Synthese (2021) 198:7923–7947
https://doi.org/10.1007/s11229-020-02556-9

THEMES FROM ELGIN

Reflective equilibrium and understanding

Christoph Baumberger[1] · **Georg Brun**[2]

Received: 30 October 2018 / Accepted: 25 January 2020 / Published online: 3 February 2020

Abstract
Elgin has presented an extensive defence of reflective equilibrium embedded in an epistemology which focuses on objectual understanding rather than ordinary propositional knowledge. This paper has two goals: to suggest an account of reflective equilibrium which is sympathetic to Elgin's but includes a range of further developments, and to analyse its role in an account of understanding. We first address the structure of reflective equilibrium as a target state and argue that reflective equilibrium requires more than an equilibrium in the sense of a coherent position (i.e. an agreement of commitments, theory and background theories). On the one hand, the position also needs to be stable between a 'conservative' pull of input commitments and a 'progressive' pull of epistemic goals; on the other hand, reflective equilibrium requires that enough of the resulting commitments have some credibility independent of the coherence of the position. We then turn to the dynamics of reflective equilibrium, the process of mutual adjustment of commitments and theories. Here, the most pressing internal challenges for defenders of reflective equilibrium arise: to characterize this process more exactly and to explain its status in relation to reflective equilibrium as a target state. Finally, we investigate the role of reflective equilibrium in Elgin's account of understanding and argue that objectual understanding cannot be explained in terms of reflective equilibrium alone. An epistemic agent who understands a subject matter by means of a theory also needs to be able to use this theory and the theory needs to meet some external rightness condition.

Keywords Reflective equilibrium · Justification · Understanding · Truth · Coherence · Theoretical virtues

✉ Christoph Baumberger
christoph.baumberger@usys.ethz.ch

Georg Brun
Georg.Brun@philo.unibe.ch

1 ETH Zurich, Institute for Environmental Decisions, Universitätstrasse 16, 8092 Zurich, Switzerland

2 Institute for Philosophy, University of Bern, Länggassstrasse 49a, 3012 Bern, Switzerland

1 Introduction

Reflective equilibrium is an account of epistemic justification which is usually traced back to the work of Goodman (1983) and Rawls (1999b) and standardly understood as involving a dynamic and a static aspect: we start out with our judgements about a subject matter, introduce systematic principles that aim at accounting for these judgements, and—since in all likelihood, there will be discrepancies between the proposed principles and our judgements—proceed by adjusting principles and judgements until we reach a state in which our judgements agree with the systematic principles and are supported (in so-called "wide" reflective equilibrium) by background theories.

Appealing to reflective equilibrium is popular in all philosophical disciplines from metaethics and bioethics to logic and philosophy of science, and some philosophers consider it to be the core method of philosophy. David Lewis, for example, writes "Our common task is to find out what equilibria there are that can withstand examination" (1983, p. x; see also Keefe 2000, ch. 2.1; DePaul 1998). Catherine Elgin, however, goes further. Her project is to develop a reflective-equilibrium based epistemology. In a series of papers, chapters and books (most notably Elgin 1983; 1996; 2014; 2017) she has presented an extensive and elaborate analysis and defence of reflective equilibrium. But although—or maybe because—her theory has the potential of moving the discussion about reflective equilibrium to another level, it has been widely ignored in the ongoing debate about reflective equilibrium.[1] This, we submit, is a glaring shortcoming of the current debate.

Particularly in her recent work (e.g. Elgin 2006; 2007; 2012; 2017), Elgin embeds her account of reflective equilibrium in a wider epistemological picture, which involves a shift in focus from propositional knowledge to the understanding of entire subject matters or topics, so-called "objectual understanding". Even though this has made her a key figure in the ongoing debate about understanding, the idea of a reflective equilibrium has hardly been taken up in this debate.

In this paper, we pursue two goals: we develop an account of reflective equilibrium which is sympathetic to Elgin's but includes a range of further developments, and we analyse its role in an account of understanding. Although our account of reflective equilibrium is based on Elgin's, we do not attempt to deal with all intended applications of Elgin's conception of reflective equilibrium. Elgin addresses understanding not only in the sciences, the humanities and philosophy, but also in the arts and everyday epistemic practices. She therefore analyses the epistemic functions of a broad range of representational devices, including theories and models, as well as non-propositional and non-verbal (e.g. pictorial or diagrammatic) symbols used literally and metaphorically (Elgin 2017, p. 89). This paper has a more limited scope in two respects. Firstly, we focus on epistemic justification in the context of understanding by means of theories, but use "theory" in a very broad sense that includes scientific theories, theory-like representations such as mathematical models, and normative and philosophical theories. Secondly, we will often simplify and address theories as systems of propositions, leaving aside non-propositional elements such as categories, diagrams and graphs.

[1] There is, for example, no substantial reference to Elgin's work in any recent survey on reflective equilibrium (see Cath 2016; Daniels 2018; DePaul 2011; 2013; McPherson 2015; Tersman 2018).

Section 2 develops an account of reflective equilibrium. We first address the structure of reflective equilibrium as a target state. In Sects. 2.2, 2.3 and 2.4, we analyse key features of Elgin's theory, specifically, the holistic character of reflective-equilibrium states, the role of theoretical virtues, and non-coherentist aspects of epistemic justification. This leads us to suggest a number of further developments, which are based on distinctions routinely overlooked in the literature. In particular, we argue that an epistemic agent's commitments about a subject matter need to be distinguished from her epistemic goals and the theory she develops. Using these distinctions, we put forward four conditions a position in reflective equilibrium must meet. The result is an account of reflective equilibrium that is substantially more elaborated than the usual descriptions in the literature. Section 2.5 then focuses on the dynamics of reflective equilibrium, the process of equilibrating. It brings us to what we take to be the most pressing internal challenges for defenders of reflective equilibrium: characterizing this process more exactly and explaining its status in relation to reflective equilibrium as a target state. In Sect. 3, we turn to the broader epistemological picture and investigate the role of reflective equilibrium in Elgin's account of understanding. We examine how far reflective equilibrium can take us in providing an explanation of objectual understanding, and whether an account of understanding requires additional conditions.

2 The structure of reflective equilibrium

Elgin summarizes her basic understanding of reflective equilibrium as a target state as follows:

> Because the elements of the resulting account are reasonable in light of one another, they are in equilibrium; because the account as a whole is as reasonable as any available alternative in light of the relevant antecedent commitments, its equilibrium is reflective (Elgin 2017, p. 66).[2]

In comparison with what we standardly find in the literature on reflective equilibrium, some points immediately stand out. At the heart of Goodman's and Rawls's account of reflective equilibrium, there is a contrast between two basic elements, namely the practice of inferring and judgements about logical validity versus principles of valid inference (Goodman 1983, pp. 62–6), and ethical judgements versus principles (Rawls 1999b, pp. 17–9). Elgin, however, does not give centre stage to such a contrast, but rather speaks of an *account,* which she takes to include not merely a set of propositions about a subject matter but a wide variety of commitments which also pertain to methods, standards and goals.[3] Another striking feature of Elgin's account of reflective equilibrium relates to the status of coherence. Whereas coherence is often taken

[2] Similar statements can be found in many instances since Elgin (1993), e.g. 1996, pp. 13, 99, 107; 2014, p. 254.

[3] In this paper, we use "account" both in the usual, non-terminological, sense (especially when speaking of "an account of reflective equilibrium") and in the technical sense introduced by Elgin. We trust that the context makes clear which sense is intended. We do not suggest that Elgin's account of reflective equilibrium is an account in her technical sense.

to be the essence of reflective equilibrium, Elgin appeals not only to coherence ("reasonable in light of one another"), but explicitly insists on a non-coherentist element when she demands that a justified account additionally needs to be appropriately tied to "antecedent commitments".[4]

In what follows, we take Elgin's non-standard theory of reflective equilibrium as a starting point for suggesting an account of reflective equilibrium which includes a range of further developments. In Sects. 2.2, 2.3 and 2.4, we argue that some additional distinctions are needed, both with respect to the elements that are involved in a reflective-equilibrium state and the conditions for such a state. In particular, we distinguish, within an account, between an agent's commitments to propositions about the subject matter at hand, the epistemic goals the agent tries to do justice to (Sect. 2.2) and the theory of the subject matter the agent develops (Sect. 2.3). As concerns the conditions for reflective equilibrium, we argue that a more perspicuous account of reflective equilibrium replaces Elgin's two conditions of reasonableness by four partly antagonistic requirements (Sect. 2.4). Section 2.5 then analyses the dynamic aspect of reflective equilibrium in more detail and investigates the relevance of the process of equilibrating for justification by reflective equilibrium.

There are, however, a number of interesting points of Elgin's conception of reflective equilibrium which we will not be able to discuss here, for example, the way in which reflective equilibrium is a matter of degree, the requirement that an account in reflective equilibrium should be at least as reasonable as any available alternative, as well as social aspects of reflective equilibrium, which become conspicuous once we realize that the commitments of others are relevant too and that striving for reflective equilibrium is, especially in the sciences, a collaborative process.

A good starting point for getting a better grasp of Elgin's s conception of reflective equilibrium is her broad notion of *account* sketched above.

2.1 Elgin's notion of an account

When Elgin characterizes reflective equilibrium as a state in which the elements of an account are "reasonable in light of each other", she is quick to underline that an account includes an entire range of different types of elements (e.g. 2017, pp. 12, 65). To begin with, there are elements which directly constitute the content of one's understanding of the subject matter in question. They include all kinds of propositions (from the singular and concrete to the most general and abstract) about the subject matter, as well as non-propositional elements such as information coded in certain non-verbal representations (e.g. diagrams) and categories used to structure the subject matter. In what follows, we call the ensemble of these object-level elements a *position.* In addition, an account also comprises the epistemology and methodology which guides the investigation of the subject matter, the standards for assessing a position, as well as the pragmatic-epistemic objective that motivates the search for a position in reflective equilibrium.

As an example, take a decision theorist who seeks to understand preferences. The objective driving her inquiry may be, for example, to develop a normative theory of

[4] This idea of a tie needs further analysis; we will undertake it in Sect. 2.4.

rational preferences and she places a high premium on standards such as numerical precision, mathematical tractability and general applicability, maybe at the expense of being representable in an easily graspable way. The position she comes up with may include commitments about particular examples (i.e. the rationality of some specific preferences of a particular person; e.g. that my preferring apples to prunes, prunes to pears and pears to apples is irrational) or general judgements (e.g. that cyclic preferences (as in the preceding example) are irrational); a mathematical notation for representing preferences (e.g. xRy, xPy etc.); theoretical statements and models (such as Sen's Property α expressing "independence of irrelevant alternatives", which holds that if, e.g. apple is my preferred choice, then it should be my preferred choice independent of whether prunes are on offer too); categories (a comparative notion of *x is weakly preferred to y*); and criteria for relating stated preferences to their formal representations.

The main reason why Elgin works with such a broad notion of an account is that it dovetails with holism about justification (e.g. 1989, p. 91; 1996, p. 13; 2017, pp. 63, 102) and with the view that reflective equilibrium is an account of epistemic justification which is general rather than tailored to a specific domain such as logic, ethics or theories of rationality.[5] This is best explained together with the metaphor of background and foreground (made prominent by Daniels 1979). The picture is this: When we develop an understanding of some subject matter, we focus on representations about this subject; they constitute the position which is in the foreground. The justification of a position, however, additionally involves background theories, which may support or tell against the position in the foreground, as well as background information which is needed to establish inferential relations between elements of the position. In the example of our decision theorist, background theories may include theories of cognitive limitations, degrees of belief and probability, as well as epistemological principles about, e.g. the conditions under which intuitions about the rationality of preferences are not trustworthy. Background information about a person's actual preferences will be needed to determine whether some given theoretical principles deem his or her preferences irrational. Holism results because what is in the background must in turn be justified by a reflective equilibrium (in which it is in the foreground) and because the distinction between background and foreground only means that the background is treated as independently justified to some degree, but not as being immune from revision on principle. Consequently, the contrast between foreground and background is established by the epistemic project at hand and reflects the fact that inquiry cannot but proceed piecemeal, even though justification is ultimately holistic; that is, a matter of a reflective equilibrium that encompasses all the epistemic agent's commitments.

In the next three sections, we argue that a more elaborate and convincing conception of reflective equilibrium can be developed if we introduce, within an account, a number of distinctions between elements that perform different functions in a reflective equilibrium.[6]

[5] Another reason is Elgin's shift from knowledge to understanding (2017, pp. 12–4). Non-propositional representations, categories, methods and standards provide important means of understanding but need to be justified as well since they can differ in cognitive merit no less than judgements can.

[6] Sections 2.2, 2.3 and 2.4 further develop our previous work on reflective equilibrium, esp. Baumberger and Brun (2017), Brun (2014; 2017).

2.2 Epistemic goals

For understanding the dynamics of equilibrating as well as the nature of reflective equilibrium as a target state, it is crucial to see that an account in Elgin's sense includes, along with commitments about the subject matter, a range of epistemic goals, which comprise coherence and further virtues well-known from the debate about theory choice (Kuhn 1977; see Douglas 2013 for an overview). Some of these virtues are generally relevant, e.g. simplicity, conceptual clarity, precision, fruitfulness, explanatory power, and scope of application; others are specific to subject matters or types of theories such as decidability in logics, identification of causal mechanisms in biology, visualizability in physics (see De Regt 2017, pp. 36–9).

Clearly, such virtues can be exhibited in various degrees and different virtues can pull in opposite directions, making trade-offs inevitable, for example, when more numerical precision can only be had at the expense of scope of application. An inquiry will therefore be guided by a configuration of epistemic goals which sets constraints on which virtues are relevant, how much weight they have and what trade-offs they admit for. Such a configuration cannot be determined by completely general epistemic considerations but depends on the pragmatic-epistemic objective of the inquiry. If, for example, an agent-based model is developed with the objective of understanding basic mechanisms of social segregation, it should be simple and applicable to a broad range of cases, but we may not insist that the model is useful to effectively determine specific segregation patterns. If, on the other hand, we want to understand and reliably predict segregation in a specific area, we will require that the model be as detailed as necessary for effectively computing sufficiently precise segregation patterns, even if this means that the model gets very complicated.

In the context of reflective equilibrium, the configuration of epistemic goals set in view of the pragmatic-epistemic objective plays a crucial role because it provides the very reason why an epistemic agent should not simply stick to whatever he was initially committed to about the subject matter. In fact, it is the key driver for the reflective-equilibrium process. Giving epistemic goals and theoretical virtues centre stage, as Elgin has been doing all along (since Elgin 1983, pp. 185–7), is all the more important as a great deal of the literature on reflective equilibrium merely speaks of "systematic principles", thereby implicitly alluding to virtues of theories without discussing their function in reflective equilibrium (there are exceptions, e.g. Keefe 2000, ch. 2.1). However, while Elgin typically refers to epistemic goals and commitments about the subject matter in the same breath (see, e.g. Elgin 2017, pp. 12, 65, 85), we distinguish them because they function differently in reflective equilibrium. The distinction allows us to isolate a key condition reflective-equilibrium states must meet: the position must do justice to the relevant configuration of epistemic goals.

Although we distinguish the epistemic goals from the other elements of an account in Elgin's sense, we do not want to deny that, as Elgin emphasizes, the epistemic goals share important features with the commitments about the subject matter. To begin with, the configuration of epistemic goals can be adjusted in the process of developing a reflective equilibrium. We may, for example, realize that we can settle for some simple general principles if we accept a cutback in precision we now consider

acceptable given the objective at hand. Often such adjustments can be interpreted as putting more or less weight on reaching a goal; in other cases a goal may also be re-interpreted. So even if inquiry starts with some configuration of epistemic goals, its exact characteristics (the relative weight and the precise interpretation of the epistemic goals) is a product of, not a precondition for, the process of developing a reflective equilibrium (Elgin 1983, pp. 186–7; 2006; 2017, p. 89). Moreover, epistemic goals can be supported or undermined by background theories about specific theoretical virtues and their cognitive merits (e.g. Elgin 1996, p. 105). This brings us back to the holistic character of justification. Understanding a theoretical virtue, say, simplicity requires to develop an account of simplicity, and this calls for a reflective equilibrium with simplicity in the foreground. In this way, we also get the resources to support or criticize epistemic goals as (un)reasonable, in much the same way as commitments concerning other subject matters.

2.3 Commitments versus elements of a theory

The role played by epistemic goals and theoretical virtues gives us also reason to come back to a feature of Elgin's theory we pointed out at the beginning of Sect. 2. At the very centre of all standard accounts of reflective equilibrium, there is a distinction between two types of elements, usually labelled "judgements" and "principles"; the process of equilibrating is then characterized as "mutual adjustment" (Goodman 1983, p. 64) or "going back and forth" (Rawls 1999b, p. 18) between judgements and principles, and the state of reflective equilibrium as an "agreement" (Goodman 1983, p. 64) or "match" (Rawls 1999b, p. 18) between the two. Elgin, however, does not rely on such a contrast between two key elements. Rather, she describes a process of striving for coherence among several types of elements and therefore refers to iterative adjustments which are not *bi*directional but rather '*omni*directional'. This accords with her emphasis on holism and gains plausibility once we recognize that the adjustments involve elements which are not about the subject matter. For example, conflicts between epistemic goals may be an incentive to modify the configuration of epistemic goals, and background theories can both induce and undergo adaptions. In the case of our decision theorist, commitments about preferences may be revised because they are not in line with background assumptions about rational actions and beliefs (as, for example, a Dutch Book argument can show).

Nevertheless, we argue that, in the context of understanding by means of theories, there are reasons to insist on distinguishing between commitments to propositions (or other representations) about the subject matter on the one hand, and elements of a theoretical system which provides the understanding on the other hand. The distinction is needed for describing the equilibrium that is characteristic for the target state, as well as for understanding the way in which epistemic goals drive the process of equilibrating. Before we can argue these points, the distinction needs to be explained in more detail.

Distinguishing between commitments about the subject matter and elements of a theory is not a matter of content. It is rather intended to capture two roles a proposition about a subject matter can play. Such a proposition can be something that the epis-

temic agent accepts, believes, takes as the best working hypothesis or is committed to in another way. Or it can be something that the epistemic agent uses for giving a systematic account of the subject matter. The two roles are neither mutually exclusive nor strictly correlated; the same proposition can but need not play both roles. For example, the proposition that rational preferences are transitive may well be something an epistemic agent is committed to and something that is explicitly stated in a theory about preferences. Addressing a proposition as a commitment amounts to treating this proposition as the object of an attitude that assigns to it an epistemic status of at least "initial tenability" (Elgin 1996, pp. 101–7) or "initial credibility" (Goodman 1972, pp. 62–3; Rawls 1999a); that is, the proposition is considered to have something speaking in favour of it or to be at least a starting point we find ourselves at (Elgin 2017, pp. 64–5).[7]. In the initial state of a reflective-equilibrium process, the epistemic agent starts with propositions about the subject matter which all have this epistemic status. As pointed out above, the reason why epistemic activities do not simply stop here is that the initial commitments, as a rule, do not do justice to the relevant epistemic goals. The epistemic agent then seeks to make progress by developing a systematic account of the subject matter. In order to do this, she has to introduce elements of a theory. This means that she has to treat certain propositions as elements of a candidate theory. Treating a proposition in this way is not the same as being committed to it even if in fact the agent has already been committed to it. In particular, treating a proposition as an element of a theory does not amount to assigning it any particular epistemic status at all.[8]

The reason why we insist on the contrast between the elements of the theory the epistemic agent introduces and her commitments is that they differ in two respects. Firstly, as argued in the preceding section, reflective equilibrium calls for a systematic account of the subject matter which exhibits certain theoretical virtues. But this requirement presupposes a distinction between those elements of a position which are elements of the theory and those which are not. After all, theoretical virtues are virtues of theories, not of just any commitments. If, for example, one epistemic goal is to find a theory that is economically axiomatized, we need to know which propositions are meant to be the axioms of the theory.

Secondly, the epistemic agent does not need to be committed to all the individual elements of the theory. During the process of equilibrating, she may introduce theories in order to see in which ways she can, given her current commitments, make epistemic progress. She does so by confronting a tentative theory with her commitments and

[7] Elgin's reference to initially tenable commitments marks an important contrast to the accounts of reflective equilibrium in the tradition of Rawls and Daniels, which restrict the 'input' to the process of equilibration to *considered* judgements. Despite her occasional use of "considered judgement" (in Elgin 1983, pp. 187, 188; 1996, pp. 12–5, 158), Elgin admits all commitments at the initial stage. The flimsy and problematic ones are weeded out by the reflective-equilibrium process, not by a 'filter' that operates beforehand. This eliminates questions about the standards for counting as considered (see also Walden 2013).

Furthermore, we use the term "commitment" for both explicitly acknowledged and merely inferred commitments

[8] Note that this contrast between commitments and elements of a theoretical system gives the distinction between judgements and principles another sense than the usual explanation in terms of particular versus general. We have discussed elsewhere (e.g. Baumberger and Brun 2017, p. 172; Brun 2017) why we consider the standard explanation to be inadequate.

then decides whether she will proceed by making changes to the theory or by altering some of her commitments. Importantly, neither being a commitment nor being an element of the theory implies a privilege not to be changed. And even if a state of equilibrium is reached, the epistemic agent may well want to treat some elements of the theory as theory-internal and in this sense need not be committed to them. Take a successful logician who has developed a formal theory of deductive validity which is in agreement with her commitments to the validity of inferences. It may well be that her theory contains much more than merely principles which deem certain inferences valid, for example, a model-theoretic apparatus or principles which decide what counts as a well-formed formula. Nonetheless, the logician need not have any commitments to what should count as a well-formed formula, and it would even seem bizarre if she adjusted the logical system to some independent conviction of what should count as a formula.

In what follows, we will use "commitment" in the narrower sense introduced in this subsection; that is, we reserve this term for the initially tenable or eventually accepted representations about the subject matter and do not use it for the other elements of an account in reflective equilibrium such as epistemic goals and background theories.

2.4 Beyond coherence

Relying on our discussion of epistemic goals, commitments and elements of theories, we can now analyse the criteria reflective equilibria must meet.

In contrast to the widespread understanding of reflective equilibrium as essentially coherentist,[9] Elgin makes it very clear that reflective equilibrium requires coherence but does not boil down to coherence. First of all, we must note that coherence, as Elgin explains it, includes not only the negative requirements of consistency and cotenability (i.e. commitments do not undermine each other, nor do elements of the theory) but also that there are relations of support within and between the commitments and the theory, as well as that the position is supported by background theories (Elgin 2017, pp. 71–3). This covers the usual understanding of the metaphor of an "equilibrium" in terms of agreement. That a position is in equilibrium only if the commitments and the theory agree means that it must be possible to infer (all and only) the commitments from the theory,[10] and so-called "wide" reflective equilibrium additionally requires that the position is supported by background theories.

However, reflective equilibrium asks for more than coherence in this sense. As we have discussed in Sect. 2.2, reflective equilibrium requires that the epistemic agent's position does justice to the relevant epistemic goals. This is a non-coherentist criterion since many epistemic goals relate to virtues of theories, which are not aspects of coherence, e.g. precision, conceptual clarity and visualizability. This point—that equi-

[9] See, e.g. Lycan (2014), Millgram (2008). Coherentist interpretations of reflective equilibrium might be inspired by Rawls's remark that "justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view" (1999b, pp. 19, 507; but cf. 1999a).

[10] "Inference", as we use it here, is not limited to deductive consequence but includes forms of defeasible reasoning such as reasoning with *pro tanto* principles. It also very often includes a transition from expressions couched in a more or less formal language to expressions in ordinary language, e.g. from a logical proof to a commitment which deems an argument in, say, Punjabi as valid.

librium requires, in addition to coherence, doing justice to epistemic goals—may be acceptable to many who see reflective equilibrium as a coherentist method. But Elgin also insists that a position in *reflective* equilibrium must have an appropriate "tie to initially tenable commitments" (1996, p. 107; see also 2017, pp. 64, 66). In what follows, we analyse Elgin's requirement of a tie to initially tenable commitments and argue that it actually should be replaced by two distinctively non-coherentist requirements which are both central to the idea of a reflective equilibrium.

One way of arguing for a requirement that involves initially tenable commitments is this: if there were no limits to altering commitments, there would be no guarantee that the process of equilibrating would not in fact change the subject. If, for example, a theory should deem unrestricted egoism, hurting others for one's own enjoyment and racist discrimination right, we would not revise our moral commitments, but rather conclude that this is not a *moral* theory. And a theory that introduces Prior's (1960) tonk-operator does not count as a theory of valid inference, because it would require us to give up the commitment that logical inference is a relation that does not hold between any two arbitrary sentences. Considerations with respect to changing the subject are not explicitly elaborated by Elgin but alluded to when she writes: "Because our initial commitments constitute our previous best guesses about the topic, we use those commitments as a threshold for assessment. The account we arrive at should be recognizable as an improvement upon them" (Elgin 2017, p. 66).

To elaborate this idea, we need to introduce the new notion of input commitments. The reason is that not only initial commitments—commitments held at the initial stage of inquiry—are relevant for securing that the process does not change the subject. After all, new commitments about the subject matter can emerge during the process of equilibrating for a variety of reasons. The epistemic agent may become aware of new information by observation, reflection, memory, testimony and so on in just the same way as she acquired her initial commitments. To accommodate this, we call the initial together with the later emerging commitments "input" commitments and contrast them with "purely inferential" commitments; that is, commitments which the epistemic agent adopts solely because she inferred them from the theory currently under construction. In contrast to input commitments, purely inferential commitments may not be used to assess whether the resulting position can be seen as providing an understanding of the original subject matter. All they could contribute to answering the subject-change worry would either be simply derived from input commitments or open the door to artefacts of the systematization.

But in what sense should the resulting commitments be "tied" to the input commitments? To explain this relation, we can adapt Elgin characterization of a "tie". The idea is not that a position in reflective equilibrium needs to include all or a certain minimum share of the input commitments, but only that it respects them insofar as "we need a reason to give [them] up" (Elgin 2017, p. 64; see also 1989, p. 91; 1996, p. 107) and the resulting account must "show why they seemed as reasonable as they did when they did" (2017, p. 67). In other words, the epistemic agent must be in a position to explain why input commitments were discarded, replaced or adjusted. Such explanations may take a variety of forms, for example, "a system may show why we were misled into accepting" the commitment (Elgin 2014, p. 254), a conflict between commitments can be resolved by giving up a commitment that is quite feeble anyway,

or a possible gain in terms of doing justice to epistemic goals may outweigh the cost of giving up a few commitments.[11]

The resulting requirement of respecting input commitments can now be characterized as follows. Since the input commitments encode our current understanding of the subject matter, the process of equilibrating may not change them unrestrictedly. Although no input commitment is immune from revision and there is no principled limit to the number of changes a process of equilibrating may implement, there is a limit to the admissible adjustments: we must be in a position to explain the changes made to the input commitments in a way that makes it plausible that the resulting position is still a position about the subject matter at hand.[12]

The condition just described may seem rather indeterminate. It does not protect any specific commitments, nor does it give us rules for deciding which commitments may (not) be adapted. This, however, does not imply that all input commitments are to be treated on a par. Yet, the question of which changes to the input commitments are compatible with addressing the relevant subject matter must be answered by arguments which relate to the subject matter as well as to the pragmatic-epistemic objective, and hence neither permits a formal ("Don't change more than x% of the commitments!") nor a context-independent answer.[13]

A second requirement that involves initially tenable commitments can be developed much more directly out of Elgin's writing. She argues that a tie to initial commitments is needed to deal with the standard charge that coherentists must hold that justification could be generated by coherence alone out of nothing.[14] Here Elgin follows Goodman (1972, pp. 62–3) and Scheffler (1954, p. 181) in arguing that, indeed, coherence cannot create credibility *ex nihilo* but only boost some independently given credibility. Justification by reflective equilibrium therefore requires that the resulting position includes at least some commitments whose credibility is independent of the coherence of the resulting position.[15] Such independent credibility can have a number of sources, for example: a commitment may accommodate some evidence given by observation, intuition or testimony, or it may be supported by background theories (Elgin 2017,

[11] Note that such explanations may not be available if we focus on individual commitments and individual elements of the theory. *Respecting* is intended to be understood as a relation between a theory and a set of commitments.

[12] Elsewhere, Brun (2017), we have argued that the criterion of respecting input commitments has its historical and systematic root in Carnap's method of explication, specifically in the similarity condition for adequate explications (see Carnap 1962, §§ 2–3; Carnap 1963, pp. 933–40).

[13] Of course, sticking to a given subject matter is not an absolute epistemic precept. An epistemic agent who fails to reach a reflective equilibrium may reasonably come to the conclusion that she should better change the subject.

[14] A related objection is that coherentists would have to accept coherent fictions as justified; see the exchange between Elgin (2014) and van Cleve (2014).

[15] Consequently, Elgin (2014, pp. 254, 267–8) holds that reflective equilibrium can be classified as a form of weak foundationalism in BonJour's (1985, pp. 26–9) terms. However, reflective equilibrium is not a form of moderate foundationalism since a commitment always needs to have more than just independent credibility in order to be justified by reflective equilibrium; first of all, the commitment must be part of a position in equilibrium.

p. 78). How much independent credibility is actually needed to meet the *ex-nihilo* objection is debated, but can be left open here.[16]

In her discussion of the *ex-nihilo* objection, Elgin gives centre stage to the tie to initially tenable commitments. One might therefore think that independent credibility can be identified with or at least be traced back to the credibility of initial commitments. And indeed, there can and often will be propositions to which the epistemic agent was committed initially and which are also part of the resulting position in reflective equilibrium enjoying the same independent credibility they initially had. Nonetheless, we submit that *independent* credibility, the credibility a commitment has independent of the coherence of the current position, must be distinguished from *initial* credibility, the credibility a commitment had at initial stage.[17]

There a number of reasons for distinguishing between independent and initial credibility, and consequently between the requirement that some resulting commitments have independent credibility and the requirement that the resulting commitments respect input commitments. On the one hand, independent credibility cannot be reduced to the credibility of initial commitments, because any input commitment may have some degree of independent credibility, irrespective of whether it is given at the initial stage or emerges later on. If, for example, new evidence becomes available during the process of developing a reflective equilibrium, the agent can form a new commitment with some degree of credibility which is due to this evidence and independent of whether the new commitment coheres with her other commitments. Such a commitment will be independently credible, but not an initial commitment. On the other hand, the independent credibility of an initial commitment need not be 'inherited' by the resulting commitments. Firstly, the credibility of an initial commitment may get lost during the process of equilibrating, for example, as a result of a consideration that leads to giving the commitment up. The same goes for independent credibility in general. It can get lost (or, for that matter, increase) if, for example, new information becomes available. Secondly, the relation of respecting input commitments does not necessarily 'transmit' credibility. For example, the current commitments can respect the input commitments even if some of the latter have been replaced by incompatible commitments, provided there is an adequate explanation for why this has been done. In such cases, the independent credibility (if any) of the new commitments cannot be traced back to the credibility of the corresponding initial commitments. Moreover, the 'transmission' of independent credibility may also be undermined by conceptual shifts and incommensurabilities as Kuhn characterized them paradigmatically with reference to, for example, the concept *mass* (Kuhn 1996, pp. 101–3). When such incommensurability is involved, a current commitment may not be identical to an initial commitment even if the two can be expressed in exactly the same words. If the change in meaning is radical enough, the credibility of the initial commitment may cease contributing to the credibility of the current commitment, which, of course, may be established in other ways, for example, by accommodating some evidence. Hence,

[16] For formal analyses see van Cleve (2011) and Roche (2012).

[17] Goodman (1972, p. 63) and Scheffler (1954, p. 181) used "initial credibility" for what we call "independent credibility". When later on reflective-equilibrium theorists (e.g. Elgin 1983; DePaul 1993) started to use "initial" to refer to the first stage of the process of developing a reflective equilibrium, the distinction got lost (see also Johnsen 2017, p. 123).

even if the independent credibility of many resulting commitments can be traced to their credibility at the initial stage, independent and initial credibility must not to be conflated.

The analysis in the preceding paragraphs has shown that two different requirements are needed in addition to coherence and doing justice to epistemic goals: credibility of the resulting position requires independent credibility, and adherence to the subject matter requires respecting input commitments.

As a result, we do not only arrive at a conception of reflective equilibrium which is substantially more complex than the alternatives we find in the literature. We can now also spell out in a structured way Elgin's basic idea that reflective equilibrium calls for an account which is internally (its elements are "reasonable in light of one another") as well as externally reasonable (it is, as a whole, reasonable "in light of the relevant antecedent commitments") (Elgin 2017:66; see the quote at the beginning of Sect. 2). Reflective equilibrium as a target state, we suggest, is characterized by four interacting conditions. There is, to begin with, the requirement that the agent's position is in equilibrium insofar as her commitments, the theory and background theories agree with one another. However, not just any such agreement will do. A position in reflective equilibrium must also do justice to epistemic goals and simultaneously respect the input commitments. These two conditions give us a second reading of the metaphor of an "equilibrium" (the first was "equilibrium" as referring to the agreement of commitments and theories): a position in reflective equilibrium must be stable between the pull of two forces, the 'conservative' pull of input commitments and the 'progressive' pull of the epistemic goals. Finally, reaching a reflective equilibrium is also a matter of securing that enough of the resulting commitments have some degree of independent credibility. This picture relies in crucial respects on the distinctions introduced in the preceding subsections. Distinguishing epistemic goals and elements of a theory from commitments about the subject matter is the basis for describing reflective equilibrium as a coherent position between the two antagonistic requirements of doing justice to epistemic goals and respecting input commitments. Distinguishing independent credibility from initial credibility allows us to single out the non-coherentist requirement that a position needs to include commitments with some independent credibility.

2.5 The process of equilibrating

So far, we have discussed reflective equilibrium as a target state, but the idea of a reflective equilibrium also includes a process of mutual adjustment of commitments and theories. In this section, we take a closer look at this process of equilibrating. Our main goal is to explain what we take to be the most pressing internal challenges for defenders of reflective equilibrium: How, if at all, can this process be characterized in a more exact way? And what exactly is the status of the process in relation to reflective equilibrium as a target state?

Most reflective-equilibrium theorists say very little about the structure of the process. From Goodman (1983, p. 64) we at least learn that the process is iterative and bidirectional, as well as that it is guided by the target of an agreement between commitments and a theory and by the 'weight'—i.e. our resistance to modifications—of

commitments and elements of the theory. Elgin no longer refers to this bidirectional structure of resolving conflicts, but describes a much broader spectrum of possible conflicts and types of adjustments (Elgin 1996, pp. 102–5, 129–34; 2017, pp. 65–6, 71–5, 82–4). Conflicts, Elgin argues, can involve not only commitments and elements of the theory, but also background theories, epistemic goals and other non-propositional elements. And instead of giving up a commitment or a principle with less weight, conflicts can, for example, be resolved by replacing all conflicting commitments (Elgin 1996, p. 103). Furthermore, progress with respect to reflective equilibrium can also be made by supplementing commitments and elements of the theory, for instance, by coming up with new commitments for cases that antecedently were unsettled or by postulating entities that complete a theory (Elgin 2017, pp. 65–6). The process, as Elgin describes it, is not deterministic. Its outcome can depend on the sequence of addressing discrepancies (it is "path-dependent") and there can be cases in which the epistemic agent can proceed in different yet equally acceptable ways (Elgin 1996, pp. 134–5; see also Bonevac 2004, pp. 369–70).

All this raises, firstly, the question of whether more specific rules, strict ones or mere guidelines, could be given for the process of adjustments, and if not, how we can explain why it is not possible. But before we can discuss this question, we must first tackle another, more basic, issue: why should an account of reflective equilibrium describe such a process at all?[18] Why not leave the nature of the process completely open and only specify the conditions for having reached a reflective equilibrium state?

Indeed, one might think that the process should be dismissed as relevant to the context of discovery only and be treated as merely a description of how we often proceed in developing a position about some subject matter. This point was in fact already addressed by Goodman (1983, p. 65), who insisted that the process of mutual adjustment is neither intended to describe how we in fact arrive at a position, nor is it to be read as a recipe one should follow in developing a position (see also DePaul 2006, pp. 599, 619; Keefe 2000, p. 42).[19] The idea is rather to spell out what is required for a justification: we need to be able to describe how the position could have been developed from the commitments we started out with, irrespective of whether we have actually devised a theory in this manner. Hence, the process is meant as a reconstruction (in the sense of Carnap 1963, p. 16), not as a description or prescription, of actual cognitive practice. This is also Elgin's view. What is needed for justification is not that the epistemic agent actually arrived at her position by going through a process of equilibrating, but only that such a process can be reconstructed (Elgin 2017, p. 64).

Now the question is what reconstructing a process of adjustments could contribute to the justification of a position. Why should we not simply say that all that matters for justification is that an epistemic agent reaches a state which meets the discussed

[18] Note that we do not ask whether we could dispense with the process-dimension entirely. This is not possible as long as the state of reflective equilibrium is described in a way that involves a difference between initial and resulting commitments.

[19] Because the process is not meant as a recipe, speaking of a *method* of reflective equilibrium may seem misleading. Nonetheless, it may often prove helpful to use such a process as a means for reaching a state of reflective equilibrium. We cannot pursue the methodological perspective further here, but see Rechnitzer (2018) for case studies on explicit applications of reflective equilibrium.

conditions, irrespective of how she has or could have reached it? Two lines of argument seem to be available at this point.

One option would be to claim that reconstructing a process of adjustments is a further necessary condition for a position to be justified, because such a reconstruction contributes something to justification which is not covered by the conditions discussed in Sects. 2.2, 2.3 and 2.4. But it is hard to see what this contribution could be, as long as the process is not described in a more specific way. That *any kind* of process must be reconstructable is an empty requirement which surely adds nothing to justification. Only with reference to specific constraints on admissible processes or specific rules for making adjustments one could argue that they ensure that the process actually contributes something to justification. Hence, working out this line of argument presupposes that one can answer the first question above by actually specifying rules for the process of equilibrating. The crucial challenge for the current line of argument, however, is to show that the process of adjustments contributes something additional to justification, in contrast to being merely instrumental in reaching a state that meets the criteria for a reflective equilibrium. It is far from obvious what this additional contribution could be. We will not pursue this strategy of defending the justificatory relevance of the process of equilibration, but rather explore the second line of argument.

A more plausible option, we submit, is to argue that the reconstructed process of adjustments is *indirectly* relevant to justification: reconstructing such a process is not an independent contribution to justification,[20] but often the only way to decide whether a position is in reflective equilibrium; that is, whether it meets the four discussed conditions. The reason is that it will often be impossible to decide directly whether a proposed position is in reflective equilibrium, even if we know all input commitments and the relevant epistemic goals. The problem, more specifically, is that we may be unable to decide whether the necessary trade-offs between the various epistemic goals and between respecting input commitments and doing justice to the epistemic goals can be defended. This problem arises because we often cannot specify an exact configuration of epistemic goals in advance. We may know, for example, that our pragmatic-epistemic objective demands that we do not give up too much precision in favour of simplicity, but how much we must finally give up is just as much a research question as the question of which commitments we will end up with in reflective equilibrium. Going through a process of equilibrating and taking context-specific 'piecemeal' decisions on adjustments—decisions which are sensitive to the given subject matter and to the pragmatic-epistemic objective and which affect only a surveyable part of a position—then seems to be the only feasible way of developing a precise configuration of epistemic goals. If this is so, there is no way of deciding whether a position is in reflective equilibrium without reconstructing a process of equilibrating.

But why should we think that reconstructing a process of 'piecemeal' adjustments is a way of defending that we have reached a position in reflective equilibrium? It is tempting to claim that we can defend it by pointing out that we have reached the state, or could have reached it, by a sequence of individual adjustments which all brought

[20] Goodman seems to concur: "in the agreement achieved lies *the only* justification needed for either [commitments and the system]" (1983, p. 64; our italics).

us closer to the target state as we conceived of it at the respective stage. But this is problematic since a sequence of improvements may well lead us into a local optimum; that is, into a state we cannot directly improve upon but only if we are ready to take steps that result in a 'temporary' worsening before they realize a further improvement (Thagard 2009).

To make room for temporary setbacks, we seem to be forced to retreat to a more liberal characterization of permissible adjustments: they only need be reasonable with respect to the goal of reaching a reflective equilibrium. And this makes it plausible that the first question about more specific rules has a negative answer. If the foregoing reflections are sound, it is hard to see how we could give more exact rules on the most general level of describing the basic idea of reflective equilibrium in abstraction from any specific subject matter and pragmatic-epistemic objective (see also Walden 2013).

Nonetheless, more specific settings may permit a more specific characterization of the process. One possibility is to consider epistemic projects which address certain subject matters or make certain assumptions about the relevant epistemic goals, the form or content of the commitments, or the relevant background theories. Another is to switch from a completely general informal characterization of reflective equilibrium to a more exactly defined account of reflective equilibrium, based on some formal model which includes simplifying assumptions. Such a model may make it possible to explore, in relation to various configurations of epistemic goals, questions such as: Are some rules for adjustments better than others because they lead more often to a reflective equilibrium? Under which conditions are proposed rules most instrumental in reaching a reflective equilibrium?

From a methodological point of view, trying to characterize exact procedures or rules of thumb for adjustments may also prove immensely useful because they give the epistemic agent a 'toolbox' of moves that she can try in her (basically unregulated) search for a reflective equilibrium. Hence, more research on processes of equilibrating is not only much needed for addressing a basic issue in reflective-equilibrium theory, it also promising from the point of view of epistemic practice.

3 The role of reflective equilibrium in an account of understanding

Reflective equilibrium finds its place in an epistemology which focuses on objectual understanding rather than on knowledge of individual facts. A position in reflective equilibrium and objectual understanding share several features, which distinguish them from knowledge (Elgin 1996, pp. 122–4). For example, knowledge can come in discrete bits, but a position in reflective equilibrium involves a comprehensive and coherent system that typically provides insights into how things hang together, which is often seen as a hallmark of understanding (Kvanvig 2003, p. 192; Riggs 2003, p. 218). Furthermore, knowledge implies truth, but idealizations and simplifying assumptions that are known to be false can be incorporated into a position in reflective equilibrium and advance our understanding because reflective equilibrium and understanding are both guided by a plurality of epistemic goals that are not exclusively truth-conducive and admit of trade-offs. Reflective equilibrium and objectual understanding fit thus well together, but what exactly is the role of reflective equilibrium in an account of

objectual understanding? We address this systematic question by discussing the role of reflective equilibrium in Elgin's account of objectual understanding.

In the introductory chapter of *True Enough*, Elgin simply identifies objectual understanding with accepting a position in reflective equilibrium, a view she traces back to her earlier book *Considered Judgment*:

(1) "[A]n understanding of a topic consists in accepting a system of commitments in reflective equilibrium." (Elgin 2017, pp. 3–4; see also 2012, p. 134)

Later in *True Enough*, she provides a sketch for an explication of objectual understanding that does not explicitly refer to reflective equilibrium:

(2) "[A]n understanding is an epistemic commitment to a comprehensive, systematically linked body of information that is grounded in fact, is duly responsive to reasons or evidence, and enables nontrivial inference, argument, and perhaps action regarding the topic the information pertains to." (Elgin 2017, p. 44; see also 2007, p. 39)

If we read (2) in the light of Elgin's discussion in chapters 3 and 4 of *True Enough*, then it can be reconstructed as involving four conditions: A *commitment condition* (understanding requires "an epistemic commitment to a comprehensive […] body of information"), a *justification condition* (the body of information must be "systematically linked" and "duly responsive to reasons or evidence"), a *rightness condition* (the body of information must be "grounded in fact") and an *ability condition* (the body of information needs to enable "nontrivial inference, argument and perhaps action").

While (2) involves four conditions, (1) seems to get along with only one. The reflective-equilibrium condition in (1) certainly implies the justification condition in (2). However, a position in reflective equilibrium involves more than a "systematically linked body of information […] that is duly responsive to reasons or evidence". In the terminology of the previous sections, it involves more than a coherent theory which is part of a position that includes commitments which are independently credible due to the accommodation of evidence. It also requires that the theory does justice to additional epistemic goals and that the resulting position respects input commitments. If we construe reflective equilibrium in (1) along the lines suggested in the previous sections, does (2) still add additional conditions? Or are the ability, the commitment and the rightness conditions too implied by a reflective-equilibrium condition? And if not, should an explication of objectual understanding in terms of reflective equilibrium include these additional conditions? We address these questions by discussing the ability, the commitment and the rightness condition in turn.

3.1 Commitment and abilities

Understanding a subject matter by means of a theory requires, according to (2), an epistemic commitment to the theory. Elgin does not provide an account of commitment, but the outlines of such an account emerge from our reconstruction of her conception of reflective equilibrium: commitment is an epistemically relevant status which comes in degrees, is not immune from revision, and not just related to truth (as, e.g. belief), but to the multitude of epistemic goals to which the position they are part of needs

to do justice. If commitment in (2) is understood in this way, a reflective-equilibrium condition implies the commitment condition. That a theory is part of a position in reflective equilibrium implies that the agent is committed to everything which can be inferred from the theory and, in this sense, is committed to the theory. This makes the commitment condition redundant in an account of understanding in terms of reflective equilibrium.

Whether a reflective-equilibrium condition also implies the ability condition depends on which specific abilities are required for understanding and which, if any, for reflective equilibrium. According to (2), an agent who understands a subject matter by means of a theory needs to be able to use this theory and apply it to the subject matter. Elgin does not specify the required abilities in detail, but based on the examples she discusses and the current literature on understanding,[21] we suggest the following systematization: The ability to use a theory involves (a) the ability to draw consequences of the theory about the subject matter at issue. If the theory is formulated mathematically, this requires (a_1) being able to calculate and interpret results accurately in order to solve quantitative problems. But since it is often possible to cheat one's way to a mathematical solution without really understanding it and the underlying theory, understanding a subject matter through a theory also requires (a_2) being able to solve qualitative problems by drawing consequences of the theory without performing exact calculations. In the case of qualitative theories this corresponds to the ability to answer questions without going explicitly through the argumentation suggested by the theory (De Regt 2017:102). In typical cases, being able to use a theory also requires (b) that one can explain aspects of the subject matter in terms of the theory. It may also require (c) that one is able to assess the conditions and limits of the theory's application. Ability (c) is particularly important in the case of idealized models, which require some awareness of how they diverge from reality and of the conditions under which the divergences are negligible for the application of the models.

A reflective-equilibrium condition only demands any abilities if it is construed internalistically. In our discussion in the previous sections we specified criteria for being in reflective equilibrium, but we left open whether the agent needs to be able to argue that her position meets these criteria. Elgin's use of "reflective" as referring to an equilibrium that we accept or endorse on reflection (Elgin 1996, p. 128; 2017, p. 66) suggests that justification by reflective equilibrium must meet an internalist requirement. We argued elsewhere that such a requirement is necessary for an account of reflective equilibrium that is plausible for objectual understanding (Baumberger and Brun 2017, p. 181). However, the internalist requirement should only demand that the agent can make it sufficiently plausible that her position is in reflective equilibrium. What counts as sufficiently plausible depends on the context; in a scientific context, for instance, more is required than in an everyday context. But in neither context does the agent need be able to show conclusively that her position is in reflective equilibrium. For example, she does not need to be able to actually infer all her commitments by way of explicit arguments from her theory, but only to make it sufficiently plausible that her theory and her commitments agree with one another.

[21] The idea that understanding requires certain abilities is often dealt with under the heading of "grasping" (e.g. Grimm 2010; Newman 2017; Baumberger 2019) or "intelligibility" (e.g. De Regt 2017; Wilkenfeld 2017); for an overview see Baumberger et al. (2017).

In the literature, it has hardly been examined which abilities an internalistic reflective-equilibrium condition requires. Here we can only suggest some initial steps towards answering this question. Being able to make it sufficiently plausible that her theory and her commitments are in agreement requires that the agent can draw characteristic consequences from her theory without performing exact calculations or without logically completely explicit argumentation (a_2), and that she can assess the conditions and limits of the application of her theory (c). The agent also needs to be able to make it sufficiently plausible that her theory does justice to a configuration of epistemic goals. If a theory with explanatory power, computability and numerical precision are among these goals, it may well be that the agent needs also be able to give explanations (b) and draw consequences by performing exact calculations or by explicitly going through the argumentation suggested by the theory (a_1). But even if this is so, it might turn out that understanding requires additional abilities not demanded by an internalistic reflective-equilibrium condition. Candidates are the ability to use the theory as a basis for actions (Elgin 2017, p. 44) and the ability to conduct thought experiments (Stuart 2018). To leave room for such a possibility, an explication of objectual understanding needs an ability condition in addition to the reflective-equilibrium condition.

Hence, objectual understanding by means of a theory requires commitment to the theory and the ability to use the theory. It might be tempting to build the ability condition into the commitment condition and suggest that being committed to a theory involves the ability to use it.[22] However, such a suggestion does not fit well into an explication of objectual understanding in terms of reflective equilibrium, since even an internalistically construed reflective-equilibrium condition implies the commitment condition but not necessarily the ability condition for objectual understanding. Moreover, commitment to a theory should also be distinguished from the ability to use it since one can be able to use a theory without committing oneself to the theory. For example, a historian of science who can explain combustion in terms of phlogiston theory does not need to commit herself to the phlogiston theory.

3.2 Rightness

From *Considered Judgment* on, Elgin has emphasized that understanding requires some kind of 'grounding'. Whether this grounding requirement is implied by the reflective-equilibrium condition, as suggested by (1), or whether an explication of objectual understanding requires an external rightness condition, as it figures in (2), depends on how the grounding at issue is understood. Concerning this question, Elgin seems to have modified her view in her later writings which culminate in *True Enough*.

[22] Elgin's notion of acceptance, which she traces back to Cohen (1992), combines commitment and abilities: "An epistemic agent who accepts a consideration is not just willing to deploy it when her ends are cognitive; she is also able to do so" (Elgin 2017, p. 19). Accepting a theory in this sense involves thus both commitment to the theory and the ability to use it.

Note, that in (1), Elgin identifies understanding with *accepting* a position in reflective equilibrium. If acceptance is understood as just explained, then (1) combines a reflective-equilibrium condition with a commitment condition and an ability condition. According to such a reading, the only possible point of divergence between (1) and (2) is the rightness condition.

In *Considered Judgment*, Elgin explains the grounding that is necessary for understanding in terms of accommodating initially tenable commitments, and thus in terms of what makes an equilibrium reflective: "To be in *reflective* equilibrium, a balanced system must be reasonable in light of what we already have reason to hold; that is, it must answer to our initially tenable commitments about the subject at hand. Such commitments, being independent of the system they tether, provide the standard we need" (Elgin 1996, p. 128; see also p. 107). We have argued that accommodating initially tenable commitments involves two non-coherentist requirements of reflective equilibrium: that the current commitments respect the input commitments and that some current commitments have a credibility that is independent of the coherence of the current position. It seems likely that the grounding required for understanding is supposed to be provided by those commitments that are independently credible because they accommodate some evidence. If so, then understanding does not need an external rightness condition besides the reflective-equilibrium condition. A position in reflective equilibrium is tied to something outside itself, but these are commitments and hence intentional states, not the facts constituting the subject matter.

In *True Enough*, Elgin argues that understanding needs to be tied to the facts and acknowledges that accommodating initially tenable commitments does not establish the required tie: "I continue to consider the tie to [initially tenable] commitments important. But we also need a tether that connects the system to the facts it pertains to." (Elgin 2017, pp. 183–4) Elgin insists that the tether does not need to be truth. Understanding thus requires an external rightness condition, but it should not be construed as a truth condition. A truth condition, she argues, cannot do justice to the cognitive contributions of science since models and idealizations are acknowledged not to be true but nonetheless partly constitutive of the understanding science delivers (Elgin 2007; 2017, pp. 57–62). Elgin suggests construing the external rightness condition rather as a true-enough condition. A theory needs only to be true enough to enable understanding, and it can be true enough if its central propositions are not even approximately true and if it contains non-propositional elements that are not even truth-apt. Elgin suggests that falsehoods and non-propositional representations can contribute to understanding if they exemplify—i.e. instantiate and refer to—important features they share with the facts. Exemplification, she urges, "provides at least as strong and stable a link to the phenomena as truth does" (Elgin 2017, p. 2).

Irrespective of how exactly a true-enough condition is spelled out, it is not implied by a reflective-equilibrium condition, because reflective equilibrium relates to true enough in a similar way as justification to truth in standard accounts of knowledge: as justification (in the case of knowledge) has to speak in favour of the belief in question being true, reflective equilibrium has to speak in favour of the position in question being true enough.

The view that an account of objectual understanding requires an external rightness condition dovetails with Elgin's characterization of her own approach as an imperfect procedural epistemology, which recognizes an external standard but cannot guarantee that the results of its procedures satisfy this standard.[23] According to the above

[23] The term "imperfect procedural epistemology" is from *Considered Judgment* (Elgin 1996, p. 20), but Elgin still seems to adhere to the idea in *True Enough* (see e.g. Elgin 2017, p .65). If our argument above is correct, then the term characterizes her current position much better than the earlier one.

suggestion, true enough is the external standard, and reflective equilibrium provides convincing but not conclusive reasons that the standard is met. The view also fits well with Elgin's (2017, p. 71, 75–6) claim that epistemic justification requires both a coherent position and that the best explanation of its coherence is that the position is true enough. Elgin (2014) substantiates this claim by an extensive example in which a theft is investigated, based on the reports of several witnesses. Elgin argues that the best explanation for the coherence of the resulting account is that it gets the facts roughly right. But Elgin does not explicitly connect this inference-to-the-best-explanation requirement to reflective equilibrium.

Here is a rough outline of how this might be done in a way which is based on our account of reflective equilibrium and fits well with Elgin's example. Obviously, the coherence of a position alone does not make it likely that the position is true enough. For an inference to the best explanation, we need to refer additionally to non-coherentist features of the position and argue that the best explanation for the position having all these features is that it is roughly true. We have identified three non-coherentist requirements: doing justice to additional epistemic goals, respecting input commitments and independent credibility. An inference to the best explanation can refer to the first requirement since some epistemic goals are likely to be truth-conducive; promising candidates might be explanatory power and comprehensiveness. But even epistemic goals such as simplicity and visualizability that are clearly not truth-conducive may play a role in such an inference since they can justify that a position is true *enough*, given the additional goals. The second requirement is irrelevant for the intended inference. That a position respects input commitments makes sure that it is in fact about the subject matter at hand, but it does not provide reasons for assuming that the position is true enough. The third requirement, however, is crucial for the intended inference. If enough commitments of a position have some credibility that is independent of its coherence and is rather due to the accommodation of evidence or derives from well-established background theories, then this clearly speaks in favour of the position being true enough. Thus, that a position is true enough is the best explanation if the position is coherent, if it does justice to the relevant configuration of epistemic goals, and if enough of its commitments have independent credibility.[24] If this is correct, then the fact that a position is in reflective equilibrium provides convincing but—since an inference to the best explanation is an non-deductive inference—not conclusive reasons for the claim that the position is true enough and consequently meets the external standard that an imperfect procedural epistemology requires.

According to this analysis, there are good reasons to ascribe Elgin the view that an account of understanding requires an external rightness condition. But, on the other hand, Elgin (2017, pp. 91–121) explicitly rejects standards that are "exogenous" to the epistemic practice and develops a deontological account of epistemic norms according to which epistemic acceptability turns on epistemic responsibility. This suggests that we should accept a position in reflective equilibrium not because it is likely to be true enough (or to meet some other external standard) but because it conforms to epistemic

[24] Note that, in this line of thought, independent credibility does not constitute the grounding necessary for understanding. It is rather part of an argument for claiming that the position in question is true enough and thus sufficiently grounded in fact for providing understanding.

norms that would emerge from the deliberations of idealized agents who are equal and free.

There is thus some tension with respect to whether, according to Elgin, an account of understanding requires an external rightness condition. We have argued elsewhere for such a condition and suggested to construe it as an evaluative criterion for assessing how good someone's understanding is rather than as a necessary condition (Baumberger 2019). In a nutshell, the key idea is that an external rightness condition seems required to accommodate the intuition that if two positions are each in reflective equilibrium to the same degree for their agents (and the agents are also equally able to use them), the position that answers better to the facts provides the better understanding of its subject matter. If a Copernican astronomer of the 16th century and her Ptolemaic contemporary both held a position that was in reflective equilibrium for them and the two astronomers were equally able to use their theories in predictions and explanations, the Copernican understood the motion of the planets better than her Ptolemaic contemporary. If this is correct, then an explication of objectual understanding requires an external rightness condition besides the reflective-equilibrium condition. Elgin is thus well advised to add in (1) the grounded-in-fact requirement that she formulated in (2).

4 Conclusion

The account of reflective equilibrium we have suggested is sympathetic to Elgin's, but it introduces a more differentiated structure, and it is significantly more elaborated than the alternatives in the literature. In particular, we have argued that reflective equilibrium requires more than a coherent position; that is, more than an agreement of commitments, theory and relevant background theories. Firstly, the position also needs to do justice to a configuration of epistemic goals. Since the epistemic goals are the key driver of the reflective-equilibrium process, they should be distinguished from the commitments about the subject matter. Secondly, Elgin emphasizes that the position must accommodate initially tenable commitments about the subject matter. We argue that this involves actually two non-coherentist requirements. In order to make sure that the process of equilibrating does not change the subject, current commitments need to respect input commitments. Input commitments involve initial commitments but also commitments that emerge during the equilibration process, for example because the agent becomes aware of new information. And since coherence cannot create justification out of nothing, enough resulting commitments must have some credibility independent of the coherence of the current position. Importantly, independent credibility cannot be reduced to the credibility of initial commitments but can have other sources as well, for example the accommodation of evidence or support by background theories.

These points relate to reflective equilibrium as a target state, but the idea of a reflective equilibrium also includes prominently a process of mutual adjustments. The process-aspect of reflective equilibrium raises, we think, the most pressing internal challenges for defenders of reflective equilibrium, namely giving a more exact characterization of the process and determining what exactly the status of the process is

in relation to reflective equilibrium as a target state. We argue that the process of mutual adjustments is only indirectly relevant to justification. Running through such a process does not in itself contribute to justification, but reconstructing such a process is often the only way to decide whether a position is in reflective equilibrium. Obviously, adjustments must be reasonable with respect to the goal of reaching a reflective equilibrium. But we do not think that more precise rules can be given in abstraction from specific subject matters and pragmatic-epistemic objectives. This does not rule out more precise descriptions of the process of equilibrating, but suggests that they should be sought for more specific settings. Further investigations are needed to determine whether, for example, formal models can be used to describe the equilibration processes more precisely for certain types of justificatory projects.

Reflective equilibrium is embedded in an epistemology of understanding. We have argued that objectual understanding cannot not be explained in terms of reflective equilibrium alone, however. An internalistic conception of reflective equilibrium can be used to analyse the justification condition and the commitment condition for understanding, but an explication of objectual understanding also requires an ability condition and an external rightness condition. The outlines of an account of commitment emerge from the suggested conception of reflective equilibrium. But this account still needs to be spelled out in detail, and more work is required to fully analyse the ability and the rightness condition. In her most recent writings, especially in *True Enough,* Elgin seems to conceive the rightness condition as a true-enough condition and suggests analysing it in terms of the exemplification of features the theory in question shares with the facts constituting the subject matter. This paper is not the place to discuss how such a rightness condition can do justice to the cognitive contributions of idealizations and non-propositional representations. Nonetheless, we hope we have shown that an account of reflective equilibrium and understanding based on Elgin's work is a strong contender in the debate on objectual understanding and epistemic justification.

Acknowledgements This paper draws on research which is part of the Project *Reflective Equilibrium – Reconception and Application* (Swiss National Science Foundation Project 150251). We thank Finnur Dellsén, Tanja Rechnitzer and the anonymous reviewers for helpful comments, and especially Catherine Elgin for all the critical discussions and support over the years.

References

Baumberger, C. (2019). Explicating objectual understanding. Taking degrees seriously. *Journal for General Philosophy of Science, 50,* 367–388.

Baumberger, C., Beisbart, C., & Brun, G. (2017). What is understanding? An overview of recent debates in epistemology and philosophy of science. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science* (pp. 1–34). New York: Routledge.

Baumberger, C., & Brun, G. (2017). Dimensions of objectual understanding. In S. R. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science* (pp. 165–189). New York: Routledge.

Bonevac, D. (2004). Reflection without equilibrium. *Journal of Philosophy, 101,* 363–388.

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA/London: Harvard University Press.

Brun, G. (2014). Reflective equilibrium without intuitions? *Ethical Theory and Moral Practice, 17,* 237–252.
Brun, G. (2017). Conceptual re-engineering: From explication to reflective equilibrium. *Synthese*. https://doi.org/10.1007/s10670-015-9791-5.
Carnap, R. (1962). *Logical foundations of probability* (2nd ed.). Chicago/London: University of Chicago Press/Routledge and Kegan Paul.
Carnap, R. (1963). Intellectual autobiography. In P. A. Schilpp (Ed.), *The philosophy of rudolf carnap* (pp. 3–84). La Salle: Open Court.
Cath, Y. (2016). Reflective equilibrium. In H. Cappelen, T. S. Gendler, & J. Hawthorne (Eds.), *The Oxford handbook of philosophical methodology* (pp. 213–230). Oxford: Oxford University Press.
Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.
Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy, 76,* 256–282.
Daniels, N. (2018). Reflective equilibrium. In: E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/spr2018/entries/reflective-equilibrium/.
De Regt, H. W. (2017). *Understanding scientific understanding*. New York: Oxford.
DePaul, M. R. (1993). *Balance and refinement. Beyond coherence methods of moral inquiry*. London: Routledge.
DePaul, M. R. (1998). Why bother with reflective equilibrium? In M. R. DePaul & W. Ramsey (Eds.), *Rethinking intuition. The psychology of intuition and its role in philosophical inquiry* (pp. 293–309). Lanham: Rowman and Littlefield.
DePaul, M. R. (2006). Intuitions in moral inquiry. In D. Copp (Ed.), *The Oxford handbook of ethical theory* (pp. 595–623). New York: Oxford University Press.
DePaul, M. R. (2011). Methodological issues. Reflective equilibrium. In C. Miller (Ed.), *The Continuum companion to ethics* (pp. lxxv–lxcv). London: Continuum.
DePaul, M. R. (2013). Reflective equilibrium. In H. Lafollette (Ed.), *The international encyclopedia of ethics* (pp. 4466–4475). Wiley-Blackwell: Malden, MA.
Douglas, H. (2013). The value of cognitive values. *Philosophy of Science, 80,* 796–806.
Elgin, C. Z. (1983). *With reference to reference*. Indianapolis: Hackett.
Elgin, C. Z. (1989). The relativity of fact and the objectivity of value. In M. Krausz (Ed.), *Relativism. Interpretation and confrontation* (pp. 86–98). Notre Dame: University of Notre Dame Press (reprinted in 1997. Between the Absolute and the Arbitrary. Ithaca/London: Cornell University Press. 176–191).
Elgin, C. Z. (1993). Scheffler's symbols. *Synthese, 94,* 3–12.
Elgin, C. Z. (1996). *Considered judgment*. Princeton: Princeton University Press.
Elgin, C. Z. (2006). From knowledge to understanding. In Stephen Herrington (Ed.), *Epistemology futures* (pp. 199–215). Oxford: Clarendon Press.
Elgin, C. Z. (2007). Understanding and the facts. *Philosophical Studies, 132,* 33–42.
Elgin, C. Z. (2012). Understanding's tethers. In C. Jäger & W. Löffler (Eds.), *Epistemology: Contexts, values, disagreement. Proceedings of the 34th international Ludwig Wittgenstein symposium Kirchberg am Wechsel* (pp. 131–146), Austria 2011. Frankfurt a.M.: Ontos.
Elgin, C. Z. (2014). "Non-foundationalist epistemology. Holism, coherence, and tenability" and "Reply to van cleve". In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 244–255; 267–271). Wiley: Malden.
Elgin, C. Z. (2017). *True enough*. Cambridge, MA: MIT Press.
Goodman, N. (1972). Sense and certainty. In *Problems and projects* (pp. 60–68). Bobbs-Merrill: Indianapolis/New York.
Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
Grimm, S. R. (2010). The goal of explanation. *Studies in the history and philosophy of science, 41,* 337–344.
Johnsen, B. (2017). *Righting epistemology. Hume's revolution*. Oxford: Oxford University Press.
Keefe, R. (2000). *Theories of vagueness*. Cambridge: Cambridge University Press.
Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In *The essential tension. Selected studies in scientific tradition and change* (pp. 320–339). Chicago/London: University of Chicago Press.
Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago/London: University of Chicago Press.
Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. New York: Cambridge University Press.
Lewis, D. (1983). *Philosophical papers* (Vol. I). New York/Oxford: Oxford University Press.

Lycan, W. G. (2014). Epistemology and the role of intuitions. In S. Bernecker & D. Pritchard (Eds.), *The Routledge companion to epistemology* (pp. 813–822). London/New York: Routledge.
McPherson, T. (2015). The methodological irrelevance of reflective equilibrium. In Chris Daly (Ed.), *The Palgrave handbook of philosophical methods* (pp. 652–674). Palgrave: Basingstoke.
Millgram, E. (2008). Specificationism. In J. E. Adler & L. J. Rips (Eds.), *Reasoning. Studies of human inference and its foundations* (pp. 731–747). Cambridge: Cambridge University Press.
Newman, M. (2017). Theoretical understanding in science. *British Journal for the Philosophy of Science, 68,* 571–595.
Prior, A. N. (1960). A runabout inference-ticket. *Analysis, 21,* 38–39.
Rawls, J. (1999a). The independence of moral theory. In *Collected papers* (pp. 286–302). Cambridge, MA: Harvard University Press.
Rawls, J. (1999b). *A theory of justice. Revised edition*. Cambridge, MA: Belknap Press.
Rechnitzer, T. (2018). *Applying reflective equilibrium. A case study in justification.* Ph.D. thesis, University of Bern.
Riggs, W. D. (2003). Understanding 'virtue' and the virtue of understanding. In M. DePaul & L. Zagzebski (Eds.), *Intellectual virtue* (pp. 203–226). Oxford: Clarendon Press.
Roche, W. (2012). Witness agreement and the truth-conduciveness of coherentist justification. *The Southern Journal of Philosophy, 50,* 151–169.
Scheffler, I. (1954). On justification and commitment. *Journal of Philosophy, 51,* 180–190.
Stuart, M. T. (2018). How thought experiments increase understanding. In M. T. Stuart, Y. Fehige, & J. R. Brown (Eds.), *The Routledge companion to thought experiments* (pp. 526–544). London/New York: Routledge.
Tersman, F. (2018). Recent work on reflective equilibrium and method in ethics. *Philosophy Compass*. https://doi.org/10.1111/phc3.12493.
Thagard, P. (2009). Why cognitive science needs philosophy and vice versa. *Topics in Cognitive Science, 1,* 237–254.
van Cleve, J. (2011). Can coherence generate warrant ex nihilo? probability and the logic of concurring witnesses. *Philosophy and Phenomenological Research, 82,* 337–380.
van Cleve, J. (2014). "Why coherence is not enough. A defense of moderate foundationalism" and "Reply to Elgin". In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 255–267; 271–273). Wiley: Malden.
Walden, K. (2013). In defense of reflective equilibrium. *Philosophical Studies, 166,* 243–256.
Wilkenfeld, D. A. (2017). MUDy understanding. *Synthese, 194,* 1273–1293.