mathematical proofs (instances of *modus ponens*, conditional proof, proof by cases, etc.). It is not an option for logicians to simply recategorize these inferential moves as unacceptable. It is for exactly this reason, as noted above, that non-classical logicians have spent so much energy attempting to show that these inferences can be sanctioned even when certain classically valid rules have been invalidated.

The robustness of this data is further shown by the historical attitude towards mathematical proofs which could not be suitably modelled and sanctioned by Aristotelian syllogistic logic. It was well known, from the sixteenth century onwards, that proofs contained within Euclid's *Elements* could not be sanctioned by syllogistic logic (Mugnai 2010). However, at no point was this considered to count against the acceptability of the inferences within these proofs, nor to justify a revision of these judgements to bring them into line with the established (logical) theory. Rather, it was recognised that (eventually) a superior theory would need to be found, and this was of course one of the motivating factors behind the development of mathematical logic. Not all logical data, therefore, has the plasticity that RE presumes.[37]

---

[36] Though, in all likelihood, advocates of predictivism won't be satisfied with this assimilation, given that they criticise abductive accounts of logic's epistemology for making the same unspecified appeals to "fruitfulness" (Martin & Hjortland 2021). However, there are other notable and important differences between predictivism and RE worth focusing on here.

[37] Of course, the robustness of this data again raises the question of logic's *normativity*, for what reassurances do we have that this inferential data is indeed reliable? Why should the theory have to conform to the data, rather than the inverse? If logic is meant to be normative for reasoning in some sense, one would expect that its laws could be held constant despite how individuals *do* in fact reason, given that the inferential data could be logically deviant. This is an interesting question and deserves further discussion elsewhere. However, two points are important to make here. Firstly, whether it is puzzling or not, the robustness of certain data—in particular, the inferences mathematicians make—is a fact of logic's methodology. We cannot ignore it simply because we admit that logic has some normative force. Secondly, at least according to predictivism, these data are treated as robust *because* they are treated as inherently reliable, which thereby respects the normativity of logic. Now, it is an interesting *why* logicians treat this data as inherently reliable, and whether they are justified in doing so (Martin

Secondly, *logics attempt to explain.* According to RE, the state of (reflective) equilibrium between the theory, data, and (in the case of *wide* RE) wider commitments, suffices to justify our belief in the theory. However, logicians desire more from their theory. They do not wish for it to simply cohere with the available data. They wish for the theory to effectively *explain* the theory's target phenomenon. In the present case, this means explaining *why* certain arguments are valid and others invalid.[38]

This desire to provide an effective explanation of the (in)validity of arguments is most forcefully shown when logicians disagree over their theory of validity whilst agreeing over the *extension* of logical consequence. What occurs in these cases are extensionally identical logics with different semantics, which are favoured in virtue of their perceived explanatory power. Two examples will suffice here.

Firstly, while we have equally well mathematically formulated model-theoretic and proof-theoretic accounts of validity that can deliver a classical consequence relation, logicians find reasons for preferring one of these over the other on the basis that it is explanatorily superior. For instance, advocates of the proof-theoretic account have argued that it is explanatorily superior because it is able to specify the discrete steps needed to demonstrate that a given argument is valid, unlike the model-theoretic account (Prawitz 1985). Further, unlike the model-theoretic account, it does not require us to have a prior notion of (possible) "models" or "cases" to determine an argument's validity, which we must if we are to make sense of quantifying over all suitable models (Etchemendy 1990). In comparison, model-theoretic accounts have been deemed explanatorily more powerful as they are able to specify the exact countermodels which demonstrate why a particular argument is invalid, as well as providing the counterfactual conditions under which alterations to an argument's logical form would make it either valid or invalid (Martin 2021a).

Secondly, within relevant logic there are (or, were) two competing research programmes: the Australian plan, which used a possible-worlds semantics, and the American plan, which used a multi-valued semantics (Read 1988). Both were capable of delivering the same logics, extensionally understood. However, the Australian plan was widely criticised for requiring the use of a star operator, known as the Routley star (Routley & Routley 1972), in order to express negation, on the basis that it was totally unclear how to comprehend what this operation on worlds amounted to (Copeland 1979; Hintikka 1981; van Bentham 1979).[39] According to these critics, given that one of the functions of a logic's semantics is to elucidate *why* the eventual argument forms come out as valid, the Australian plan's semantics missed their mark (Copeland 1979). Both examples serve to show that logicians desire more from their theory than being in reflective equilibrium with the available data. They also require it to effectively *explain* the target phenomenon.

Lastly, it can be rational to continue to endorse a theory even when there is existent data that is *inconsistent with the theory* and revising the data *isn't currently an option*. This is a product of two features of logic's methodology. Firstly,

---

& Hjortland 2022: Sect. 4), however such a concern seems no different in kind to general fallibilist concerns over whether *any* research area is justified in treating a given set of data as robust. We thank an anonymous referee for raising this concern.

[38] Note, this is not the same as explaining why *individuals find certain inferences (un)acceptable*. This would be to confuse the data with the phenomenon (Martin & Hjortland 2021). It is for this reason that the explanatory role of logics provides some evidence against non-factualist accounts of logic (such as logical expressivism) which conceive of logic's laws as simply the codification of those inferences individuals find (un)acceptable. In such cases, there is *no phenomenon* beyond the data *to explain*. Here again is one of those rare occasions in which a feature of logic's epistemology provides evidence against a particular metaphysical account of logic.

[39] This disillusionment with the Routley star isn't universal (see Restall 1999). However, that isn't our point. Rather, the fact this disillusionment even occurs highlights that explanatory power is a criterion for a successful logical theory.

that the process of evaluating logics is inherently competitive. Thus, it is rational to endorse a logic which has been more (predictively) successful than its existent competitors, even if it has faults. Secondly, that some of the data and wider theoretical commitments that inform our logical theories are robust, meaning that they cannot be revised just because they contradict our standing theory. Yet, according to RE, if our theory and data clash, they *must* be brought into line in order to establish an equilibrium, whether through revising our theory or the data.

We have already cited one example of this phenomenon in the history of logic, with the recognised failure of Aristotelian syllogistic logic to sanction important inferences within mathematical proofs. This recognised failure, however, wasn't enough to automatically bring about a revision in the theory or the data. After all, it would be wholly irrational to simply reject an otherwise successful theory because of some (admittedly important) anomalies, when no better theory is available to us. In this sense, logical theorizing is no different to that in the sciences.

A similar example is the present situation with certain logico-semantic paradoxes and classical logic. Classical logic is an extremely successful theory, but is seemingly incompatible with other well-evidenced commitments, such as a transparent truth-predicate and semantic closure, as shown by the self-referential paradoxes. However, even while admitting this incompatibility, it may still be wholly unclear how to resolve the problem exactly. For while there may be a reticence to revise either of these wider commitments given how well-evidenced they are, the cost of revising our logic is considered just far too great (Williamson 2017). In this case, revising to a weaker non-classical logic in order to avoid these unsavoury consequences is deemed to be tantamount to throwing the baby out with the bathwater. In this circumstance, it can be rational to (and logicians *do*) continue to hold the established theory even if one has no clear answer as to how to address these apparent anomalies. As long as the anomalies are not so pervasive as to undercut the theory's overall theoretical strengths they can be tolerated until a better theory, or resolution, comes about.

5.2 Tension, or No Tension?

Here then we have three aspects of logic's epistemology which predictivism brings to the foreground that seem incompatible with RE about logic, at least as traditionally conceived. This means that, despite meeting the four criteria set out for an optimal version of RE about logic in Section 4, predictivism isn't the success for RE that advocates might have hoped. Given this, in order to explain away these potentially challenging features of logic's epistemology, advocates of RE about logic face a choice. Either they must:

(i) Show that, contrary to appearances, none of these three putative features of logic's epistemology *actually are* aspects of logic's epistemology. This would result in the rejection of logical predictivism, and subsequently require the advocate of RE to provide us with another suitable model of logic's epistemology which meets the four criteria for an optimal version of RE for logic; or,

(ii) Admit these three aspects of logic's epistemology but argue that contrary to appearances they are in fact *consistent with* RE about logic. Yet, this option comes with the potential risk of both amalgamating RE about logic with predictivism, and significantly widening the notion of "reflective equilibrium" beyond what was originally proposed by Goodman in his initial presentation of the position.

We have no particular view here on which option the advocate of RE should ultimately take. We shall, however, briefly note the potential complications that arise from choosing one particular version of this latter option, given that we expect

some will be tempted by it. Could the advocate of RE about logic propose that the process of RE itself was never intended to deliver us with an epistemology of *fully-fledged logical theories*, but rather is simply an account of the *initial process* in the development of the correct *logical principles*, whether understood as valid argument forms or rules of inference? These principles would then serve, so the response goes, as the building blocks for a fully-fledged logical theory, with this latter theory-construction governed not by RE but rather another process (such as that outlined by predictivism). This reply would have the further potential benefit of explaining away any apparent tension between RE about logic and the three features of logic's epistemology highlighted above, given that the latter only arise at the level of *theory choice*.[40]

Assigning RE this more limited role is certainly an interesting proposal and has been suggested before by Woods (2019a) when attempting to differentiate RE from logical abductivism. However, despite this, a few problems with the proposal are apparent. Firstly, restricting RE to this role would mean that it should ultimately be integrated into a more comprehensive epistemology of logic (such as predictivism or abductivism), with the latter treating the mechanisms RE proposes as constituting the initial stages of theory-building in their own proposal. In which case, it's unclear what the benefit of discussing RE as a standalone epistemology of logic would be. In particular, the proposed comprehensive epistemology of logic, of which RE constitutes a *stage*, would be able to explain what RE about logic can and more. Further, it is unclear that most advocates of RE about logic would be happy with this proposal, as they seem reticent to assign RE this more limited role (see, for example, Brun 2020). While this reticence isn't a conclusive reason to not interpret RE in this fashion, we should certainly take it into account.

Secondly, it is unclear how to conceive of the "logical principles" that would result from this initial RE process. After all, if RE fails to deliver a fully-fledged theory, they cannot be in the object language of the logic. They must be expressed in some non-formal or semi-formal language instead, such as (regimented) English. Yet, the business of logic is to provide a *formal* account of validity, using mathematical apparatus, which can only be done within the framework of a fully-fledged logical theory. At best, RE would be providing us with *quasi*-logical principles, using something like regimented English. Further, given that (as we argued above) "validity" is a technical term defined within logics, RE would not be providing us with the *valid* logical principles, but at best something like principles of *quasi*- or *intuitive*-validity. Certain advocates of RE about logic seem to be aware of this point (Peregrin & Svoboda 2017), and it is another potential reason that advocates of RE would not want to resign RE to this more limited role.

Relatedly, restricting RE to this role still implies that the process delivers us with some final, fixed, accepted logical principles, and that the responsibility of these further theory-building stages in the process are simply to construct a fully-fledged formal logic which respects these correct principles. Yet, as predictivism shows, theory-building is not such a linear process. There are mechanisms in logical theory choice which require us to *reject* logical principles even if we do not have judgements about their instances, such as *ad hoc rejection*. This means that the set of "accepted" principles will alter throughout the process of theory-building—they do not remain fixed. Indeed, in some cases, we can have positive evidence for a logical principle through judgements regarding instances but subsequently have to reject the principle because it conflicts with a fundamental feature of the theory (as with the relevant logician's case of *Mingle* noted earlier). The same, of course, may also happen when we begin to combine other commitments with our logical theory. This means that we could initially deem all of classical logic's rules of inference as valid, in virtue of cohering with our judgements over their instances, but then subsequently come to reject some of them because of the complications that arise with the

[40] We thank an anonymous referee for suggesting this potential reply.

logico-semantic paradoxes. Thus, we cannot treat the logical principles (whatever they are) that result from this initial stage of logical theorizing as somehow fixed. This means that even given this restricted role, RE would not necessarily be delivering with us the "valid" logical rules, because the process of theorizing can lead to their revision.

Yet, even if we were to put these concerns over interpreting RE about logic in this more restricted fashion to one side, it is unclear that this would make RE about logic compatible with the three features of logic's methodology we highlighted above. While assigning RE this more restricted role would allow the proposal to sidestep the need to accommodate the explanatory function of logics, which only occurs at the level of *theory* rather than individual principles, the same is not true of the remaining two features of logic's methodology. Both the *robustness of data* and the recognition of *existent contradictory data* for even our accepted logical principles/theory do not arise only at the level of theory-choice. They call into question the more restricted proposal that assessing logical *principles* is a matter of required *mutual adjustment*. After all, if some data is robust, as much of logical practice suggests, then logical theorising is not a simple case of symmetric give and take between data and principles, which is often how RE is characterised. Further, as has been shown by the case of syllogistic logic, it may well be that while we recognise a clash between this robust data and our currently accepted logical principles, the best course of action is to provisionally accept this inconsistency until some theoretical innovation comes along, given that the current theoretical costs of rejecting either the data or the principles would be too high. Again, mutual adjustment into a state of equilibrium is not always mandatory or rationally possible, even if we just restrict ourselves to principles.

This brief consideration of just one of the options facing advocates of RE shows that it will not be a simple matter to either explain away these prominent features of logic's epistemology or show that RE is consistent with them.

6. Conclusion

In this paper, we have sought to evaluate RE about logic. A position which, though relatively popular, is still underspecified in certain important regards. We began by highlighting those aspects of the position that require further clarification, and subsequently identified how an optimal version of RE about logic should answer these requests for clarification. We then presented another existent model of logic's epistemology, logical predictivism, and showed how it fulfils our four criteria for an optimal version of RE about logic. This, on the face of it, was good news for the advocate of RE. However, we then showed that predictivism brings to the fore three aspects of logic's epistemology which run contrary to the principles of RE, at least traditionally conceived. This leaves advocates of RE about logic with a challenge: either (i) show that (contrary to appearances) these are not aspects of logic's epistemology and provide a new better optimal version of RE about logic, or (ii) admit these features of logic's epistemology, at the potential cost of amalgamating RE with predictivism and widening the notion of RE beyond what Goodman originally proposed.[41]

[41] **Acknowledgments:** I am grateful to audiences at the European Network for the Philosophy of Logic and the Sociedade Brasileira de Filosofia Analítica for their feedback on earlier versions of this paper, as well as insightful discussions with Massimiliano Carrara, Filippo Ferrari, Ole Hjortland, and Ulf Hlobil. I would also like to thank two anonymous referees for their extremely useful suggestions. This work was supported by a PNRR grant, under the European Union's NextGenerationEU research and innovation programme, as well as the DFG (Deutsche Forschungsgemeinschaft), project no. 288923097, FOR 2495.

References

Anderson SR (1971) On the Role of Deep Structure in Semantic Interpretation. *Foundations of Language*, 7: 387–96.

Anderson AR & ND Belnap (1975) *Entailment: The Logic of Relevance and Necessity (Vol. I)*. Princeton, NJ: Princeton University Press.

Ayer AJ (1936) *Language, Truth and Logic*. New York, NY: Dover.

Baderin (2017) Reflective Equilibrium: Individual or Public? *Social Theory and Practice*, 43: 1–28.

Barrio E, F Pailos & D Szmuc (2020) A Hierarchy of Classical and Paraconsistent Logics. *Journal of Philosophical Logic*, 33: 93–120.

Baumberger C & G Brun (2021) Reflective Equilibrium and Understanding. *Synthese*, 198:7923–47.

Beall, Jc (2013) Shrieking Against Gluts: The solution to the 'just true' problem. *Analysis*, 73: 438–45.

Belnap ND (1960) Entailment and Relevance. *The Journal of Symbolic Logic*, 25: 144–46.

Benzmüller C, X Parent & L van der Torre (2020) Designing Normative Theories for Ethical and Legal Reasoning: LogiKEy framework, methodology, and tool support. *Artificial Intelligence*, 287: 103348.

Blake-Turner C & G Russell (2021) Logical Pluralism without the Normativity. *Synthese*, 198: 4859–77.

Bogen J & J Woodward (1988) Saving the Phenomena. *The Philosophical Review*, 97: 303–52.

BonJour L (1998) *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

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

Brun G (2014b) Reflective Equilibrium Without Intuitions? *Ethical Theory and Moral Practice*, 17: 237–52.

Brun G (2020) Conceptual Re-Engineering: from explication to reflective equilibrium. *Synthese*, 197: 925–54.

Burgess JP (1983) Common Sense and "Relevance". *Notre Dame Journal of Formal Logic*, 24: 41–53.

Copeland BJ (1979). On when a Semantics is not a Semantics: Some reasons for disliking the Routley Meyer semantics for relevance logic. *Journal of Philosophical Logic*, 8: 399–413.

Cook R T (2010) Let a Thousand Flowers Bloom: A tour of logical pluralism. *Philosophy Compass*, 5: 492–504.

Cummings R (1998) Reflection on Reflective Equilibrium. In: MMR DePaul and W Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* (pp. 113–27). Lanham: Rowman & Littlefield.

Curry HB (1942) The Inconsistency of certain Formal Logics. *Journal of Symbolic Logic*, 7: 115–7.

Dalrymple M (2001) *Lexical Functional Grammar* (Syntax and Semantics Series, vol. 34). New York: Academic Press.

Daniels N (1979) Wide Reflective Equilibrium and Theory Acceptance in Ethics. *Journal of Philosophy*, 76: 256–282.

DePaul MR (1993) *Balance and Refinement: Beyond coherence methods of moral inquiry*. London: Routledge.

dos Santos CF (2021) Intuitions, Theory Choice and the Ameliorative Character of Logical Theories. *Synthese* 199: 12199–223.

Etchemendy J (1990) *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.

Field H (2008) *Saving Truth from Paradox*. Oxford: Oxford University Press.

Foley R (1993) *Working Without a Net: A study of egocentric epistemology*. Oxford: Oxford University Press.

Gabbay DM (1994). What is a Logical System? In: DM Gabbay (eds.) *What is a Logical System?* (pp. 71–120). Oxford: Oxford University Press.

Glanzberg M (2015). Logical Consequence and Natural Language. In: C Caret & O Hjortland (eds.) *Foundations of Logical Consequence* (pp. 179–216). Oxford: Oxford University Press.

Goel V, B Gold, S Kapur & S Houle (1997) The Seats of Reason? An imaging study of deductive and inductive reasoning. *NeuroReport* 8: 1305–10.

Goodman N (1983) *Fact, Fiction, and Forecast*, 4th ed. Cambridge, MA: Harvard University Press.

Grove A (1988) Two Modellings for Theory Change. *Journal of Philosophical Logic*, 17: 157–170.

Halldén S (1949) *The Logic of Nonsense*. Uppsala: Uppsala Universitets Arsskrift.

Hintikka J (1981) Intuitions and Philosophical Method. *Revue Internationale de Philosophie*, 35: 74–90.

Hjortland J (2019) What Counts as Evidence for a Logical Theory? *Australasian Journal of Logic*, 16: 250–82.

Keefe R (2008) Vagueness: Supervaluationism. *Philosophy Compass*, 3: 315–24.

Kripke S (1975) Outline of a Theory of Truth. *Journal of Philosophy*, 72: 690–716.

Kuhn T (1977) *The Essential Tension*. Chicago, IL: University of Chicago Press.

Lewis D (1983) *Philosophical Papers* (Vol I). Oxford: Oxford University Press.

Lewis CL & CH Langford (1932) *Symbolic logic*. London: Dover Publications.

Maddy P (2007) *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.

Martin B (2021a) Anti-Exceptionalism about Logic and the Burden of Explanation. *Canadian Journal of Philosophy*, 51: 602-18.

Martin B (2021b) Identifying Logic Evidence. *Synthese*, 40: 9069–95.

Martin B (2021c) Searching for Deep Disagreement in Logic: The Case of Dialetheism. *Topoi*, 40: 1127-38.

Martin B (2022) The Philosophy of Logical Practice. *Metaphilosophy*. 53: 267-83.

Martin B & OT Hjortland (2021) Logical Predictivism. *Journal of Philosophical Logic*, 50: 285–318.

Martin B & OT Hjortland (2022) Anti-Exceptionalism as Tradition Rejection. *Synthese*. Online first: 10.1007/s11229-022-03653-7

Martin B & OT Hjortland (Forthcoming) Evidence in Logic. In M Lasonen-Aarnio & C Littlejohn (eds.), *Routledge Handbook for Philosophy of Evidence*. London: Routledge.

McMullin E (2008). The Virtues of a Good Theory. In: S Psillos & M Curd (eds.) *The Routledge Companion to Philosophy of Science* (pp. 498–508). London: Routledge.

Meyer RK (1971) *R*-Mingle and Relevant Disjunction. *Journal of Symbolic Logic*, 36: 366.

Mugnai M (2010) Logic and Mathematics in the Seventeenth Century. *History and Philosophy of Logic* 31: 297–314.

Peacocke C (1976) What is a Logical Constant? *Journal of Philosophy*, 73: 221–40.

Pelletier FJ, R Elio & P Hanson (2008) Is Logic all in our Heads? *Studia Logica*, 86: 1–65

Peregrin J & V Svoboda (2017) *Reflective Equilibrium and the Principles of Logical Analysis: Understanding the Laws of Logic*. London: Routledge.

Peregrin J & V Svoboda (2021) Moderate Anti-Exceptionalism and Earthborn Logic. *Synthese*, 199: 8781–806.

Prawitz D (1985) Remarks on some Approaches to the Concept of Logical Consequence. *Synthese*, 62: 153–71.

Priest G (2006a) *Doubt Truth to be a Liar*. Oxford: Clarendon Press.

Priest G (2006b) *In Contradiction: A Study of the Transconsistent*, 2nd ed. Oxford: Clarendon Press.

Priest G (2016) Logical Disputes and the a Priori. *Logique et Analyse*, 59: 347–66.

Quine WVO (1951) Two Dogmas of Empiricism. *Philosophical Review*, 60: 20–43.

Rawls J (1971) *A Theory of Justice*. Cambridge, MA: Belknap Press.

Read S (1988) *Relevant Logic*. Oxford: Blackwell.

Resnik MD (1999) Against Logical Realism. *History and Philosophy of Logic*, 20: 181–94.

Restall G (1999) Negation in Relevant Logics (How I Stopped Worrying and Learned to Love the Routley Star). In: DM Gabbay and H Wansing (eds.), *What is Negation?* (pp. 53–76). Dordrecht: Springer.

Rossberg M & S Shapiro (2021) Logic and Science: Science and logic. *Synthese*, 199: 6429–54.

Routley R & V Routley (1972) The Semantics of First Degree Entailment. *Noûs*, 6: 335–59.

Russell G (2015) The Justification of the Basic Laws of Logic. *Journal of Philosophical Logic*, 44: 793–803.

Russell G (2019) Deviance and Vice: Strength as a theoretical virtue in the epistemology of logic. *Philosophy and Phenomenological Research*, 99: 548–63.

Russell G (2020) Logic isn't Normative. *Inquiry* 63: 371–88.

Schroeder-Heister P (2006) Validity Concepts in Proof-Theoretic Semantics. *Synthese*, 148: 525–71.

Shapiro S (2000) The Status of Logic. In: C Peacocke and PA Boghossian (eds.), *New Essays on the A Priori* (pp. 333–66). Oxford: Oxford University Press.

Sher G (1991) *The Bounds of Logic*. Cambridge: MIT Press.

Slater H (1995) Paraconsistent Logics? *Journal of Philosophical Logic*, 24: 451–4.

Steinberger F (2020, Winter) The Normative Status of Logic. In: EN Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Online: https://plato.stanford.edu/archives/win2020/entries/logic-normative/

Tennant N (2004) An Anti-Realist Critique of Dialetheism. In: G Priest, Jc Beall & B Armour-Garb (eds), *The Law of Non-Contradiction: New Philosophical Essays* (pp. 355–84). Oxford: Clarendon Press.

Varzi A (2007) Supervaluationism and its Logics. *Mind*, 116: 633–75.

Warren J (2020) *Shadows of Syntax*. Oxford: Oxford University Press.

Weisberg M (2013) *Simulation and Similarity: Using models to understand the world.* Oxford: Oxford University Press.

Williamson T (1994) *Vagueness*. London: Routledge.

Williamson T (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson T (2017) Semantic Paradoxes and Abductive Methodology. In: B Armour-Garb (ed), *The Relevance of the Liar* (pp. 325–346). Oxford: Oxford University Press.

Woods J (2019a) Against Reflective Equilibrium for Logical Theorizing. *Australasian Journal of Logic*, 16: 319–41.

Woods J (2019b) Logical Partisanhood. *Philosophical Studies*, 176: 1203–24.

Woody AI (2015) Re-orienting Discussions of Scientific Explanation: A Functional Perspective. *Studies in History and Philosophy of Science*, 52: 79–87.

Wright C (1985). Inventing Logical Necessity. In: J Butterfield (ed.), *Language, Mind, and Logic* (pp. 187–209). Cambridge: Cambridge University Press.

Zardini E (2011) Truth without Contra(di)ction. *Review of Symbolic Logic*, 4: 498–535.