Synthese (2024) 203:188
https://doi.org/10.1007/s11229-024-04626-8

ORIGINAL RESEARCH

Logical abductivism on abductive logic

Filippo Mancini[1]

Received: 6 December 2023 / Accepted: 6 May 2024 / Published online: 31 May 2024

Abstract
Logical abductivism is the epistemic view about logic according to which logical theories are justified by abduction (or Inference to the Best Explanation), that is on how well they explain the relevant evidence, so that the correct logical theory turns out to be the one that explains it best. Arguably, this view should be equally applied to both deductive and non-deductive logics, abduction included. But while there seems to be nothing wrong in principle in using abduction to determine the correct logical theories of deduction and induction, things might be more complicated regarding logical theories of abduction. We may wonder whether allowing for an abductive justification of a theory of abduction is an epistemically legitimate move, since here circularity casts its shadow and makes the situation darker. This is the issue to which this work is devoted. I will defend that, to be effective, an abductive justification for a theory of abduction calls for a justification of abduction in advance, which we do not yet have.

Keywords Abduction · Inference to the best explanation · Logical abductivism · Circularity · Suasive and explanatory arguments

1 Introduction

Over the past few years, logical epistemology has been gaining much attention. There has been considerable discussion on many epistemic issues about logic,[1] and one of them in particular has attracted the interest of philosophers: its methodology. Several reasons can be found behind this specific renewed interest, but arguably the most

[1] Some examples are the properties traditionally associated to logic, such as *apriority*, *generality*, *necessity*, and *formality*. But beyond the discussion around these issues—as well as the one relevant to this work, namely, the methodology of logic—, the recent focus on the epistemology of logic is evidenced by two major debates that have emerged in the literature: the one on the normative status of logic and the one on logical monism/pluralism. On the former, see e.g. Russell (2020) and Oza (2020). As for the latter, see e.g. Ferrari and Orlandelli (2019), Stei (2020) and French (2021). For some significant investigations into the relationship between logical pluralism and the normativity of logic, see e.g. Ferrari and Moruzzi (2020) and Steinberger (2019).

✉ Filippo Mancini
fmancini@uni-bonn.de

[1] Department of Philosophy, University of Bonn, Bonn, Germany

relevant is the large number of logical theories available today. For as far as different logics are designed to serve distinct purposes, no epistemic issue emerges. But when incompatible logics aim to model the same matter, the question arises as to which one to choose. And this is exactly what happens with what most philosophers consider the *raison d'être* (or canonical application) of logic: validity (valid reasoning). To better understand this point and to lay the ground for the discussion, let me put on the table and make explicit some important basic assumptions.

The word "logic" is ambiguous and appropriate disambiguation is needed to start the investigation. A large proportion of philosophically minded logicians think of it as "a theory about what follows from what and why",[2] and I do the same in this paper. Following Priest (2014), this way of understanding logic corresponds to (or at least is part of) *logica docens*—i.e., "the logic that is taught" (Priest, 2014, p. 212)—, and must be distinguished from what it is a theory of, i.e. *logica ens*, that is "what *is* actually valid: what really follows from what" (Priest, 2014, p. 212, italics in original). Thus, in line with this, the aim of logicians is to develop a[3] logical theory (or logic) capable of accounting for *logica ens*, that is modeling valid[4] reasoning, discriminating between valid and invalid (forms of) inferences and explaining why it is so. Such a theory, call it T_L, aspires to be correct (or, with some flexibility, true) in the following sense: T_L is correct if and only if (henceforth, iff) every inference that is T_L-valid is also actually valid—or "genuinely valid", following Field (2015). It is worth noting that in this context correctness should not be confused with soundness, although they are related. They are two distinct notions for the following reasons: first, correctness depends on the match between T_L-validity and genuine validity, whereas soundness depends on that between T_L-validity and T_L-provability; second, correctness is a property of logical theories, whereas soundness is a property of formal logical (or proof) systems. A logical theory—in the use of this term here—is not just a logical system, i.e. a mathematical structure consisting of a formal syntactic apparatus in parallel with a formal semantics. Although they usually include one, logical theories are more than a logical system: they "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

[2] This is Graham Priest's characterization of logic, as it appears in many of his writings, e.g. Priest (2006a, p. 196), Priest (2021, p. 3207), and Priest (2014, p. 212). I am very sympathetic with it, mainly because of its completeness and generality. It is complete because (1) it identifies the logical consequence relation as the main matter of logic ("what follows from what"), (2) it defines logic as a theory, abstracting it from its own practice (i.e. logic is not, in this sense, a discipline), and (3) it makes explicit the explanatory import it has *qua* theory ("and why"). It is general because, while excluding those views of logic that are not centered on logical consequence, it includes different options as to what exactly are the relata of such relation (whether propositions, sentences, assertions, etc.). These are the reasons why I find this characterization suitable to represent the view of logic held by most.

[3] Here, I will not get into the debate between *logical pluralism* and *logical monism*, about whether there is only One True Logic or many. Although it is important, I will keep this issue aside since it does not play an essential role in the issue we are about to discuss.

[4] Given our general discussion targeting both deductive and non-deductive logics, it should be noted that the notion of validity that is meant here is the most general one, covering both deductive validity and inductive strength.

the theory connects to the phenomenon" (Martin, 2021, p. 23).[5] Further, a second and more demanding property for a logical theory that logicians may also aspire to is comprehensiveness, that is the parallel notion to completeness: T_L is comprehensive iff every inference that is genuinely valid is also T_L-valid. Whether the development of a comprehensive logical theory should be considered as one of the essential aims of logic (as a discipline) can be a source of dispute. Instead, I take correctness as a necessary property a logical theory must have to do what it is designed to do—i.e. to account for validity.[6]

As an example, consider classical propositional logic (CPL) and the propositional fragment of the *Logic of Paradox* (LP).[7] As sentential logics, their scope is quite limited, but nonetheless one may wonder which of these two logical theories is more correct than the other. Several inferences turn out to be both LP-valid and CPL-valid, for example Contraposition: $\alpha \rightarrow \beta \vdash_{\text{CPL}} \neg\beta \rightarrow \neg\alpha$ and $\alpha \rightarrow \beta \vdash_{\text{LP}} \neg\beta \rightarrow \neg\alpha$. Suppose we know that this inference is genuinely valid. Then, while this makes both CPL and LP correct with respect to it, Contraposition does not tell us much about their general correctness, and on whether we should prefer one over the other. However, other inferences are not evaluated in the same way by the two theories, and they can help tipping the scale in favor or against them—differences can make a difference! Likely, a paraconsistent logician will point out that LP's non-explosive feature—i.e., $\alpha, \neg\alpha \nvdash_{\text{LP}} \beta$—fits evidence better than CPL. For cognitive agents do hold contradictory beliefs sometimes, but nonetheless they reject triviality, and this may count as evidence that Explosion is genuinely invalid. Therefore, while LP would be correct with respect to it, CPL would not, since $\alpha, \neg\alpha \vdash_{\text{CPL}} \beta$. On the other hand, Disjunctive Syllogism—i.e., $\alpha \vee \beta, \neg\alpha \vdash \beta$-is LP-invalid and CPL-valid, and most logicians take it to be a genuinely valid inference. Therefore, this would make CPL correct and LP incorrect with respect to it. How then should we choose between these two logical theories?

Logical abductivism gives us the following answer: by abductive reasoning, that is Inference to the Best Explanation (IBE).[8] This view about logic is finding favor with an increasing number of philosophers and logicians in recent times—e.g., Williamson (2013, 2017), Priest (2006a, 2014, 2016), Maddy (2014), and Russell (2014, 2015)—, namely the majority of those who endorse the broader epistemological view known

[5] See also Hjortland (2019, §2) and Haack (1976, p. 223) for a clear characterization of logical theories, although in the latter the term used is "formal systems".

[6] This somehow traces the priority of soundness over completeness, as made explicit by e.g. Dummett (1973, p. 290).

[7] Priest (1979, §III).

[8] To be fair, it should be noted that this is not the only way to define logical abductivism. Some—e.g., Woods, in his still unpublished *Logical Abductivism*—prefer a broader characterization along the lines of the following: logical abductivism is the view that we select our logical theories by abductive methods. Then, they not only include IBE among these methods, but also reflective equilibrium and other possible ones. The debate is in its early stages, and it is still not clear whether, for example, reflective equilibrium is abductive in nature, whether it is something other than IBE—something that Woods (2019) tries to disprove—, let alone whether it is the correct means of justification for logical theories. Net of this, the best choice is to proceed by assuming the narrower and more widely accepted definition of logical abductivism given above.

as *anti-exceptionalism about logic*.[9] More precisely, logical abductivism claims that logical theories are justified on the basis of IBE, that is on how well they explain the relevant evidence, so that the correct logical theory turns out to be the one that explains it best. What exactly counts as evidence, apart from some genuinely valid inferences, needs to be better clarified (see Sect. 2), and the state of the art on IBE and its justificatory power are still contentious. Nevertheless, logical abductivism is a rather clear and well-defined view and has recently been one of the most discussed topics in logical epistemology.

There is one issue, however, that could pose a problem for logical abductivism, or at least rings alarm bells: circularity. Such a potential problem is mentioned in Martin (2021, p. 2), and discussed in Douven (2022, Chap. 6) and Priest (2021), among others. To see that, let us assume the usual[10] taxonomy of inferences, according to which there are three exclusive and exhaustive types: deductive, inductive and abductive inferences. Thus, logical theories can be classified accordingly, so that there are theories of deduction, induction and abduction. Now, logical abductivism has been primarily discussed with respect to logical theories of deduction, but arguably the very same view also applies to logical theories of the other domains, by virtue of a kind of uniformity principle: it would be rather ad hoc, and therefore undesirable, if logical theories of different types required a different source of justification. But here the issue arises. While there seems to be nothing wrong in principle in using abduction to determine the correct logical theories of deduction and induction,[11] things might be more complicated regarding logical theories of abduction. For we may ask: is it an epistemically legitimate move to allow for an abductive justification of a theory of abduction? Here, circularity seems to cast its shadow, making the situation darker. This is the issue to which this work is devoted.

The present paper is structured as follows. In Sect. 2 I will briefly review abduction and quickly address a few important points related to it, as well as to logical abductivism. In Sect. 3 I will examine and systematize the way circularity affects arguments

[9] Rooted in Quine (1976), anti-exceptionalism about logic (AEL) claims that "logic does not require its own epistemology, for its methods are continuous with those of science" (Read, 2019, p. 1). Such a characterization is rather vague and can be better specified (see e.g. Martin and Hjortland (2022) who propose an effective systematization of AEL based on the rejection of at least some of the features traditionally ascribed to logic). As a matter of fact, AEL is a collection of epistemic theses about logic, so that it results in a family of different perspectives rather than a unique theory, depending on which of these theses are endorsed. Logical abductivism is one of them, so that it counts as a distinct but related position to AEL. In particular, logical abductivists are anti-exceptionalists, but some versions of AEL can do without logical abductivism. See e.g. Martin and Hjortland (2022) on the mutual relationship between these two positions.

[10] Such a logical tripartition is not straightforward. To mention a few different accounts, Lipton (2017) takes induction to be equivalent to non-deduction, so that abduction turns out to be one particular inductive inference. Instead, Harman (1965) claims that *enumerative induction*, that is routinely classified as an inductive inference, would instead be nothing more than IBE. Finally, bayesians hold that no explanatory considerations are involved in IBE, which makes it an inductive inference as for Lipton (2017), but for a different reason.

[11] This too can be challenged, however. For it can be argued that, even if IBE is necessary in justifying logical theories, it is not sufficient. That is, other inferential deductive or inductive moves are necessary in addition to IBE. Therefore, if this is true, we may doubt that a legitimate inferential justification of logical theories of deduction or induction can be accomplished. Be that as it may, here I take logical abductivism to hold that abduction is both necessary and sufficient for the justification of logical theories, and leave the discussion on whether this is so for future investigation.

depending on their pragmatic function—i.e., in which cases circularity is vicious. Finally, in Sect. 4 I will analyze the circularity involved in the abductive justification of a theory of abduction.

2 Abduction

In this section, I will briefly review abduction and mention some possible theories of this type of reasoning. Note that "[p]recise statements of what abduction amounts to are rare in the literature on abduction" (Douven, 2021, §2), likely because it is a rather controversial form of inference, although ubiquitous in our everyday life, and crucially in scientific practice.[12] However, we can establish some firm points. Abduction—in the way this notion is used in current literature—is equivalent to Inference to the Best Explanation. It is non-deductive, since the necessity-condition of truth preservation from premises to conclusion is dismissed, non-monotonic and ampliative. In this regard, abduction is like induction, but it differs from this in appealing to explanatory considerations.[13] More precisely, IBE is an inference $E_1, ..., E_n \vdash_{\text{IBE}} H_{best}$ whose premises state some evidence (or data) for which the conclusion, H_{best}, is supposed to provide the best explanation. A quite often quoted instance of IBE is the following argument: (E) there are wolf paw-shaped tracks in the snow, therefore (H_{best}) a wolf has recently passed this way. Now, H_{best} is not the only available explanation of E. Other competing hypotheses, call them H_i, can accommodate the data: e.g., (H_1) one of my prankster friends might have walked by on purpose leaving those tracks to fool me, since he knows I am afraid of wolves. However, H_{best} explains E better than all H_i, H_1 included—i.e., H_{best} is the best explanation of E. As such, evidence E justifies the hypothesis H_{best} through IBE.

IBE allows not only a simple hypothesis (i.e., a single proposition), but a whole theory to be inferred as a conclusion. This might appear suspicious, since in general theories are not single propositions, but sets of (even infinitely many) related propositions. However, such worry should not concern us here, and following Priest (2021, p. 3210) we can take such a conclusion "to be of the form 'T is true' (in whatever sense of 'true' is deemed relevant)." For convenience, from now on I will be writing just $E_1, ..., E_n \vdash_{\text{IBE}} T$, where T is the theory that best explains $E_1, ..., E_n$. Thus, we can express logical abductivism as the view that $E_1, ..., E_n \vdash_{\text{IBE}} T_L$, where T_L is the best logical theory. Assuming the above tripartition of inferences, T_L can be a theory of deduction ($T_{ded.}$), induction ($T_{ind.}$), or abduction (T_{IBE}). Again, while there is nothing

[12] In this regard, a relevant and fun fact is that on the same day this sentence was written, the Euclid spacecraft was launched into orbit with the purpose of gathering evidence of dark matter and dark energy. Crucially, their existence has never been empirically proven, but has been postulated by IBE: that is, physicists and astronomers consider the existence of dark matter and dark energy to be the hypothesis that best explains a large number of astronomical observations that have been made so far. Based on such instance of IBE, a huge collective effort has been made and large amounts of funding have been dedicated to this ESA project. This fact alone proves the relevance that IBE has in science.

[13] This is what explanationists claim. On the contrary, bayesians argue that no explanatory consideration plays an inferential role and serves as a guide for drawing conclusions, thus rejecting the legitimacy of IBE and holding that bayesian confirmation theory is all we need.

wrong in principle with $E_1, ..., E_n \vdash_{\text{IBE}} T_{ded.}$ and $E_1, ..., E_n \vdash_{\text{IBE}} T_{ind.}$, we want to further investigate whether $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is legitimate.[14]

Before we move on, there are still three issues that I want to address very briefly as they are of some relevance to our discussion. The first one is about the notion of explanation. A theory of abduction is not, in general, independent of what explanation is taken to be. For depending on how we account for latter, we may get different results as to which the theory that best explains the evidence is. This is hardly deniable, but the details of any specific model of explanation are not important here. This is because we do not have to evaluate any specific T_{IBE}. Rather, due to the level of abstraction of the issue we want to examine, we will talk about different theories of abduction in general, without needing to refer to any of them in particular, and thus without having to presuppose any theory of explanation. Second, one might wonder whether there are actually different theories of abduction from which to choose the best one. In fact, there are. I will not discuss them here, but at least two different theories are outlined in Priest (2021, §5) and developed in Priest (2006a, Chaps. 7, 8).[15] Finally, it is essential for logical abductivism to make clear what exactly the evidence from which to infer the correct logical theory is. This issue has been little explored so far. Among the few works that I know which discuss it, Hjortland (2019), Martin (2021) and his still unpublished *Reflective Equilibrium in Logic* are the most recent ones. Here, we take logical evidence to be whatever serves as a constraint for a logical theory.[16] An analogy with theories from the empirical sciences can help provide clarity. Before anything else, physical and biological theories, for example, must face the "tribunal of sense experience" (Quine, 1982, p. xii): they must be adequate to empirical data, whether measurements or observations in general. Similarly, logical theories must be adequate to 'something', which is exactly what we call logical evidence. But while the notion of empirical evidence is rather clear, what counts as logical evidence seems to be much more elusive. In this regard, we can ask and leave open the following question: what is the tribunal that logical theories must face of? Nonetheless, we do have some good candidates for logical evidence that come directly from the practice of this discipline. As already mentioned, a first type of data that a logical theory should accommodate are some strong intuitions that experienced logicians may have about the validity of particular inferences—i.e., of some specific instances of certain inference schemas.

[14] It should be noted that the premises $E_1, ..., E_n$ are meant to be different for $T_{ded.}$, $T_{ind.}$, and T_{IBE}—i.e., the logical evidence for theories of distinct logical domains is different. For instance, in the case of T_{IBE}, the data include judgments about some good and bad IBEs, which play no role in justifying $T_{ded.}$ and $T_{ind.}$.

[15] According to the first one, the best hypothesis, or theory, is the one with the highest weighted average of the scores it gets for each relevant criterion of evaluation (e.g., simplicity, power, etc.). More precisely, as such, this is not a single theory, but rather a family of many possible theories. For once the set of criteria is established, we can change the specific weights of each of them—representing their relative importance in assessing the goodness of the theory—resulting in a different theory of abduction each time. As for the second one, the best theory is the one with the highest conditional probability given the way it performs on every relevant criterion.

[16] For the sake of completeness, it should be noted that this characterization does not fit with some conceptions of logical evidence that can be traced throughout history. For instance, some logical rationalists considered our intuitions about logical propositions to provide defeasible evidence for logical truths, with no further reference to logical theories (Bealer, 1998; Chudnoff, 2011). Thus, it seems to be a *prima facie* viable move also to accept that logical evidence is evidence for particular logical truths, while denying that it serves as a constraint on logical theories. Thanks to one of the reviewers for pointing this out.

This might be the case, for example, of some widely accepted mathematical proofs that instantiate *reductio ad absurdum*. Based on them, a logician may regard *reductio* as genuinely valid, and try to develop a logical theory T_L such that *reductio* is T_L-valid. But even if experts' intuition can play an important role in justifying logical theories, it is not the only source of logical evidence, and must be considered along with others, so much so that intuition may prove fallible: upon broader evaluation, what it suggests as valid may not turn out to be so. A second type of evidence comes from paradoxes, or more precisely from the way a logical theory manages to handle them. Because of their counter-intuitive, and often contradictory, conclusions, these seemingly sound arguments are usually taken to be problems to be solved. Loosely speaking, paradoxes represent *negative evidence*—with the important exception of dialetheism, for which some paradoxes would count as *positive evidence*—, that is 'phenomena' that a logical theory should make disappear by preventing them from emerging as its own consequences, or more generally should somehow accommodate. Some well-known examples are the logical paradoxes,[17] such as the Liar paradox and its many revenge forms, but also Curry's and the Sorites paradoxes. Further, a third and last example of logical evidence is represented by some features exhibited by natural languages and ordinary reasoning. This is the case, for example, of semantic closure, of the absence of a layered hierarchical structure in natural languages, and of intuitive judgements about the meaningfulness of certain natural-language sentences, that have been taken as evidence in favor of dialetheism, and thus of a logical theory that meets the dialetheic requirements.

To conclude, even if the issue of logical evidence is far from settled, no doubt logicians have relied on some sort of logical evidence in developing, testing and evaluating logical theories, so that what may be controversial is not whether there is logical evidence, but what this evidence is. Thus, the existence of some premises $E_1, \ldots, E_n$ based on which a logical theory can be justified through IBE—as dictated by logical abductivism—can be safely assumed.

3 Circularity and its forms

Circularity is a specific feature of some arguments, and it comes in two distinct forms: *premise-circularity* and *rule-circularity*.[18] Premise-circular, or question-begging, arguments are such that their conclusion also appears (sometimes covertly) as one of their premises. They are considered bad arguments, but in a sense that needs to be clarified. For $P \vdash P$ is deductively valid, so that logical consequence is not guilty of any crime here. In other words, begging the question is not a *logical* fallacy. What is wrong with premise-circular arguments is that they fail "to increase the degree of reasonable confidence which one has in the truth of the conclusion" (Sanford, 1972, p. 198). The rationale here is that often we are required to do more than just arguing validly: "[t]he question then is not the formal validity of the argument [...], but how the argument is supposed to be used in context to fulfill a purpose of discourse" (Walton,

[17] See Priest (2006b, Chaps. 1, 2).

[18] This distinction dates back to Braithwaite (1953).

1994, p. 112). In particular, a crucial function of arguments is the *suasive* (or justificatory, or doubt-reducing) function,[19] which is at play in any process of inferential justification. In a suasive argument, the truth of the conclusion is not yet established, and the purpose is precisely to justify it on the basis of *more certain* premises via reliable inferential moves. As an example, consider a middle school student who has never before wondered whether prime numbers are infinite. When the math teacher asks him what he thinks of that mathematical statement, he will likely answer he has no idea. Then, the teacher presents him with Euclid's proof—i.e. a deductive argument. Presumably, the student will readily accept all the premises, as they are quite intuitive, and after carefully going through each of its steps, become convinced that prime numbers are truly infinite. So, in this context Euclid's proof played a suasive role: it was intended to serve as a means of justification for the mathematical statement at stake. But then, premise-circular arguments cannot do that. For the degree of certainty inherited by the conclusion by means of an argument is at best equal to the degree of certainty of the most dubious premise. So, in case we have the very same sentence both as premise and as conclusion, the argument is not doubt-reducing. In this respect, premise-circular arguments are useless in contributing to the justificatory goal of discourse, and can be labeled as 'pragmatic fallacies'.[20] Premise-circularity is therefore vicious in suasive arguments, in the sense that it makes it impossible to realise the function for which such arguments are intended.

The same may be true for rule-circular arguments, but not always. A rule-circular argument is one that makes use of a rule of inference, call it R, to infer something about R itself.[21] Consider the following example, where MP is for *modus ponens*:

(P_1) MP is a rule of inference that has been employed by many logicians.
(P_2) If many logicians have employed an inference rule, then it is intuitive.
(C_1) Therefore, MP is intuitive.

This argument is an instance of MP where the conclusion tells us something about MP, which makes it a rule-circular, or more precisely an MP-circular, argument. As should be clear, several R-circular arguments are then possible, depending on what their conclusion states about R. For instance, as conclusion we may have "R is an intuitive inference rule", or "R is a primitive rule of system S". In such cases, if the argument is suasive, rule-circularity is not vicious. The reason is that, to fulfill the doubt-reducing function, the argument must only rely on rules of inference that have been already justified, and hence known to be reliable.[22] Provided R is, there is no

[19] This pragmatic characterization of arguments can be found e.g. in Dummett (1973), who distinguishes two kinds of arguments, *suasive* and *explanatory* arguments: "suasive argument is one the purpose of which is to persuade someone of the truth of its conclusion by deriving it from already-accepted premises, while an explanatory argument is one the purpose of which is to explain the already-accepted truth of its conclusion by deriving it from premises the truth of which may not be acknowledged in advance" (Haack, 1982, pp. 218–219).

[20] For a different view on this subject, cf. Bergmann (2004).

[21] Carter (2017, p. 138) further distinguish between *narrow* and *wide* rule-circularity, but such a distinction is not relevant for the present work.

[22] What it means for an inference rule to be reliable is a question as important to elucidate as it is difficult. Reliability is often made to coincide with truth-tracking, but this issue deserves to be discussed separately, which we cannot do here.

problem in using it to justify some of its other features, as long as these do not affect its reliability—i.e., do not imply anything contravening the reliability of R. For, if this were the case, the reliability condition of R might subsequently fall through. Since, arguably, R's being intuitive or primitive has no consequence on its reliability, suasive R-circular arguments for "R is an intuitive inference rule" or "R is a primitive rule of system S" are not vicious. On the contrary, when, for example, the justifiedness of R is what the argument aims to establish, R-circularity becomes vicious: any attempt to justify R which makes use of R fails to do what it is intended to do.[23]

Regarding this last point, however, some philosophers think differently.[24] Among those who belong to this group, Psillos (1999, Chap. 4) is the one who has defended more extensively that R-circular arguments are not vicious even when they aim to justify the reliability of R. His defence of rule-circularity is specifically designed to validate his abductive argument for abduction, but it is very general and can be reconstructed as follows. The problem with this kind of suasive rule-circular arguments is that one has to assume the reliability of the rule invoked in the argument in order to draw it as a conclusion. But then, the proponent of the argument must prove the conclusion before accepting and using that rule, which is something she cannot do unless she first accepts the reliability of the rule. To this objection, Psillos (1999, p. 80) replies "by denying that any assumptions about the reliability of a rule are present, either explicitly or implicitly, when an instance of this rule is used", and that "[n]or should the reliability of the rule be established before one is able to use it in a justifiable way." The reason is that the need for a proof of reliability depends on the epistemological perspective one adopts, and among those available, *externalism* (about justification) is one option. This account "sever the alleged link between being justified in using a reliable rule of inference and knowing, or having reasons to believe, that this rule is reliable" (Psillos, 1999, p. 80). This means that, according to externalism, all that matters is whether the rule *is* reliable, and not whether we have a proof of that. If this condition is met, even in the case we do not know it, the rule-circular argument can serve its purpose and provide a justification for the reliability of the rule.

I find this answer unconvincing. First, there are well-known reasons why one might not want to endorse externalism.[25] But apart from that, suppose it is a fact that the rule at stake is reliable, but we cannot know it. Thus, rule-circularity would not be vicious. But that would be completely irrelevant, since we could not know that either. The crucial point here is that circularity is an epistemic problem, which has to do with whether we can, or cannot, use some inferential moves to change our doxastic state. In other words, it is an 'internal' issue. As such, an externalist move may well be legitimate, but it basically deems rule-circularity as insignificant, or as not an issue at all, which I do not. Therefore, *contra* Psillos (1999), I restore the link between being justified in using a reliable rule of inference and having reasons to believe that this rule is reliable, and my whole discussion will be conducted in this internalist spirit.

[23] For a more extended discussion on this kind of vicious circularity see e.g. Vogel (2008). To mention some relevant philosophers who endorse this view, see Dummett (1973, pp. 295–296) and Priest (2021, p. 3212).

[24] See e.g. Braithwaite (1953), van Cleve (1984), Papineau (1993), and Psillos (1999).

[25] I will resist entering the huge discussion between *externalism* and *internalism* about justification. See Pappas and Littlejohn (2023) for a general overview.

Moreover, in further developing his defence, Psillos (1999) seems to concede some point to the internalist concerns about his externalist strategy, and says that, even if the proponent of a R-circular argument for the reliability of R does not have to assume that R is reliable in advance, it is essential for her not to have any evidence that R is unreliable:

> If one knew that a rule of inference was unreliable, one would be foolish to use it. This does not imply that one should first be able to prove that the rule is reliable before one uses it. All that is required is that one should have no reason to doubt the reliability of the rule—that there is nothing currently available which can make one distrust the rule.
>
> Psillos (1999 p. 83)

According to Psillos, this is exactly the case for IBE: there would be no doubt about its reliability. Now, I take his condition of not using inference rules that are known to be unreliable as entirely acceptable. After all, it would be a mark of irrationality to use e.g. *modus morons* while knowing that it is not reliable. But as should be clear from what has already been said, like the internalists I regard this condition as necessary but not sufficient. In addition to this, I take his belief about the reliability of IBE at least controversial, if not plainly wrong. The extensive debate on abduction has also involved philosophers who have actually questioned the reliability of this mode of inference, for example Van Fraassen (1985, 1989). In fact, this is precisely the reason why we need a justification for it, and why some attempts to provide one have already been made.[26]

The moral is that in suasive arguments, rule-circularity is not always vicious, but can be depending on their conclusion. For later convenience, we can represent R-circular arguments in general as in (1), and the vicious R-circular argument that aims to prove that R is reliable, as in (2)[27]:

(1) $P_1, ..., P_n \vdash_R R$ *is such and such*
(2) $P_1, ..., P_n \vdash_R R$

This concludes our discussion on circularity in suasive arguments. Though, the suasive function is not the only one arguments can have.

In addition to suasive arguments, there are *explanatory* arguments. In this case, the truth of the conclusion is already accepted, so that the purpose is not suasive, but rather explanatory: the argument is intended to provide an explanation of *why* the conclusion is true. Consider the following example. Suppose that I am an astronomer and you are a chemist, and that I show you the light spectrum of GN-z11, a very distant galaxy. You are able to recognize the main absorption lines, since their relative distances are exactly as predicted by chemistry, but you notice that the whole spectrum of GN-z11 is red-shifted. And while acknowledging that it is so, you do not understand why and ask me for an explanation. Thus, I reply:

[26] For example, Boyd (1984) and Douven (2002), among others.

[27] Here, $\vdash_R$ means that the derivation of the conclusion from the premises involves at least one step in which the rule R is employed.

(P_3) Our universe is expanding, which makes all of its objects receding from us, especially those that are very distant like GN-z11.
(P_4) According to the Doppler effect, if a light source is receding from an observer, she will detect its spectrum red-shifted.
(C_2) Therefore, the light spectrum of GN-z11 is red-shifted.

In offering this argument, I have not tried to persuade you that the light spectrum of GN-z11 is red-shifted; you are already convinced of that. Instead, I have provided you with an explanation of the fact noted, that is the conclusion of the argument. Thus, this argument served an explanatory purpose. Moreover, note that you may also not know the premises. Suppose for example that you know P_4, but not P_3. Then, you can take my explanation as a guide to infer P_3 via IBE, that is $P_4, C_2 \vdash_{\text{IBE}} P_3$, and thus convince yourself that P_3 is true. Otherwise, you might know both premises, but before receiving my explanation you did not realise how they could be combined to provide the explanation you were looking for.

For the record, I believe that such a pragmatic classification of arguments into suasive and explanatory is exhaustive.[28] But even if it were not, this would not have much impact on our discussion. Instead, what is important is whether circularity, and especially rule-circularity, is vicious in the case of explanatory arguments, i.e. if it makes it impossible to realise the function for which explanatory arguments are intended. So, the question is: do premise-circularity and rule-circularity conflict with the purpose of explanatory arguments?

As for premise-circularity, we can rephrase the question as follows: is it possible for the same sentence to be both the *explanandum* and part of the *explanans*? This specific issue is not really relevant for us, but I believe that the answer is negative. For I take it as a plausible principle that explanation is a strictly asymmetric process: if x explains y,

[28] The reason is that, as noted by Haack (1982, p. 220), "the distinction between explanatory and suasive arguments is relative to the beliefs of the parties concerned" about the conclusion, and specifically to the doxastic state of the one to whom the argument is intended. Therefore, there are two options: either she believes the conclusion or not. Consequently, either the argument plays an explanatory or a suasive function. However, the matter might not be so simple for at least two reasons. First, beliefs are not just two-state entities (i.e., believe and not believe), but can have a whole range of values depending on how firmly one believes. Thus, arguments may have different functions than just explain and persuade, for example in case of intermediate degrees of belief in the conclusion. This is what Psillos (1999) seems to suggest. With the words of Douven (2022, p. 161): "Psillos makes it clear that the point of philosophical argumentation is not always to convince an opponent of one's position. Sometimes the point is, more modestly, to assure and reassure oneself that the position that one endorses, or is inclined to endorse, is correct." Therefore, in this case arguments would not be attempts to convince the opponent to accept their conclusion. Rather, they could be thought of as justifying the conclusion "from within the perspective of someone who is already sympathetic toward" it. But whether this function is distinct from the suasive one is disputable. For one can take this alleged different use of arguments as an attempt to *better convince oneself*, and therefore still suasive. However, as observed by one of the reviewers, it is hard to deny that this case does differ, in some way, from the attempt to persuade someone through an argument who does not accept its conclusion. Whether the rationale behind this difference justifies the identification of an additional function of arguments remains a possible option. Essentially, this amounts to the second potential difficulty I mentioned: that is, the doxastic state about the conclusion of the one who receives the argument might not be all that matters for a comprehensive pragmatic classification of arguments, and other criteria may be relevant as to what the function of the argument is. But as interesting as this discussion is, I must stop here and leave it for another occasion.

then it is not the case that y explains x, for every x and y.[29] But asymmetric relations are necessarily irreflexive. Hence, it is never the case that x explains itself. More generally, the explanandum cannot be also part of the explanans: i.e., it is not allowed to explain P by resorting, among other sentences, to P itself. Thus, premise-circular explanatory arguments turn out to be always vicious. An extensive investigation of the concept of explanation would be required for a better and more conclusive answer, but since this is not a significant case for our issue, we can leave it aside and move on.

On the contrary, the explanatory function of arguments is more welcoming towards rule-circularity. To see that, consider this example. Suppose we both know that most people find MP intuitive, but I wonder why it is so and ask you for an explanation. Then, you offer me the following argument:

(P_5) An inference rule becomes intuitive for someone if they use it often and successfully.
(P_6) Statistical studies have shown that most people use MP very often and successfully.
(C_3) Therefore, most people find MP intuitive.

Arguably, this argument counts as a good explanation for C_3 (perhaps not particularly insightful, but still successful), and it is just an instance of MP. Therefore, $P_5, P_6 \vdash C_3$ is an example of a non-viciously rule-circular explanatory argument. That is, at least some instances of $P_1, ..., P_n \vdash_R R\ is\ such\ and\ such$ that are explanatory are not vicious. What might appear more contentious, instead, is whether R-circular explanatory arguments for the reliability of R are vicious. But, again, the answer is negative. Consider the following example:

(P_7) If an inference rule has no counterexamples, then it is valid.
(P_8) MP has no counterexamples.
(C_4) Therefore, MP is valid.

Provided the one to whom this argument is addressed agrees (based on a different justification) that MP is valid, $P_7, P_8 \vdash C_4$ fulfills its explanatory function entirely: if we know that MP is valid, there is no problem in explaining why this is so by using MP. Thus, in general: for any explanatory instance of $P_1, ..., P_n \vdash_R R\ is\ such\ and\ such$, using R to explain why $R\ is\ such\ and\ such$ is legitimate, since the explanation process can be carried out untouched by rule-circularity. To put it in an analogy, using R to explain why and how R works is "no more suspect than using observation to study the structure and function of the eye" (Lipton, 2017, p. 12), once we are persuaded

[29] The asymmetry of explanation is not only plausible, it is quite a standard principle in the literature on explanation. It is widely accepted that most kinds of explanation are asymmetric, e.g. causal explanation. For a recent paper on this topic see e.g. Khalifa et al. (2021). Nonetheless, as one of the reviewers pointed out, throughout the history of philosophy, some self-explanatory phenomena have been proposed—e.g., existence of God or some irreducible metaphysical facts, like *autonomous* facts in Dasgupta (2016)—, which would represent a violation of Irreflexivity, and with it, Asymmetry. An in-depth analysis of such cases goes beyond the scope of this paper, but arguably the kinds of explanans/explanandum that are relevant to the issue we are addressing (i.e., logical evidence and theories) do not fall within the domain of (alleged) self-explanatory phenomena. In other words, even if the explanation were not irreflexive in general, it is a plausible assumption that it is when restricted to the class of explanans/explananda that we are dealing with here.

Table 1 Viciousness of circular arguments depending on their pragmatic function

Arguments	Premise-circular	Rule-circular with conclusion affecting R's reliability	with conclusion not affecting R's reliability
Suasive	**vicious**	**vicious**	not vicious
Explanatory	**vicious**	not vicious	not vicious

Here, "vicious/not vicious" have to be read as "*always* vicious/not vicious"

that eyesight is reliable. Therefore, rule-circularity is not vicious in explanatory arguments, and this results in an interesting asymmetry between suasive and explanatory arguments with respect to it.

However, a new problem arises in this regard. Haack (1982, §I, b–c) shows that even if rule-circular explanatory arguments are not viciously circular, they can be defective in a different and important way. For "if the validity of *modus ponens* could be explained by means of an argument which used *modus ponens*, then, presumably, the validity of *modus morons* [...] could be explained by means of an argument which used *modus morons*" (Haack, 1982, p. 221). That is, both reliable and unreliable inference rules may be used to explain their own reliability. In this sense, rule-circular explanatory arguments are "indiscriminating" (ibid.), and until a justification is provided for the inference rules involved in them, we lack sufficient guidance as to which explanation we should consider correct. This is why Haack says that "explanation presupposes justification" (ibid.). We will return to this point in the next section, but for now this concludes our systematic analysis of circularity and whether or not it is vicious depending on the pragmatic purposes of the arguments. We can then summarise the main results of this section as in Table 1.

3.1 The case of soundness proofs

An important example of argument whose function has been discussed in the literature is the standard soundness proof for a deductive logical system, say S. In this short section, I briefly discuss this case since we will need it to address an important analogy in Sect. 4.

Presumably, in such a metalogical argument one (or more) of the inference rules of S, say R, is used to prove that S is sound. But for S to be sound means that all of the inferences provable in S are valid—according to the agreed semantics—, R included. Therefore, R is employed to prove its own validity, which makes the soundness proof R-circular.

Some philosophers—e.g., Dummett (1973) and Priest (2021)—judge this circularity as not vicious, essentially because soundness arguments would have an explanatory function, and not a suasive one. According to this interpretation, an R-circular soundness proof does not have any persuasive purpose about the validity of R. Rather, it counts as an explanation of why R is valid, provided we agree in advance that it is so. As such, it would not be viciously circular.

As long as we keep in mind the above warning of Haack (1976), I completely agree. I only find that the issue is slightly more complicated in the following way: a soundness argument is suasive, but allows an 'explanatory reading' with respect to the inferences of S that it employs. Let me try to provide some clarity. I believe that proofs of soundness are suasive in nature. For they aim to establish a result that is not known prior to them. We may conjecture that S is sound, but cannot know it before we prove it. Therefore, an argument for the soundness of S seems to provide a justification—i.e., to play a suasive function—for the validity of all the inferences that are provable in S. However, this may be true with respect to some of the inferences of S, but not to all of them, and precisely not to those that are used in the soundness proof. As for them, the function of the proof is not suasive, but explanatory, since we already accepted their validity. Then, in light of the systematization we have developed, we can classify soundness proofs as arguments with form $P_1, ..., P_n \vdash_R R$ *is such and such* that are explanatory at least with respect to R—even if suasive with respect to other inferences of S. Therefore, under this explanatory reading, metalogical arguments for soundness are rule-circular but not vicious.[30]

4 Logical abductivism on abductive logic

Let us come, then, to our problem: is there anything epistemically wrong with an abductive justification of a theory of abduction?

A first important clarification step is made by Priest (2021, §6). The issue we are facing might seem analogous to that which arises when trying to justify induction by induction itself. But this is not quite the case. For it is one thing to use IBE to justify IBE, quite another to use IBE to justify a theory of IBE, T_{IBE}. As an example of the first case, consider the following attempt of justification: evidence that scientific theories are successful is best explained by the hypothesis that IBE is a reliable form of inference, which therefore turns out to be likely true by virtue of IBE itself. Following Sect. 3, we can represent it as $E_1, ..., E_n \vdash_{\text{IBE}}$ IBE, agree on its suasive function, and conclude that it is viciously circular on the basis of the above.[31] But this is not the issue we are interested in here. Instead, we want to know whether the circularity involved in an abductive justification of a theory of abduction—to which logical abductivism is committed—is vicious.

To begin with, note that $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is an instance of $P_1, ..., P_n \vdash_R R$ *is such and such*, where $E_1, ..., E_n$ are $P_1, ..., P_n$, IBE is R, and the conclusion T_{IBE} can be rephrased as "T_{IBE} is the correct theory of IBE", which instantiates "R *is such and such*". Therefore, without much surprise, $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is certainly rule-circular, or better IBE-circular. The question, then, is whether this IBE-circularity is vicious or not, which in turn depends on whether the function of the

[30] Of course, the situation is reversed in case the proof of soundness is intended to convince someone about the validity of all the rules of the system, including those used in the proof. This would result in the suasive vicious kind of argument described above.

[31] On the problem of the justification of IBE see e.g. Boyd (1984), Enoch and Schechter (2008), and Carter and Pritchard (2017).

argument is explanatory or suasive, and in the latter case, on what the impact that the conclusion of the argument has on the reliability of IBE is.

Priest (2021) argues that $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is not vicious:

> A deductive explanatory argument for deductive inferences—as found, for example, in standard soundness arguments in metalogic—does not beg the question; neither does an inductive explanatory argument of induction. What we are dealing with concerning abduction is finding the best theory which explains how and why it works. In the same way, then, using abduction to determine what follows from what abductively begs no questions.
>
> Priest (2021, pp. 3212–3213)

The way I understand this passage is that $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ allows an explanatory reading in the same way as soundness arguments, and therefore it is not vicious. More precisely, $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is for all intents and purposes a suasive argument: it aims to establish which, among some competing theories of abduction, is the correct one, which is something we can not know, according to logical abductivism, prior to this IBE. After all, there is a very general and interesting reason why this is not an explanatory argument: IBEs can never be used in the explanatory way. This should be clear if we recall what an IBE is: an inference in which we infer the hypothesis, or theory, that best explains the evidence in the premises. Therefore, while the logical direction goes from the premises to the conclusion, the explanatory direction goes from the conclusion to the premises: that is, the conclusion of an IBE is the explanans, whereas the premises represent the explanandum. However, in the explanatory arguments the explanatory 'dynamics' is exactly the opposite. In them, it is the conclusion that requires the explanation that is provided by the premises—i.e., the explanatory direction goes from the premises to the conclusion. So, if IBEs could be used explanatorily, it would result that both the premises explain the conclusion and vice versa. But, as already mentioned, explanation is asymmetric. Therefore, IBEs are unable to perform the explanatory function as it is described in Sect. 3, which further supports that $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is a suasive argument. However, it is not vicious. The reason for this is that, in justifying T_{IBE}, IBE does not play any suasive role with respect to IBE itself, but only an explanatory one. For T_{IBE} is not intended to provide any justification for IBE, but only an explanation. T_{IBE} would only explain why and how IBE is reliable, since we already accept, in this view, that IBE is so.

Now, I fully agree with that, but what I contend is that such an explanatory reading puts an end to the matter. More specifically, I argue that this explanatory reading of $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is incomplete by virtue of what Haack (1982) has pointed out about Dummett's explanatory interpretation of soundness proofs for deductive systems: explanation presupposes justification.

Consider the following example given by Douven (2022):

> Suppose that some scientific community relied not on abduction but on a rule that we may dub "Inference to the Worst Explanation" (IWE), a rule that sanctions inferring to the *worst* explanation of the available data. We may safely assume that the use of this rule would lead mostly to the adoption of very unsuccessful

> theories. Nevertheless, the said community might justify its use of IWE by dint of the following reasoning: "Scientific theories tend to be hugely unsuccessful. These theories were arrived at by application of IWE. That IWE is a reliable rule of inference—that is, it usually leads from true premises to true conclusions—is surely the worst explanation of the fact that our theories are so unsuccessful. Hence, by application of IWE, we may conclude that IWE is a reliable rule of inference."
>
> Douven (2022, p. 160)

Like IBEs, IWEs also cannot serve an explanatory function, and for the same reason discussed above: explanation is asymmetric. Therefore, we cannot use the argument proposed by the community imagined by Douven as being explanatory. Nonetheless, we can take a cue from his example to devise another one that may serve our case.

Suppose that some logicians from the same community are logical abductivists, but in their own way: they believe that logical theories are justified on the basis of IWE, that is on how bad they explain the relevant evidence, so that the correct logical theory turns out to be the one that explains it worst. Then, they use IWE to select the correct (i.e., worst) logical theory of IWE, that is: $E_1, ..., E_n \vdash_{\text{IWE}} T_{\text{IWE}}$. This argument is (i) rule-circular, (ii) suasive and (iii) not vicious, just as $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is, and for the very same reasons: (i) it is rule-circular because it employs IWE to infer something about IWE itself, (ii) it is suasive because it is intended to justify T_{IWE}—besides the fact that it cannot be explanatory—, and (iii) it is not vicious because, in justifying T_{IWE}, IWE does not play any suasive function with respect to IWE itself, but only an explanatory one, since the whole community, logicians included, already believe that IWE is reliable. Thus, this suggests that logical abductivists who rely on IWE and those who rely on IBE are not in significantly different situations from an epistemic perspective. In other words, in the absence of a justification for IBE, we are not justified in assuming that logical abductivism based on IBE is significantly better than the way the logicians from Douven's community justify their logical theories: if $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ is legitimate, so it seems to be $E_1, ..., E_n \vdash_{\text{IWE}} T_{\text{IWE}}$.

Following Psillos (1999), one can reply that there are reasons to disbelieve the reliability of IWE, since it "would lead mostly to the adoption of very *unsuccessful* theories" (my italics). Therefore, $E_1, ..., E_n \vdash_{\text{IWE}} T_{\text{IWE}}$ turns out to be defective, as it does not meet the condition of employing only inference rules that we have not reasons to disbelieve. But then, as previously mentioned, there are reasons to doubt also IBE. Plausibly less, but still there—e.g., some false theories have often been justified by means of IBE. Thus, what this situation calls for is a justification for IBE, which we do not yet have, in order for logical abductivism to be a plausible (i.e., non-defective and effective) view of the justification of logical theories.

4.1 Changing epistemological perspective: the case of knowledge-first epistemology

As we clarified in Sect. 3, the discussion conducted and the conclusions it led to adhere to an internalist approach to justification, which, clearly, is not the only available epistemological perspective and is open to challenge. Among the alternatives, one epistemology that may be particularly relevant to consider in relation to the central issue of this paper is the *knowledge-first epistemology* proposed by Williamson (2000).[32] In this section, we carry out a preliminary investigation of how the problem of the lack of justification for IBE, and thus of logical abductivism on abductive logic, might be framed within this view. This analysis is not intended to be exhaustive and is instead positioned as an initial step towards a more in-depth future examination that would extend the present work. We will begin by describing the most salient aspects of this alternative epistemological perspective.

Knowledge-first epistemology overturns the common assumption that belief is conceptually prior to knowledge. According to Williamson (2000, §1.3), there is no possible conceptual analysis of knowledge, that is, knowledge cannot be reduced to more basic notions. One might think that, since for p to be known, it is necessary that p be true and believed, such an analysis must then be possible. But acknowledging some necessary feature for a concept—e.g., being true and believed—does not contradict the impossibility of its analysis. Indeed, the general warning is that "the existence of conceptual connections is a bad reason to postulate an analysis of a concept to explain them" (Williamson, 2000, p. 33). Thus, knowledge-first epistemology takes knowledge—to be understood as propositional knowledge—to be the "unexplained explainer" (Williamson, 2000, p. 10), that is the starting point to elucidate the other epistemic concepts, such as belief, justification, and evidence, and therefore the conceptually prior notion on which to develop epistemology. Nonetheless, a modest positive characterization of knowledge can be given, even though one that is not an analysis of it in the traditional sense. Williamson (2000, Chap. 1) takes knowing to be the most general[33] factive[34] stative[35] propositional attitude, and therefore, a mental state of a subject. Consequently, the knowledge of a subject S is just the body of propositions which S knows—i.e., the collection of true propositions that are the contents of S's attitudes of knowing. Believing, on the other hand, is a different (non-factive) propositional attitude, which is nevertheless related to knowing in the following way: if S knows that p then S believes that p.

[32] The relevance of knowledge-first epistemology to the content of this paper was appropriately suggested by one of the reviewers.

[33] In a few words, "the most general" means that the factive stative propositional attitude of knowing applies if any factive stative propositional attitude applies at all.

[34] Here "factive" means that the propositional attitude of the subject S towards p is not to be understood merely as a relation between S and p, but it also depends on whether p is true. Therefore, the attitude of knowing has also a non-mental component—i.e., truth. More clearly: "[a] propositional attitude is factive if and only if, necessarily, one has it only to truths" (Williamson, 2000, p. 34).

[35] According to Williamson (2000, §1.4), not all factive attitudes are mental states (e.g., forgetting, which is a process, and therefore not a state). He then refers to those factive attitudes that constitute a mental state as stative.

A further key feature of knowing is that it is not *luminous*. This means that it is not the case that whenever S knows (or fails to know) that p, S is in a position to know that she herself knows (or fails to know) that p.[36] To put it in different words, knowing can be often a *transparent* attitude, meaning that we can have cognitive access to it, but not always: "[o]ne can know something without being in a position to know that one knows it" (Williamson, 2000, p. 114)—and the schema can be iterated indefinitely to higher-order knowledge (e.g., to know that one knows that one himself knows something). For ease of discussion, let's agree to say that S knows (or fails to know) that p *opaquely*—i.e., non-transparently—whenever S knows (or fails to know) that p without being in a position to know that she knows (or fails to know) that p. A straightforward and not surprising consequence of opaque knowledge, but one worth emphasizing, is that it cannot be identified. For suppose by *reductio* that S can identify her opaque piece of knowledge. Therefore, there is something, call it p, such that (1) S knows that p, (2) S cannot know that she knows that p, and (3) S can identify that it is precisely p that she knows and that she cannot know that she knows. But (3) entails that S can know that she knows that p, which contradicts (2).

Moving forward, knowledge-first epistemology equates evidence with knowledge: "one's total evidence is simply one's total knowledge" (Williamson, 2000, p. 9). This allows for the understanding of justification in terms of knowledge, the latter being "what justifies" instead of "what gets justified" (ibid.). More precisely, knowledge becomes the only source of justification for beliefs: "knowledge, and only knowledge, justifies belief" (Williamson, 2000, p. 185). A belief would be justified just in case it receives support from—to be understood as being subject to an increase in its probability by—the evidence, i.e., knowledge. As Williamson (2000, p. 9) makes clear: "justification is primarily a status which knowledge can confer on beliefs [...] without themselves amounting to knowledge. Knowledge itself enjoys the status of justification only as a limiting case [...]." Moreover, within this perspective, any piece of knowledge, such as the proposition p, can be questioned just by removing p from one's existing knowledge base and assessing it relative to the remaining knowledge—that is, one's independent evidence. Therefore, "one can lose as well as gain knowledge" (Williamson, 2000, p. 10).

Clearly, knowledge-first epistemology is not exhausted by the brief description we have just provided. Nonetheless, the features we have outlined should be sufficient to begin sketching an answer to the following question: what conclusions can be drawn regarding how this alternative epistemology might frame the issue of logical abductivism that we have pointed out in this paper? What we have argued for in Sect. 4 is the need to have a justification for IBE in order for logical abductivism to actually accomplish what it aspires to do. However, for a knowledge-first epistemologist there might be no need for that. To see why, take p to be "IBE is reliable". Arguably, for logical abductivism on abductive logic to be interpreted as a non-defective (i.e. discriminating) explanatory argument we need one of the following two options: either we know that p or we justifiably believe it—i.e., either p is knowledge or it is a justified belief. For in both cases the explanatory reading of $E_1, \ldots, E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ would turn out

[36] This essentially amounts to denying the epistemic principle most commonly known as the KK-principle: $Kp \rightarrow KKp$.

to be discriminating, since if p is knowledge, then is true, and if p is a justified belief, then is at least likely true. The crucial point is that for knowledge-first epistemology, both these options seem to be available. Let's proceed by examining the two cases separately.

Suppose that, for some subject S, p is knowledge. Then, we have two further options: either S knows that p transparently or opaquely. Now, recall that knowing is a mental state and that arguably we do not have access to mental states other than our own. That means that we cannot rule out that there actually is some subject who knows that p transparently. However, based on the extensive debate on the topic, I would be inclined to consider this scenario as quite marginal, and therefore to dismiss it. For even the most staunch defenders of the reliability of IBE would probably not be willing to claim they know that p and know that they know that p.[37] On the contrary, I find the second sub-case to be more plausible: it is possible that some subject S knows that IBE is reliable without being in a position to know that she knows that. This just means that p is an instance of opaque knowledge, whose existence is guaranteed by knowledge-first epistemology. Even though this scenario does not correspond to the kind of attitude I personally experience with respect to p—as will be clarified shortly—I find it relevant[38] in this context and therefore not to be rejected.

Now, to the second case. Is it possible that, for some subject S, S justifiably believes that p? As previously observed, we do not have a satisfactory justification for IBE, which would lead us to answer the question negatively. For the lack of a justification for IBE excludes that there is some subject S who knows that p is a justified belief. That is, it excludes that p is a transparently justified belief for S. If we were working within an epistemology in which knowing is luminous, the matter would end here. However, this is not true for knowledge-first epistemology. For since knowledge justifies beliefs and lacks luminosity, it follows that we are not always in a position to know what our evidence is, and consequently that we may have justified beliefs without being in a position to know that such beliefs are justified. In other words, knowledge-first epistemology allows for opaquely justified beliefs: that our beliefs are justified can sometimes be an opaque matter for us. And this might be precisely the case for p. After all, this position seems very close to—if not exactly the same of—the externalist stance of Psillos (1999) that we discussed in Sect. 3.

To sum up, knowledge-first epistemology seems to save the day in the following sense. Since it is possible that we know that p opaquely, or that our believing in p is opaquely justified, it is possible for p to be true, or at least likely true, and therefore for the explanatory interpretation of $E_1, \ldots, E_n \vdash_{\text{IBE}} T_{\text{IBE}}$ to be discriminating, although unbeknownst to us. Moreover, the lack of a justification for p would no longer constitute a problem, since the pursuit of such a justification would be an impossible[39] and therefore futile endeavor.

An attempt to challenge the conclusions of this line of thought can be made in the following way. I will discuss it very briefly, leaving a more detailed analysis for another

[37] However, if I am mistaken and this situation is precisely where some advocates of IBE find themselves, this means there is an additional way in which this first case occurs, that is exactly what I am arguing for.

[38] What I mean by "relevant" here is "representative of the situation in which a significant portion of the experts on the topic may find themselves."

[39] Since this is an opaque scenario where we are not in a position to know.

occasion. Consider not-p, i.e., "IBE is not reliable". For reasons of symmetry, if it is possible that we know that p opaquely, or that our believing that p is opaquely justified, then it is equally possible that we know that not-p opaquely, or that our believing that not-p is opaquely justified. Of course, p and not-p are contradictories, so that it is not possible to know or believe both. In other words, knowing that p and knowing that not-p are also mutually exclusive, and the same is for believing. Therefore, it will also not be possible for both p and not-p to be opaquely known, and for both p and not-p to be opaquely justified beliefs. But then, which of these competing possibilities should we accept? Recall that, as emphasized earlier, opaque knowledge cannot be identified, and the same goes for opaquely justified belief. Thus, this lead us to conclude that the question just posed is simply an insoluble dilemma—which is perhaps a less palatable conclusion than we might have hoped.

To conclude, one last observation. Note that knowing or justifiably believing either that p or that not-p do not exhaust the epistemic space. We may instead fail to know either that p or that not-p, as well as fail to justifiably believe either that p or that not-p, both transparently and opaquely. Moreover, it is precisely within this range of possibilities that I think the most plausible ones are found. For instance, if I were to describe my own epistemic state regarding p, I would say that I fail to know that p transparently: I fail to know that IBE is reliable and I know that I fail to know that IBE is reliable. Again, knowing something and failing to know it are incompatible: for any subject, it is not possible to do both. Therefore, we should consider these cases in relation to those previously examined and draw the appropriate conclusions. But that is material for another story.

4.2 The 'real problem of circularity' of logical abductivism on abductive logic

This work was mainly motivated by Priest (2021). Here, the problem of circularity of logical abductivism when applied to non-deductive inferences is also discussed, but it goes beyond what is done in the present paper. Let me provide some explanation to clarify how these two works relate each other.

After addressing some preliminary issues and introducing the matter at hand, Priest (2021) argues that the kind of circularity involved in an abductive justification for T_{IBE} is not vicious (§6). This is essentially what we have done, only through a more in-depth analysis, which has allowed us to highlight an additional important fact: logical abductivism requires in advance a justification of abduction to be effective. But while we stop our investigation here, Priest (2021) goes further and shows that, even if circularity is not a problem from a pragmatic point of view, it could nonetheless lead to a different and very intriguing problematic situation.

Most logicians agree that logic (T_L) is normative. Arguably, this is equally true for $T_{ded.}$, $T_{ind.}$ and T_{IBE}. Thus, we should submit to the authority of T_{IBE} and follow its dictates: i.e., we should reason abductively in the way T_{IBE} prescribes. But for a logical abductivist, this means that she has to comply with T_{IBE} also in the justification of the correct logical theory of abduction, i.e. $E_1, ..., E_n \vdash_{\text{IBE}} T_{\text{IBE}}$. But what if this results in a different theory of abduction, say T^*_{IBE}? More clearly, what if T^*_{IBE} turns out to be better (and therefore more likely true) than T_{IBE} when running IBE according

to T_{IBE}? Well, one might take this as a good reason to consider T^*_{IBE} actually better than T_{IBE}, and therefore to consider this theory, i.e. T^*_{IBE}, as the correct one. But then, what if T_{IBE} turns out to be better than T^*_{IBE} when applying T^*_{IBE}? In this case, "[w]e will be forced into a game of rational ping pong" (Priest, 2021, §7), which is clearly a dead-end situation that does not seem to be permissible. This is what Priest (2021) calls 'the real problem of circularity' in the case of logical abductivism when applied to abductive logic.

While bringing out such an interesting problem, as well as helping to clarify and examine it, Priest (2021) does not offer a solution. And certainly, the present paper is not the place to try to address it. I just wanted to mention it to show that the investigation on circularity in the specific case of logical abductivism applied to abductive logic is not yet exhausted, and deserves further efforts in the future.

5 Conclusion

This paper presents three results, one main and two auxiliary. The main result is that logical abductivism on abductive logic-i.e., $E_1, \ldots, E_n \vdash_{\text{IBE}} T_{\text{IBE}}$-is rule-circular but not vicious. However, it is still in need of a justification for IBE to be a plausible epistemic view. To support this conclusion, I systematically analyzed circularity, showing in which pragmatic situations it is vicious, and in which it is not. The results of this analysis are shown in Table 1. Finally, based on the asymmetry of explanation, I argued that IBEs can serve a suasive function, but not an explanatory one.

As we have mentioned, however, the need to find a justification for IBE seems to depend significantly on the general epistemological perspective we adopt. As we began to explore in Sect. 3, and especially in Sect. 4.1, a change in such a perspective can have several consequences, including a significant impact on the legitimacy of logical abductivism. This should not be too surprising. For logical abductivism is itself an epistemological view, albeit one limited to the justification of logical theories, so that it connects to other epistemic notions and depends on the way they are conceived. *Prima facie*, this might be taken as evidence supporting a top-down approach to the epistemology of logic. That is: only after we have clarified which general epistemological theory is correct, or at least most adequate—and this includes understanding fundamental epistemic notions such as knowledge, belief, and justification—can we proceed with the investigation into the epistemology of logic. But whether this is really the case requires a separate investigation.

Acknowledgements I would like to extend my gratitude to Elke Brendel, as the supervisor of my research project, for her insightful comments on earlier versions of this work, and for her guidance. I also express my appreciation to Gerhard Schurz and his research team for inviting me to present and discuss the main contents of this paper at the University of Düsseldorf. Furthermore, I wish to thank the audiences of the Inductive Metaphysics Conference (August 9–11, 2023, University of Düsseldorf), the BoBoPa (Bologna-Bonn-Padova) seminars, the Philosophy of the Formal Sciences Workshop (December 4, 2023, University of Padova), and the European Network for the Philosophy of Logic. Special thanks are due to Massimiliano Carrara, Sebastian Speitel, Stefano Pugnaghi, Filippo Ferrari, Ben Martin, Andrea Strollo, Sebastiano Moruzzi and Giuseppe Spolaore for their expertise and valuable insights. I would also like to thank two anonymous referees for their extremely useful comments on the previous versions of this paper. This work was supported by the DFG (Deutsche Forschungsgemeinschaft), research unit FOR 2495.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The author has no conflict of interest to disclose.

References

Bealer, G. (1998). Intuition and the autonomy of philosophy. In M. DePaul & W. Ramsey (Eds.), *Rethinking intuition: The psychology of intuition and its role in philosophical inquiry* (pp. 201–240). Rowman & Littlefield.

Bergmann, M. (2004). Epistemic circularity: Malignant and benign. *Philosophy and Phenomenological Research, 69*(3), 709–727.

Boyd, R. (1984). The current status of scientific realism. In J. Leplin (Ed.), *Scientific realism* (pp. 41–82). University of California Press.

Braithwaite, R. (1953). *Scientific explanation*. Cambridge University Press.

Carter, J. A., & Pritchard, D. (2017). Inference to the best explanation and epistemic circularity. In K. McCain & T. Poston (Eds.), *New essays on inference to the best explanation: Best explanations* (pp. 133–149). Oxford University Press.

Chudnoff, E. (2011). What intuitions are like. *Philosophy and Phenomenological Research, 82*(3), 625–654. https://doi.org/10.1111/j.1933-1592.2010.00463.x

Dasgupta, S. (2016). Metaphysical rationalism. *Noûs, 50*(2), 379–418.

Douven, I. (2002). Testing inference to the best explanation. *Synthese, 130*, 355–377.

Douven, I. (2021). Abduction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.). Metaphysics Research Lab, Stanford University.

Douven, I. (2022). *The art of abduction*. MIT Press.

Dummett, M. (1973). The justification of deduction. In M. Dummett (Ed.), *Truth and other enigmas.* Oxford University Press.

Enoch, D., & Schechter, J. (2008). How are basic belief-forming methods justified? *Philosophy and Phenomenological Research, 76*(3), 547–579.

Ferrari, F., & Moruzzi, S. (2020). Logical pluralism, indeterminacy and the normativity of logic. *Inquiry: An Interdisciplinary Journal of Philosophy, 63*(3–4), 323–346. https://doi.org/10.1080/0020174x.2017.1393198

Ferrari, F., & Orlandelli, E. (2019). Proof-theoretic pluralism. *Synthese, 198*(Suppl 20), 4879–4903. https://doi.org/10.1007/s11229-019-02217-6

Field, H. (2015). What is logical validity. In C. R. Caret & O. T. Hjortland (Eds.), *Foundations of logical consequence.* Oxford University Press.

French, R. (2021). A dialogical route to logical pluralism. *Synthese, 198*(Suppl 20), 4969–4989.

Haack, S. (1976). The justification of deduction. *Mind, 85*(337), 112–119. https://doi.org/10.1093/mind/lxxxv.337.112

Haack, S. (1982). Dummett's justification of deduction. *Mind, 91*(362), 216–239.

Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review, 74*(1), 88–95.

Hjortland, O.T. (2019a). What counts as evidence for a logical theory? *Australasian Journal of Logic*, *16*(7), 250–282. https://doi.org/10.26686/ajl.v16i7.5912

Hjortland, O. T. (2019b). What counts as evidence for a logical theory? *The Australasian Journal of Logic, 16*(7), 250–282.

Khalifa, K., Millson, J., & Risjord, M. (2021). Inference, explanation, and asymmetry. *Synthese, 198*, 929–953.
Lipton, P. (2017). Inference to the best explanation. In *A companion to the philosophy of science* (pp. 184–193). Blackwell.
Maddy, P. (2014). *The logical must: Wittgenstein on logic*. Oxford University Press.
Martin, B. (2021). Identifying logical evidence. *Synthese, 198*(10), 9069–9095.
Martin, B., & Hjortland, O. T. (2022). Anti-exceptionalism about logic as tradition rejection. *Synthese, 200*(2), 1–33.
Oza, M. (2020). The value of thinking and the normativity of logic. *Philosophers' Imprint, 20*(25), 1–23.
Papineau, D. (1993). *Philosophical naturalism*. Blackwell.
Pappas, G., & Littlejohn, C. (2023). Internalist vs. externalist conceptions of epistemic justification. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Metaphysics Research Lab, Stanford University.
Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic, 8*(1), 219–241.
Priest, G. (2006a). *Doubt truth to be a liar*. Oxford University Press.
Priest, G. (2006b). *In contradiction*. Oxford University Press.
Priest, G. (2014). Revising logic. In P. Rush (Ed.), *The metaphysics of logic* (pp. 211–223). Cambridge University Press.
Priest, G. (2016). Logical disputes and the a priori. *Logique et Analyse, 236*, 347–366.
Priest, G. (2021). Logical abductivism and non-deductive inference. *Synthese, 199*(1), 3207–3217.
Psillos, S. (1999). *Scientific realism: How science tracks truth*. Routledge.
Quine, W. v. O. (1976). Two dogmas of empiricism. In S. Harding (Ed.), *Can theories be refuted?: Essays on the Duhem-Quine thesis* (pp. 41–64). D. Reidel Publishing Company.
Quine, W. V. O. (1982). *Methods of logic*. Harvard University Press.
Read, S. (2019). Anti-exceptionalism about logic. *Australasian Journal of Logic, 16*(7), 298. https://doi.org/10.26686/ajl.v16i7.5926
Russell, G. (2015). The justification of the basic laws of logic. *Journal of Philosophical Logic, 44*(6), 793–803.
Russell, G. (2020). Logic isn't normative. *Inquiry, 63*(3–4), 371–388.
Russell, G. K. (2014). Metaphysical analyticity and the epistemology of logic. *Philosophical Studies, 171*(1), 161–175.
Sanford, D. H. (1972). Begging the question. *Analysis, 32*(6), 197–199.
Stei, E. (2020). Disagreement about logic from a pluralist perspective. *Philosophical Studies, 177*(11), 3329–3350.
Steinberger, F. (2019). Logical pluralism and logical normativity. *Philosophers' Imprint, 19*(12), 1–19.
van Cleve, J. (1984). Reliability, justification, and the problem of induction. *Midwest Studies in Philosophy, 9*(1), 555–567.
Van Fraassen, B. C. (1985). Empiricism in the philosophy of science. In P. Churchland & C. Hooker (Eds.), *Images of science* (pp. 245–308). University of Chicago Press.
Van Fraassen, B. C. (1989). *Laws and symmetry*. Clarendon Press.
Vogel, J. (2008). Epistemic bootstrapping. *The Journal of Philosophy, 105*(9), 518–539.
Walton, D. N. (1994). Begging the question as a pragmatic fallacy. *Synthese, 100*(1), 95–131.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.
Williamson, T. (2013). *Modal logic as metaphysics*. Oxford University Press.
Williamson, T. (2017). Semantic paradoxes and abductive methodology. In B. P. Armour-Garb (Ed.), *Reflections on the liar* (pp. 325–346). Oxford University Press.
Woods, J. (2019). Against reflective equilibrium for logical theorizing. *The Australasian Journal of Logic, 16*(7), 319–341.