

ORIGINAL ARTICLE

The surplus value of knowledge

Wolfgang Spohn 

Department of Philosophy, University of
Konstanz, Konstanz, Germany

Correspondence

Wolfgang Spohn, Department of Philosophy,
University of Konstanz, Konstanz 78457,
Germany.

Email: wolfgang.spohn@uni-konstanz.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/
Award Numbers: Grant EXC 2064/1
390727645 Grant SP279/21-1, Grant
SP279/21-1 420094936

Abstract

The *Meno* problem, asking for the surplus value of knowledge beyond the value of true justified belief, was recently much treated within reliabilist and virtue epistemologies. The answers from formal epistemology, by contrast, are quite poor. This paper attempts to improve the score of formal epistemology by precisely explicating Timothy Williamson's suggestion that 'present knowledge is less vulnerable than mere present true belief to rational undermining by future evidence'. It does so by combining Nozick's sensitivity analysis of knowledge with Spohn's fact-asserting epistemic interpretation of conditionals. Accordingly, the surplus value of knowledge lies in a specific kind of stability of knowledge, which differs, though, from that claimed by other so-called stability analyses of knowledge.

KEYWORDS

epistemic conditionals, fact-asserting conditionals, Meno problem, modal accounts of knowledge, stability theory of knowledge, value of knowledge

1 | INTRODUCTION

Why do we strive for knowledge? Epistemology should be able to come up with a clear and concise general answer. Such an answer was given by Savage (1954, p. 118), as exercise 15. Its importance was appreciated only later in papers and chapters which discussed, or were even titled, 'the value of knowledge' (see, e.g., Good, 1966 and Skyrms, 1990, ch. 4). Savage's theorem said that the expected utility of cost-free relevant true information is always positive. That is, a bit more precisely, the maximal expected utility of deciding right away is at most as high as the maximal expected utility of first receiving true information and then deciding accordingly, provided the information is cost-free; equality only holds, when the information is irrelevant to the decision. This gives a perfect explanation of the instrumental value at least of true belief.¹

¹Skyrms (1990, pp. 94f.) traces the basic idea back to an undated handwritten note of Frank Ramsey from the 1920s with the header 'Value of knowledge'. I should add that true belief may also have a purely epistemic value or a value in itself (see section 8 of Pritchard et al., 2018), a point I shall not pursue here.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Theoria* published by John Wiley & Sons Ltd on behalf of Stiftelsen Theoria.

However, so far it does only this. It might be taken to explain the value of knowledge, if knowledge were merely justified true belief (JTB) and if we want to provide an explanation only for rational people who have only (subjectively) justified beliefs. However, in contemporary philosophy, the insight that knowledge is more than JTB even for rational people became overwhelming with Gettier (1963). This raises the question: what is the *surplus* value of knowledge over the value of JTB?² This question is utterly pressing. Without a good answer there is no interest in knowledge beyond JTB, and we could dump Gettierology, which has so decisively advanced the theory of knowledge. Of course, the prevailing intuition is that there must be some positive answer. But it is not so clear how to provide one.

In the last 20 or more years, this question has been widely discussed in the literature—see Greco (2011) and Pritchard et al. (2018)—however, mainly within reliabilist and virtue conceptions of knowledge. I have no desire to add to this discussion, which appears utterly indeterminate to me. There is also the idea that the surplus value of knowledge is an instrumental one, similar to that of true belief. Hawthorne and Stanley (2008), Schulz (2017), Williamson (2005) and others have suggested, but not deeply elaborated, the idea that knowledge grounds our decisions and actions in a specific way improving upon the grounding offered by credences. I won't pursue this idea, either. One should think that there is also a purely epistemological answer.

In short, I want to restrict attention to formal epistemology. This is the only area where one may hope to find precise answers. However, it is surprisingly silent on the issue,³ and I can offer only little comparative discussion at the end. This is perhaps because formal epistemology is mainly of the belief-first type and less interested in knowledge and because it is dominated by Bayesianism, which is constitutionally unable, as it were, to address issues of knowledge.⁴ This paper attempts to improve on the formal side.

As is well known, our question is just the *Meno problem* raised already by Plato in his discussion of the example of finding the way to Larissa. There, Socrates first confuses Meno by contesting the practical surplus value of knowledge over true belief and then moves on to confirm some value of knowledge in a quite obscure remark:

Since the true beliefs, too, for as much time as they stay with you, are fine possessions and bring about entirely good results. But they are unwilling to stick around for long, and they run away out of the soul of a man and so are not worth much, until someone binds them by working out the reason. (*Meno*, 97e/98a)

Plato's Larissa passage has been read in different ways. It has been read as asking for the value of knowledge beyond that of true belief. The response can then be understood as confirming the JTB analysis given in the *Theatetos*; the value is provided by the 'reason'. This might be taken as a sufficient answer. However, this is not how I want to read the passage here. As indicated, I am focusing here on epistemically rational persons. I am not interested in the advantage of being rational over not being rational. Epistemically rational persons hold only rational or subjectively or, in Lehrer's terms, personally justified beliefs. Let us not try to find a stronger notion of justification beyond subjective justification. So, for the purpose of this investigation, true belief and JTB fall into one; 'justified' is just pleonastic. Hence, the surplus value of knowledge over true belief and over JTB poses the same problem. And the first answer to the *Meno* problem won't do then.

Indeed, the Larissa passage has also been read as going beyond the JTB analysis by adding a new idea, namely that true beliefs need to be 'bound' by someone to become knowledge, that

²The literature often refers only to the value of knowledge. But this invites confusion with the older literature I just mentioned.

³Perhaps five of the 150 references of Pritchard et al. (2018) can be assigned to the formal side.

⁴As argued disappointedly already by Plantinga (1993, ch. 6). The basic point, of course, is that Bayesianism has great difficulties (as highlighted by the lottery paradox) and to some extent even no interest in accounting for belief *simpliciter*.

is, that knowledge is distinguished by some kind of stability. ‘Present knowledge is less vulnerable than mere present true belief to *rational* undermining by future evidence’ (Williamson, 2000, p. 79, his italics). This kind of stability may, and will usually, have positive practical consequences, but as such it is a purely epistemic value not referring to agency.⁵ No doubt, this is a very plausible idea, but Williamson does not add much to his assertion.⁶ This paper focuses on this second reading of the Larissa passage and has the modest aim to precisely explicate this stability idea.⁷

What the surplus value of knowledge might be depends, of course, on how we conceive of the surplus meaning of knowledge over JTB. Here, I shall take the so-called sensitivity analysis of knowledge of Nozick (1981, sect. 3.I) as my starting point. Reliability or virtue accounts of knowledge do not lend themselves to a formal treatment. And I will assume that, as far as our issue is concerned, the sensitivity analysis stands *pars pro toto* for other modal analyses of knowledge like the safety analysis of Pritchard (2005) and Sosa (2000) and the normality analysis of Freitag (2013). I could have started from them as well. These analyses have been extensively discussed in the literature. I won’t engage here in their pros and cons. We would lose ourselves in that discussion. I submit that in case these analyses need amendments, these amendments could also be added to my explication of the stability idea.

My point about the modal analyses will indeed be a different one. Namely, they all assume that some specific conditional or counterfactual relation between facts and beliefs is characteristic of knowledge. They have diverging explanations of the conditional used by them. But again, I do not want to discuss this in detail. How to understand conditionals is a huge issue by itself, far beyond the scope of this paper. Rather, I want to directly build on the epistemic account I have given in Spohn (2015), which has not been used in the analyses of knowledge, but seems particularly germane for our purposes. This will allow us to offer a novel explication of Williamson’s quote above. Thus, the upshot of the paper is roughly: Nozick + Spohn Williamson.

However, this is not to say that the upshot is one that promotes knowledge-first epistemology. The paper will rather stay within the confines of belief-first epistemology. Let me explain: In answering our title question it is important to distinguish the third- and the first-person perspective. They give rise to three value questions: two from the third-person perspective: (i) Why is it valuable *for the ascriber*, when he is able to ascribe knowledge of *p* to the knower and not just (justified) true belief in *p*? (ii) Which benefit *for the knower* can the ascriber state by ascribing knowledge of *p* to the knower instead of merely (justified) true belief in *p*? And one from the first-person perspective: (iii) What do *I* gain from knowing *p* instead of merely truly (and justifiedly) believing *p*? (I shall subsequently skip the ‘justified’ in parentheses.)

Our title question is presumably understood as (ii) and (iii) in the two perspectives and not as (i). This is why (i) may sound unexpected or even beside the issue. Clearly, (iii) is the most relevant question. In the end I want to know what the surplus value of knowledge is *for me*. I need an answer within the first-person perspective, if I am to have any motive for seeking knowledge beyond true belief. Is it useful to study question (ii) as well? If the benefit found by the ascriber is accessible to the knower as well, then an answer to (ii) immediately translates into an answer to (iii). (ii) goes beyond (iii) only if this benefit can be found only from outside and not from inside. I don’t want to speculate what such a benefit might be. (Perhaps knowers have an evolutionary advantage over true believers?) Hence, I put question (ii) aside.

This still leaves us with question (i), which is, against first appearance, relevant as well. The ascriber, too, may benefit from the detection of knowledge. If by learning a knower’s

⁵Pace Greco (2011, p. 221), who interprets the passage in Williamson (2000) as pointing to a practical value of knowledge.

⁶He gives a kind of proof of his claim on p. 79: ‘If your cognitive faculties are in good order, the probability of your believing *p* tomorrow is greater conditional on your knowing *p* today than on your merely believing *p* truly today’. A probabilistic formalisation is added in a footnote. This argument will run into the difficulties specified here in footnote 25.

⁷See also Rott (2004) for more discussion of what stability might mean.

knowledge of p the ascriber himself should acquire knowledge of p , then the benefit for the ascriber is the same as that for the knower, and an answer to (iii) entails an answer to (i). However, the transfer of knowledge may not be so easy, and without a stance towards the transfer of knowledge, (i) is independent of (iii). I don't want to take such a stance here; the transfer of knowledge is a different issue.

Besides its independence, however, there is a methodological advantage in discussing (i). The first-person perspective is well known to harbour confusion. (i) may have a clearer answer, and then we may hope to transfer this answer to (iii), that is, to the case of self-ascription of knowledge. So, below I shall always pursue questions (i) and (iii) in parallel. The dialectic point of this move will become fully clear in Section 5.

So, this will be my strategy for approaching the first-person question (iii), which, to repeat, is the one we ultimately want to answer. Note, however, that in the first-person perspective, we are occupied with the difference between mere believing p and believing that one knows p . The third person can ascribe me knowledge of p . I can certainly do so as well. Still, I thereby only express to believe that I know p .⁸ Hence, it is this difference between 'I believe p ' and 'I believe I know p ' which I am interested in.⁹ This is the reason why this paper will move within belief-first epistemology.

The distinction of the perspectives already helps to clarify a potential misunderstanding of Plato's stability idea. You may lose knowledge, simply due to your unavoidable forgetfulness. You may also rationally give up your knowledge of p and come to believe $\neg p$; this is the possibility mentioned in the above quote of Williamson. Still, you can never revise p as an item of knowledge. That is, you can never change from knowing p to knowing $\neg p$. If you now know $\neg p$, you couldn't have known p before. In this sense, knowledge is stable. These assertions hold within the third-person perspective. However, this is not the intended sense of stability. It follows trivially from the factivity of knowledge.

In the first-person perspective this misunderstanding cannot arise. Today I believe I know p , and tomorrow I believe that I didn't know p or even that I now know $\neg p$. Prima facie, beliefs to know something are revisable in the same way as beliefs in general. If stability is to have substantial content, it should be able to explain that there is still some gradient of stability between believing that one knows p and merely believing p .

So, the plan of the paper is this: In Section 2 I will expand a bit on the importance of distinguishing between the third- and the first-person perspective. In Section 3 I shall introduce Nozick's sensitivity analysis and explain why it provides a poor answer to our question if combined with the Stalnaker/Lewis analysis of conditionals, which may be taken as kind of standard. This motivates me to introduce, in Section 4, the epistemic account of conditionals I have put forward in Spohn (2015) and in particular the so-called 'circumstances are such that' reading of conditionals. Here I prefer to call it the *fact-asserting* interpretation of conditionals. Section 5 then provides, as promised, the explication of Williamson's quote, that is, of the idea that the surplus value of knowledge consists in a specific kind of stability. Section 6 concludes with a few comparative remarks.

2 | THIRD-PERSON AND FIRST-PERSON PERSPECTIVE

To get attuned to our issue, let's briefly discuss two examples how *not* to explain a surplus value of knowledge and observe how possible answers depend on the perspective. Let me briefly justify our assumption to exclusively deal with epistemically rational persons. This seems legitimate, since we are engaged in normative epistemology, which tries to say how things are within

⁸If we accept the knowledge norm of assertion, then, of course, I express knowledge of knowing something when ascribing knowledge to myself. However, we should not rely here on this contested norm. And we need not assume more than that assertion expresses belief.

⁹This sounds as if the difference would be one between a first-order and a second-order epistemic attitude. I would like to avoid this association. The German phrase 'Ich glaube zu wissen, dass p ' sounds less second-order and more like a single, not composite attitude towards p . In any case, I will later analyse this phrase as a first-order attitude.

the normative ideal. And certainly, epistemic rationality, whatever it is precisely, belongs to the normative ideal; we shall assume more specific things later on. However, it seems clear that rational persons only hold justified beliefs. This is not specific, either, since the notion of justification is quite unclear in turn. Still, it means that justification is used here as an internalistic notion, since rationality is supposed to be so as well. But this is how justification was understood in the JTB analysis all along. Otherwise, amendments of the JTB analysis might have been superfluous. In any case, the Meno problem holds for (ideally) rational persons as well. It is not a problem generated by human deficiencies.

So, let's first attend to the JTB analysis itself. Its inability to account for the surplus value directly shows in both the third-person and the first-person perspectives. When I ascribe you a JTB, then I think that you are a rational believer—but we can take this for granted now—I ascribe you a belief, and I share it. Knowledge ascriptions, one should think, go beyond this. Likewise within the first-person perspective. Here, the JTB analysis is unable to distinguish between believing p and believing that one knows p . I take my beliefs to be true—this is what it means to believe—and I rationally take all my beliefs to be justified. Hence, I can presently not distinguish between belief and knowledge from inside—if knowledge were just JTB. However, the difference between believing that one knows p and merely believing p clearly exists. I presently believe that I know that the sun was shining yesterday, but I only believe that it will shine tomorrow and don't believe that I know this.

Secondly, let's look at how a reliabilist conception might explain the surplus value of knowledge. Within the third-person perspective, the reliabilist has an easy and plausible answer. If I ascribe you knowledge of p in the sense of your true belief in p being produced by some (particularly) reliable process of belief formation (in contrast to mere belief in p), the value for me is obvious: Then I do not merely state that you share my belief. Rather, the more reliable your belief formation, the more I myself can rely on it. Your knowing p is thus much stronger support for my belief in p than your merely believing p . This is quite a satisfactory answer to question (i) from the third-person perspective.

However, things don't look so good from the first-person perspective, which, as said, is the one of proper interest. The reliabilist can easily reproduce his answer within this perspective. Obviously, I prefer beliefs which I take to be produced by some particularly reliable process and thus to constitute knowledge. There is a problem, though: How can I distinguish reliable from less reliable belief formation from inside? Only by having more or less, and more or less strong, reasons or evidence for my beliefs. This includes reasons concerning various ways of receiving evidence; for example, I can rationally learn that evidence is the more trustworthy, the better the light conditions of my observation. The evidence may come directly through my senses, or it may come from my fellow humans. It also includes higher-order evidence that may concern the ways of processing ordinary evidence. However this reasoning machinery for processing various kinds of evidence works in detail, in the end it is just about acquiring more and stronger reasons in order to have more and firmer opinions about the issue at hand. This means, however, that within the internal perspective, the reliabilist's answer reduces to the platitude that more certain or better justified beliefs are better than less certain or less well justified beliefs. There is nothing specific about believing that one has knowledge. Surely, we must guard our guiding stability idea from reducing to the same platitude. The lesser vulnerability to undermining evidence must not simply mean better reasons or higher certainty. I shall try to give a more informative description.

To emphasise the point: Goldman and Olsson (2009) have offered an explanation of the surplus value of knowledge from a reliabilist perspective: 'under reliabilism, the probability of having more true belief (of a similar kind) in the future is greater conditional on S 's *knowing* that p than conditional on S 's *merely truly believing* that p ' (p. 28, their italics). They add that probability is to be understood here in some objective sense. Thus, the knowledge ascriber is to gain from his ascription; he can then rely more on S 's further belief formation. This is certainly

plausible (see also the additional defence of Koscholke, 2021). However, this explanation runs directly into the above dilemma. So far, this states a value for the ascriber, as is perhaps not surprising, since the reliabilist account of knowledge expressly takes the third-person perspective. But what does *S* herself make of it? For *S*, true belief reduces to belief, mere belief is still belief with reasons (since *S* is rational), and *S*'s belief that her belief in *p* is reliably produced then reduces to *S* having more or stronger reasons for her belief in *p* than in the case of 'mere' belief in *p*. Clearly, this is valuable regarding *p*, and then it is valuable also regarding future beliefs that are located in a similar evidential situation. However, it comes to the same platitude as above.

3 | NOZICK'S SENSITIVITY ANALYSIS OF KNOWLEDGE

We said that the surplus value of knowledge depends on the surplus characteristic of knowledge over JTB. I propose to pursue our issue on the basis of the sensitivity analysis of Nozick (1981, sect. 3.I)—because it is one of the main contenders within Gettierology, because we may take it as a paradigm of the aforementioned family of modal analyses of knowledge, and because I am sympathetic to it; I will build on it later on. According to it, *x* knows that *p* iff the following conditions hold (in which I use obvious formalisations):

- (JTB) *p*, and *x* believes *p*, justifiably (but this is pleonastic now), that is, $B_x(p)$,
 (1) if *p* had not been the case, *x* would not have believed *p*, that is, $\neg p \triangleright \neg B_x(p)$, and
 (2) since *p* is the case, *x* believes *p*, that is, $p \triangleright B_x(p)$.

For the time being, I use \triangleright for the conditional in an indeterminate way. $p \triangleright q$ may be read as 'if *p*, *q*', but this only means that \triangleright is just as ambiguous as the ordinary language conditional. In the context of (1), \triangleright is intended as a counterfactual conditional. In the context of (2), by contrast, \triangleright is something like a factual conditional, which I have expressed here by the colloquial 'since'. In other words, the fact *p* is not only (1) a necessary, but also (2) a sufficient condition of the belief in *p*. Obviously, the resulting account of knowledge crucially depends on the interpretation of \triangleright . In any case, the intuitive idea here is that our beliefs, if they are to be knowledge, must be responsive or sensitive to the actual world in some specific way.

Nozick refrains from nailing down the counterfactual to a causal reading.¹⁰ This is why I have used 'since' in (2) instead of the more clearly causally loaded 'because'. When one reads (1) and (2) causally, they resemble the causal analysis of Goldman (1967). However, Goldman, too, refrains from outright requiring that the belief in *p* must be caused by the fact *p* itself in order to count as knowledge of *p*. Knowledge of *p* may also be had when the fact *p* and the belief in *p* are causally connected by a common cause in an appropriate way.¹¹ I would be happy with a strictly causal reading.¹² However, for the sake of greater generality, we may leave this open, even at the cost of somewhat greater indeterminacy.

In any case, the sensitivity analysis is just as clear as its use of the conditional. This holds *mutatis mutandis* also for the safety and the normality analysis mentioned above, which make

¹⁰An important source of his reservation is mathematical and ethical knowledge, to which a causal analysis of knowledge can obviously not apply. Clearly, these are very different kinds of knowledge (if we grant ethical knowledge at all). I am dealing here only with empirical knowledge.

¹¹See Goldman (1967, p. 369). This is a dangerous move, though. Given Reichenbach's common cause principle, it reduces knowledge to a positive correlation between fact and belief. But then all our true predictions, if they are more than random guesswork, would be knowledge. This would be too liberal. I am not sure whether the reference to 'appropriate ways' blocks this objection.

¹²The touchstone of a strictly causal reading is, of course, its denial of knowledge of the future, since a cause must precede (or be simultaneous with) its effect. Nozick (1981) does not address this issue. Goldman (1967, pp. 364f.) is perfectly aware of it and wants to allow knowledge of the future. However, he simply begs us to grant its possibility and discusses 'a reasonable candidate' for it. I have not found good arguments in favour of this possibility; it is apparently left to intuition.

crucial use of the conditional in their own ways. This is an unhappy dependence, at least from the point of view of formal epistemology. We know that the conditional idiom is fundamental, ubiquitous and most variegated. And we know that the philosophical and linguistic literature has produced a great variety of formal accounts in the last 50 years, none of which enjoys resounding acceptance. The fear is that this predicament radiates to the derivative account of the surplus value of knowledge. Therefore, let me focus on this dependence and not on the rich discussion of other pros and cons of the modal analyses.

So, let's look at the standard Stalnaker/Lewis account of truth conditions of counterfactuals (see Lewis, 1973, p. 16), which is probably the most familiar one. Nozick also refers to it, though in a non-committal way. According to it, (1) is true in the actual world @ iff $\neg B_x(p)$ is true in all $\neg p$ -worlds closest or most similar to @. And (2) is true iff $B_x(p)$ is true in all p -worlds closest or most similar to @. If we assume *centering*, that is, only @ is closest to @, then (2) reduces to (JTB). If only *weak centering* holds, that is, at least @ is closest to @, then (2) asserts that (JTB) holds indeed in all closest p -worlds. So, the belief in p is, as it were, not accidentally held in @. And (1) says that if we minimally change @ to falsify p , $B_x(p)$ ceases to hold as well. That is, in all worlds at least as close to @ as the closest $\neg p$ -worlds, the belief in p , if held, is true. So, the belief in p not only happens to be true in @ according to (JTB); there is a certain guarantee of its truth.

However, the similarity or closeness of worlds on which this account builds is of little help in understanding conditionals. Basically, there are two attitudes towards this similarity. One can either take it as a dummy (then closeness is the more neutral term), the formal properties of which are only to systematise all the conditionals we find acceptable.¹³ But then similarity does not further our understanding of conditionals. Rather, we are left with our intuitive understanding of them, on which it is too shaky to build philosophical theories. Or we take the talk of similarity seriously, as Lewis (1979) has done perhaps most prominently by stating four rules of thumb which govern our similarity judgements about worlds and hence our acceptance of counterfactuals. However, these rules seem entirely useless for our present purposes. How many and how big miracles and how large a variation of particular fact does it require for $B_x(p)$ to hold and p to fail or vice versa (where, to recall, Lewis takes a miracle to be a violation of natural laws, which he takes to be exceptionless, not *ceteris paribus* laws)? A hopeless question. The standard account of conditionals does not help much in understanding knowledge via (1) and (2).

This analytic deficiency extends to our issue of the surplus value; this is why I am discussing this here. If the modal analyses are understood from the third-person perspective, the truth guarantee given by (1) and (2) seems useful to the knowledge ascriber. The knower truly believes p in all closest p -worlds, and indeed her belief, if held, must be true in all worlds at least as close as all $\neg p$ -worlds. One may well interpret this as support of the ascriber's own belief in p . Still, this is quite a vague description of the surplus value.

Things don't get better in the more relevant first-person perspective. The fact that the knower believes that she knows p now translates into the fact that she believes that p , that, according to (2), she truly holds this belief in all closest p -worlds, and that, according to (1), this belief would be true in all worlds at least as close as the $\neg p$ -worlds. So, roughly, she thereby believes that the world could differ quite a bit from what it actually is and her belief in p would still be true. This kind of assurance, one might go on to say, has a certain epistemic value. Again, this does not sound wrong, but quite obscure.

Why are our findings so far so unsatisfactory? I think this is due to the fact that we have not yet done anything to spell out the interaction between the similarity or closeness relation

¹³I find clues in Lewis (1973, sect. 4.2) for this interpretation, when he speculates about the covariation of the vagueness of conditionals and the vagueness of similarity. However, in the end he tends to the realistic attitude explicated in Lewis (1979).

between worlds and the epistemic dealings of the knower. We need to focus on this interaction, if we want to better understand the epistemic surplus value of knowledge.

4 | INTERPRETING THE CONDITIONAL EPISTEMICALLY

How can we get a hold on this interaction? The epistemology of realistically interpreted conditionals, like that of modality in general, is quite a dim affair. In my view, we should directly turn to epistemic interpretations of the conditionals. Adams (1965, 1975) was the first to provide such an interpretation in probabilistic terms. However, we have excluded Bayesian approaches to our issue due to their problems with representing belief. The most obvious candidate for our concerns is belief revision theory as first and still paradigmatically displayed in book length in Gärdenfors (1988). However, for our purposes it is advisable to immediately turn to ranking theory, which is richer in expressiveness than belief revision theory, in ways needed below. Let me briefly introduce it. We can dispense with the formal machinery; some basic facts suffice for our purposes.

For us, the most useful basic notion is that of a so-called two-sided ranking function ρ_x of the subject x (see Spohn, 2015, p. 10; there I used the letter τ). ρ_x is defined on an algebra of propositions, and its range is the set of integers. $\rho_x(p) > 0$ expresses that x believes that p , that is, that x takes p to be true; that is, it defines $B_x(p)$. $\rho_x(p) < 0$ expresses x 's disbelief in p , that is, that x takes p to be false, and $\rho_x(p) = 0$ represents suspense of judgement or unopinionatedness.¹⁴ Belief thus explicated satisfies the standard axioms of doxastic logic; that is, beliefs are consistent and deductively closed. At the same time, ρ_x expresses degrees of belief; the larger $\rho_x(p)$, the stronger the belief or the weaker the disbelief in p .

A crucial point is that conditional ranks are defined as well.¹⁵ Thus, $\rho_x(q | p)$ represents x 's degree of belief in q given p , and $\rho_x(q | p) > 0$ expresses that x believes q conditional on p ; we abbreviate this by $B_x(q | p)$. Similarly for conditional disbelief and conditional suspense. Given the workings of ranking theory, conditional belief thus explicated satisfies all eight standard axioms of belief revision according to belief revision theory (see, e.g., Gärdenfors & Rott, 1995, pp. 53f). In particular, conditional belief satisfies *rational monotony*: $\neg B_x(\neg q | p) \rightarrow (B_x(r | p \wedge q) \leftrightarrow B_x(q \rightarrow r | p))$.¹⁶ This law is all we will need below. We may dispense with more deeply engaging in the mathematics of ranking theory.

The universally accepted basic idea for an epistemic interpretation of the conditional is the Ramsey test. It roughly says that I accept the conditional $p \triangleright q$ if I would accept q under the supposition p . Ranking theory translates the Ramsey test simply into conditional belief; by asserting ' $p \triangleright q$ ', x expresses his conditional belief $B_x(q | p)$. We must refrain from saying that x thereby also expresses an unconditional belief $B_x(p \triangleright q)$ in a conditional proposition. This is known to be a philosophically and formally highly problematic step. What are conditional propositions? How to avoid the ensuing triviality or impossibility results?¹⁷ We better avoid all these issues by talking only about conditional belief and not about conditional propositions.

However, we should amend the Ramsey test. Usually, by asserting a conditional $p \triangleright q$, one expresses a positive relevance of p for q . It is odd to assert the conditional if one thinks that

¹⁴There is no point here in engaging in the subtle differences between suspense of judgement and unopinionatedness or ignorance, although they certainly exist. Ranking theory also provides a stricter understanding of belief as $\rho_x(p) \geq z > 0$, paired with a more liberal understanding of unopinionatedness; see Spohn (2015, p. 9).

¹⁵See Spohn (2015, p. 10). I take this to be essential because it allows a dynamics of belief to be explained, just as probability kinematics is based on conditional probabilities.

¹⁶ \rightarrow represents material implication. Rational monotony corresponds to the axioms K^*7 and K^*8 for belief revision; see Gärdenfors and Rott (1995, p. 54). I shall accept it here, along with ranking theory, without further discussion. This is not the place to engage in the discussion about weaker alternatives.

¹⁷See, for example, Stalnaker (1984, ch. 7) for a paradigmatic discussion of conditional propositions and Rott (2011) for ways out of these trivialisation and impossibility results.

q holds, anyway. This is also Nozick's motive for requiring (1) and (2). He can do no better, though, than requiring that the antecedent is a necessary and sufficient condition for the consequent. Ranking theory provides a more general notion of positive relevance. p is positively relevant to q for x iff, for x , p is a *reason* for q (this is my preferred term), that is, if q is more credible or less incredible given p than given $\neg p$, that is, if $\rho_x(q | p) > \rho_x(q | \neg p)$. This includes the case where p is a sufficient condition or reason for q ($\rho_x(q | p) > 0 \geq \rho_x(q | \neg p)$) and the case where p is a necessary reason for q ($\rho_x(q | p) \geq 0 > \rho_x(q | \neg p)$), but it also allows, for example, for the case where p strengthens an already existing belief in q . In this case, p is an additional or supererogatory reason for q ($\rho_x(q | p) > \rho_x(q | \neg p) > 0$). Allowing this is welcome. So, by asserting $p \triangleright q$, x may as well express that p is a reason for q , that is, $\rho_x(q | p) > \rho_x(q | \neg p)$.¹⁸

In Spohn (2015), I have proposed still another variant of the Ramsey test, which I called the 'circumstances are such that' reading of conditionals and which I shall call here the *fact-asserting epistemic interpretation*; this is perhaps the more graphic term. It will be the one of which we make crucial use below. It is designed in particular for interpreting counterfactual or subjunctive conditionals. The idea is first articulated in Adams (1975, p. 131) who defines an 'expected assertibility value' for counterfactuals in order to treat them within his probabilistic semantics, a proposal which seems neglected in the literature with the exception of Skyrms (1981). My proposal is just a translation of Adams' idea into ranking theory, in the context of which it is more intelligible.

According to this interpretation, one makes a factual claim by asserting ' $p \triangleright q$ ', though it depends on one's epistemic state which factual claim this is. More precisely, one claims, that is, expresses one's belief, that conditions or circumstances are such that, given these conditions *and* the supposition p , one would believe q , and—if one adds the relevance idea—that conditions are such that p is positively relevant to q . I will call these conditions the *support* of the conditional. For instance, when I claim 'if I would drop the vase, it would break', I thereby express my unconditional belief in its support, that is, that the state of the world is such that I believe (more strongly) in the vase's breaking given such a state and my dropping of the vase (rather than my not dropping it). That is, I express my belief that none of the conditions obtains which in my view would impede the breaking of the vase, such as a soft ground, an unbreakable constitution of the vase, sufficiently weak or no gravitation, and so forth. This reading makes use of the idea underlying the Ramsey test and its amendment by relevance, but clearly goes beyond as well, insofar as there is no unconditional factual claim associated with the original Ramsey test (or its amendment by relevance). Indeed, according to this reading our dispute about a (counterfactual) conditional would be a purely factual dispute about the obtaining circumstances, provided we agree on the relevant conditional beliefs.

The possible circumstances alluded to cannot be entire possible worlds. If I assert ' $p \triangleright q$ ' as a counterfactual, I thereby express my belief in $\neg p$, but $\neg p$ must not belong to the circumstances conditional on which I believe q given p .¹⁹ Indeed, which kinds of circumstances are referred to depends on the context, as I have exemplified in Spohn (2015, pp. 16f.) with Quine's famous pair 'if Cesar would have been in command in the Korean war, he would have thrown atomic bombs' versus 'if Cesar would have been in command in the Korean war, he would have used catapults'. The first makes a factual claim about Cesar's psychological condition (which may be wrong; Cesar was perhaps not so ruthless a person), and the second makes a technological claim about Cesar's times (which may be wrong again; perhaps, unbeknownst to us, the Romans already had sort of tanks).

There on pp. 20f., I mention the 'history is such that' reading as a special case and argue that it is the appropriate reading for causal conditionals. When I assert 'if p had not happened, q would not have happened, either' and intend to convey that p is a (necessary) direct cause of

¹⁸For a complete logic of the conditional expressing the sufficient reason relation, see Raidl (2021, sect. 7).

¹⁹This is the original problem of cotenability of Goodman (1947), which initiated the modern exploration of counterfactuals.

q , then I thereby express my belief that the history of q without p is such that I believe $\neg q$ given this history only under the additional condition $\neg p$; this includes the relevance of p to q . That is, the possible circumstances are now possible histories (up to q without p).

To vary the above example: I claim that if I had dropped the vase, it would have broken, thereby expressing my belief that my clutching the vase was a (necessary) cause of its remaining intact. This is now interpreted as the belief that, according to my conditional beliefs, nothing in the actual course of events up to the time where the vase is still intact would have prevented the breaking of the vase, given I had dropped it.

Such a causal reading may be the preferred one for Nozick's sensitivity analysis. However, as mentioned, Nozick himself is not quite determinate about this. So, let's stick to talking of (relevant) conditions or circumstances instead of histories. This is vaguer, but might accommodate those who favour a more liberal reading of conditionals. For us, the crucial point is that the fact-asserting interpretation clarifies which factual claim is associated with a conditional.

Other accounts of conditionals don't clearly specify such a factual claim. This is why I am emphasising the fact-asserting interpretation. The original Ramsey test obviously does not do so at all. Next, if a conditional sentence is taken to assert a conditional proposition, then the claim is a modal, not a factual one. This is particularly clear in the Stalnaker/Lewis approach where a conditional proposition (the truth condition of the conditional sentence) says that the actual world is such that the similarity spheres around it have such and such a structure. If the structure of similarity spheres supervenes on the facts in the actual world, as claimed by Lewis in his program of Humean supervenience, then indeed a conditional proposition would reduce to a factual proposition. Taking this line would, however, commit us to a (too?) strong philosophical thesis and would still not specify the associated factual proposition, because the supervenience relation is hardly specified. (Above, we noted already the uselessness of Lewis' (1979) four rules of thumb for our present purposes.) Similarly, Edgington (1995, p. 283) speaks of, but only vaguely characterises what she calls the (factual) basis of a conditional.

Let us state this fact-asserting interpretation a bit more formally. We refer to the set $C = \{c_1, \dots, c_n\}$ of possible conditions or circumstances relevant to a given conditional. C may depend on the context of utterance. The conditions in C are maximally specific, as far as required, and they are mutually exclusive and jointly exhaustive, that is, exactly one of them must obtain. Let person x assert the conditional ' $p \triangleright q$ '. According to the Ramsey test, x would thereby simply express his conditional belief $B_x(q | p)$, that is, $\rho_x(q | p) > 0$. According to the amended Ramsey test, we add the relevance idea, and then x moreover expresses that p is a reason for q , that is, $\rho_x(q | p) > \rho_x(q | \neg p)$.

According to the fact-asserting interpretation, by contrast, x expresses his belief that one, that is, the disjunction, of those conditions obtains under which he can maintain these conditional beliefs and ranks. This disjunction is the *support*²⁰ of the conditional. That is, he expresses his belief $B_x(c)$, where the support is $c = \bigvee \{c_i \in C \mid B_x(q | p \wedge c_i) \wedge \rho_x(q | p \wedge c_i) > \rho_x(q | \neg p \wedge c_i)\}$.²¹ This entails $B_x(q | p \wedge c)$.²² One might expect that this and $B_x(c)$ entail $B_x(q | p)$. However, this is not the case. Rather, $B_x(q | p \wedge c)$ and $B_x(c | p)$ entail $B_x(q | p)$.²³ And $B_x(c)$ and $\neg B_x(\neg p)$ entail the required premise $B_x(c | p)$.²⁴ In other words, the inference from $B_x(c)$ and $B_x(q | p \wedge c)$ to $B_x(q | p)$ may fail only if $B_x(\neg p)$. This means that the fact-asserting interpretation entails the Ramsey test, given the additional premise $\neg B_x(\neg p)$. If, however, the supposition p is belief-contravening (which is the case when ' $p \triangleright q$ ' is a counterfactual), the Ramsey test and the fact-asserting interpretation may fall apart.

²⁰This is, arguably, Edgington's basis.

²¹We build here the fact-asserting interpretation on our amended Ramsey test. This is the one we will need below. Of course, we could also build it, for example, on the original Ramsey test. See Spohn (2015) for the many possible ramifications.

²²See assertion (14) of Spohn (2015, p. 17).

²³If the premise $B_x(c | p)$ is weakened to $\neg B_x(\neg c | p)$, the inference is an instance of rational monotony.

²⁴Again, this is a direct instance of rational monotony.

There is a remarkable stability in the assertion of conditionals under the fact-asserting interpretation that is not found in Ramsey test conditionals. In general, x may reverse his (conditional) beliefs arbitrarily many times. That is, we may have $B_x(q | p)$, $B_x(\neg q | p \wedge r_1)$, $B_x(q | p \wedge r_1 \wedge r_2)$, and so forth. Thus, x would start asserting ' $p \triangleright q$ ' according to the Ramsey test, but would claim ' $p \triangleright \neg q$ ' after learning r_1 by conditionalisation and return to ' $p \triangleright q$ ' after additionally learning r_2 by conditionalisation, and so forth. These pieces r_i of information may well be jointly logically compatible with p ; it does not contradict p to receive all of them. And they may move within the context given by the set C of conditions. For instance: If I walk outside, I will get wet (because it rains). What if my path is roofed? But the roof is holey! Still, I have my umbrella. And so on. In this way, the conclusion about getting wet may continue vacillating. This instability of our beliefs is our everyday experience, even though we are reluctant to change our minds too often. This observation, also known as Sobel sequences, corresponds to Lewis' (1973, sect. 1.3) observation of the variable strictness of the subjunctive conditional. It is, moreover, the qualitative analogue of Simpson's paradox haunting probabilistic inference.

By contrast, x 's assertion of ' $p \triangleright q$ ' is much stabler under the fact-asserting interpretation. As said, x thereby expresses his belief $B_x(c)$ in the support c of the conditional along with all the conditional beliefs and rank comparisons defining this support. And nothing of this changes as long as x receives information about the conditions in C that is *logically compatible* with the support c . That is, after learning such an information r_1 by conditionalisation, the conjunct r_1 is added to the conditions of those conditional beliefs and rank comparisons defining the support. But the addition is redundant, and those conditional beliefs and rank comparisons continue to hold for all conditions $c_i \in C$ entailing $c \wedge r_1$. Thus, under the fact-asserting interpretation, ' $p \triangleright q$ ' can still be asserted with the support of $c \wedge r_1$. In short, the reason for this stability is that we do not refer to unstable conditional beliefs of the form $B_x(q | p)$, as the Ramsey test does, but to the stabler conditional beliefs of the form $B_x(q | p \wedge c_i)$ given maximally specific conditions.

Does this observation mean that there are no Sobel sequences within the fact-asserting interpretation? No. It still allows asserting ' $p \triangleright q$ ', changing to ' $p \triangleright \neg q$ ' after supposing r_1 or learning r_1 by conditionalisation, and so forth. But this is only possible when the supposition or information r_1 disproves the support c of the initial assertion ' $p \triangleright q$ '. And this must be allowed too, of course; the support may turn out wrong. The gain in stability of conditionals under the fact-asserting interpretation is not unshakeable, but it exists. I will elaborate on the significance of this observation below.

5 | THE SURPLUS VALUE OF KNOWLEDGE

I argued that we must attend to epistemic interpretations of the conditional if we want to assert a surplus value of knowledge proceeding from Nozick's sensitivity analysis (or other modal accounts of knowledge). I have thus presented three such interpretations: the Ramsey test, its relevance amendment, and the fact-asserting interpretation. Do they really help?

Well, the Ramsey test as such is not so productive. If we follow this test, then, within the third-person perspective, the knowledge ascriber x expresses (3) by claiming (JTB), (1) and (2) about the knower z :

- (3) (a) $B_x(p)$, (b) $B_x(B_z(p))$, (c) $B_x(\neg B_z(p) | \neg p)$, and (d) $B_x(B_z(p) | p)$.

(3a) and (3b) correspond to (JTB), and (3c) and (3d) to (1) and (2). (d) runs empty, because it is entailed by (a) and (b); under the supposition of something believed, the beliefs just stay the same. (c) helps a little. It says: under the supposition $\neg p$ the ascriber believes that the knower would not believe p . One might interpret this as additional assurance for the ascriber's own belief in p through the knower's belief in p .

Things look worse in the first-person perspective, where we identify z with x . Indeed, we meet a crucial problem for our strategy of approaching our title question via epistemic interpretations of the conditionals (1) and (2). (3c) raises the question, if taken literally: Assuming $\neg p$, what should I (x) believe about $B_x(p)$? $\neg B_x(p)$, of course, and indeed $B_x(\neg p)$. But this seems to be the very same question as: Assuming $\neg p$, what should I believe about p ? $\neg p$, of course. The Ramsey test thus seems to reduce either to the truth $B_x(\neg p) \rightarrow B_x(\neg p)$ of propositional logic or to the assertion $B_x(\neg p) \rightarrow B_x(\neg B_x(p))$, which follows from positive introspection $B_x(q) \rightarrow B_x(B_x(q))$. (3d) seems true for the same trivial reason. It seems that we must stay away from iterations of the belief operator with respect to the same subject; the Ramsey test is not made for such auto-epistemological applications and does not make sense there.

However, if so, our strategy breaks down. It at best produces trivialities instead of an answer to question (iii) from the first-person perspective.²⁵ What to do? Well, the above trivialities emerge only if we refer my (x 's) auto-epistemological reflections to the present moment. Therefore, I suggest to understand these reflections as being about (slightly) in the past or in the future. Thus, my self-ascription of knowledge of p is now taken to express that I believe that I knew, or will know, p a little time from now. That is, I express:

$$(3') \text{ (a) } B_x(p), \text{ (b) } B_x(B'_x(p)), \text{ (c) } B_x(\neg B'_x(p) \mid \neg p), \text{ and (d) } B_x(B'_x(p) \mid p).$$

Here B'_x describes x 's beliefs at a slightly different time than B_x . Now, (3'c) and (3'd) make sense. And they make sense, even if I assume to receive no evidence relevant to p between the two times. I may simply be a bit uncertain about whether I believed, or will believe, p . And then I might or might not have the conditional beliefs (3'c) and (3'd). In short, my suggestion is to turn question (iii) by the little time shift into question (i) from the third-person perspective. My past or future self resembles a third person from my present point of view. Thus, we can directly transfer our answer from the third-person to the first-person perspective. This is the thrust of treating question (i) as superordinate to question (iii). The answer to question (iii) provided by the Ramsey test then is still as unsatisfactory as it was to question (i). At least, though, we have found a way of avoiding auto-epistemological trivialities.

So, what about the relevance amendment of the Ramsey test? Now the knowledge ascriber x expresses (4) by claiming (JTB), (1) and (2) about z :

$$(4) \text{ (a) } B_x(p), \text{ (b) } B_x(B_z(p)), \text{ and (c) } \rho_x(B_z(p) \mid p) > \rho_x(B_z(p) \mid \neg p).$$

(4c) summarises Nozick's (1) and (2). It indeed improves upon them a bit. According to (1) and (2), x takes p to be a necessary and sufficient condition of $B_z(p)$, as manifested in x 's conditional beliefs. However, we mentioned that positive relevance of p for $B_z(p)$ also allows p to be an additional reason for $B_z(p)$. In this case, (1) would be wrong. In acquiring knowledge of p , one need not thereby acquire the belief in p . One may have believed p before. For example, you believe that it's raining tomorrow (because you believe the weather forecasts), but tomorrow you see and thus come to know that it's raining. So, as an ascriber I believe that you believe it to rain tomorrow, anyway, whether or not it is actually going to rain. (4c) accommodates this case.²⁶

Unfortunately, (4) does not change much regarding the surplus value of knowledge. Note that positive relevance is always symmetric; that is, (4c) is equivalent to $\rho_x(p \mid B_z(p)) > \rho_x(p \mid \neg B_z(p))$. Hence, (4c) represents that z 's belief in p is a reason for x to himself believe in p . This was at most suggested by (3c), whence (3) was unsatisfactory. It is pleasing that (4c) makes

²⁵Note that the quote of Williamson in footnote 6 runs into the same problem, if the probabilities referred to are the knower's own subjective probabilities. If we interpret them as a kind of objective probabilities, then the quote does not offer an answer to question (iii).

²⁶Nozick (1981, pp. 179ff.) worries about such cases as well. However, Nozick is bound to stick to (1) and (2), because he does not talk of degrees of belief as ranking theory does. Rather, he tries to solve them by alluding to various methods of acquiring knowledge.

this point explicit. However, (4c) does not do more. This makes clear that (4) hardly goes beyond the mere ascription of true belief (the subjective justifiedness is still taken for granted). We generally have the defeasible presumption that our fellow humans' beliefs are true (but surely, we know all the caveats and weaknesses of this presumption). And we generally take the agreement of others as some confirmation of our opinions. (4c) only says that this general presumption is taken by x to apply in the case at hand. There is nothing distinctive in (4) concerning z 's knowledge of p .

Within the first-person perspective (where $z = x$), we again have the problem that (4c) seems to be an auto-epistemological triviality, if it makes sense at all. We should eschew auto-epistemological applications of ranking theory, too. However, the same move as above avoids this predicament. Referring the knowledge self-ascription to a slightly different time, it expresses:

$$(4') \text{ (a) } B_x(p), \text{ (b) } B_x(B'_x(p)), \text{ and (c) } \rho_x(B'_x(p) | p) > \rho_x(B'_x(p) | \neg p).$$

This move turns the instantaneous auto-epistemological certainty that I take my present beliefs to be true into the defeasible presumption that my past or future beliefs are true. Thus, (4'c) says that I take $B'_x(p)$ to be a reason for p . This sounds eminently reasonable. However, the objection above applies here as well. (4') does not specifically address my knowledge of p .

As announced, things finally change for the better with the fact-asserting interpretation of conditionals. If x ascribes to z knowledge of p , x thereby claims (JTB), (1) and (2) and, according to this interpretation, expresses (5) relative to a given set C of possible conditions:

$$(5) \text{ (a) } B_x(p), \text{ (b) } B_x(B_z(p)), \text{ and (c) } B_x(b), \\ \text{where } b = \bigvee \{c_i \in C \mid B_x(B_z(p) | p \wedge c_i) \wedge \rho_x(B_z(p) | p \wedge c_i) > \rho_z(B_z(p) | \neg p \wedge c_i)\}.$$

In general, we spoke of the support of a conditional under the fact-asserting interpretation. Now it is perhaps more appropriate to call b the pertinent *knowledge background*, or perhaps more correctly, the *knowledge ascription background*. This knowledge background b is a factual proposition which is true or false, even though it is defined in a subject-relative way, that is, defined relative to the conditional beliefs of x . The ascriber x believes in the background b , but he may be wrong, just as his knowledge ascription may be wrong.

The crucial point is that (5) allows me to redeem my announcement to give an informative explication of the guiding stability idea taken from Plato and the quote from Williamson. Let us first look at the kind of stability that automatically accrues to belief. According to the pragmatic theory of truth, we will believe the truth in Peirce's extremely counterfactual limit of inquiry.²⁷ Hence, when I now believe that p , that is, that p is true, I believe that I will forever keep believing p in that limit of inquiry.²⁸ In this sense, I believe my belief to be stable. Let's call this the stability *appendant* to each belief. It is, however, our everyday experience that this appendant stability of beliefs, which is only believed, is spurious. The to and fro schematically described above may occur any time. Actually, many beliefs are not stable. Having found reason to give up my belief in p , I believe that this belief may or may not return. And having found reason to even believe $\neg p$, then I also believe in the appendant stability of the latter belief. Indeed, such instability may not only actually occur; I can conceive of this instability already now. That is, I can specify many possible reasons against my belief in p . It is only that I do not possess any of these reasons, and I deny that any of these reasons will obtain.

²⁷The interesting interpretation of the pragmatic theory is the reverse one: that truth is what we believe in the limit of inquiry. This is not our present topic, though.

²⁸Of course, I am sure I will die and mankind will become extinct before reaching the limit. As said, the limit of inquiry is extremely counterfactual in various ways.

After this preliminary observation, let us see what is added by belief and knowledge ascriptions. Again, let us take up the third-person perspective first and put me into the role of the ascriber. So, I (x) ascribe the belief in p to some subject z (perhaps because she asserts that p). As said, the presumption is that this is a reason for me to believe p by myself. Or I may already independently believe that p and thus ascribe her a true belief, which simply expresses agreement from my side. Either way, I believe in p and thus in the appendant stability of my belief in p . But does it help with the actual instability of this belief? Well, due to this presumption, her belief in p is a (weak) reason for my belief in p , and an additional one in case I had this belief before. This is the situation represented by (4). It may make it a bit harder to give up my belief in p . However, this is a very weak stabilising factor. Otherwise, the instability is the same as before. When I find reasons to give up or even deny p , I usually do not reject those reasons with reference to her believing that p . Rather, those reasons will usually work against the presumption that her belief in p is true.

Does the situation change when I ascribe knowledge of p to z ? Yes. Again, I thereby express agreement with her. But I also believe in the knowledge ascription background b , which is defined by my conditional beliefs as specified. So, her belief in p is a reason to believe p myself conditional on the background b , and indeed also unconditionally.²⁹ In the case of my belief ascription, this reason for my belief in p may be, as just observed, defeated in any ordinary way, also according to my beliefs. However, in the case of my knowledge ascription, the situation is believed by me to be much more stable. That is, no evidence logically compatible with the background b , even if it speaks against this background or against p itself, can defeat this reason for my belief in p . In this sense, her belief in p is a stable reason for my own belief in p . This is the guarantee resulting from my knowledge ascription. It need not be a particularly strong guarantee. But it is more of a guarantee than was provided by (3) or (4). Surely, I may also receive evidence logically *refuting* the background b . Then, this set-up would indeed be defeated. My knowledge ascription may be proven wrong. But it is much more difficult to do so.

I should point out, though, that this guarantee needs a qualification. Strictly speaking, it depends on the granularity of the assumed possible conditions. These conditions could be more or less fine-grained, describing, for example, only possible Roman weapons or entire histories minutely described. The guarantee concerning evidence logically compatible with the background holds only for evidence representable by (and thus not more fine-grained than) those possible conditions. The stability of my reason for my belief in p does not extend to possible evidence not so representable. In other words, the guarantee is the stronger the more fine-grained the possible conditions referred to in the knowledge ascription are.

So far, I have discussed from the third-person perspective my benefit when I can ascribe knowledge to others. But what is the benefit in the case of self-ascription of knowledge? This is what we want to know. Alas, (5) again turns into trivialities, when we identify x and z . The background b , then, is the set C of all conditions, that is, the tautology, because all conditions c_i trivially satisfy the defining clause for b . Hence, for a third time, we need to take the move of referring the knowledge self-ascription to a slightly different time. That is, we let x ponder not about his present belief $B_x(p)$, but about his (slightly) earlier or later belief $B'_x(p)$; we need not write this out again as a clause (5').

Then, however, it is clear what the surplus value of knowledge is for me myself. It is the same as in the third-person ascription of knowledge. If I believe that I knew, or will know, that p , then my belief in p is more stable than my mere belief in p . Unlike the latter, my belief in p can then not be defeated by any evidence logically compatible with my knowledge background b . This gain in safety, or loss in revisability, makes for the epistemic surplus value. I must add, though, that this value is again relative to the granularity of the pertinent possible

²⁹We observed above that $B_x(b)$ and $B_x(p \mid b \wedge B_z(p))$ entail $B_x(p \mid B_z(p))$ given the further premise $\neg B_x(\neg B_z(p))$, and this additional premise is satisfied by the knowledge ascription.

conditions. The value is greater or lesser, depending on the conditions being more or less fine-grained.

Let me illustrate this account. It explicates, for example, the Larissa case. I stand at a cross-road. I can take one of three ways. I believe that the left one leads to Larissa, and so I take it. However, I start hesitating as soon as counter-evidence shows up. The path takes unexpected turns which seem to go in the opposite direction; there are confusing signs at the roadside; the path narrows to what appears to be a dead end; and so forth. And then I might lose my belief and turn back.

But now you are my guide, and I believe that you *know* that the left way leads to Larissa. You apparently believe so, you take the left way, I believe you, I follow you—and I even follow you when those irritating incidents occur. Why? My knowledge ascription background consists, for example, in my belief that you took the left way last week and arrived at Larissa (so you know the way), whatever the conditions unknown to me. So, whatever irritations come along the way, they are logically compatible with my knowledge background, and my belief that this is the way to Larissa remains unshattered. The case is different when you start hesitating or when I start suspecting that you want to lure me into a trap; then you apparently do no longer believe that this is the way to Larissa, or never did. Or you may tell me that it's actually the first time that you've taken this way; this would logically contradict my knowledge background.

The same applies in the case of the self-ascription of knowledge. I believe that I know that the left way is the one to Larissa. My knowledge background may be that I have taken this path last week and arrived at Larissa, or that you, whom I believe that you know the way, tell me: 'Take the left way!', and so forth, again whatever the conditions, the details of the path. And therefore, however these details turn out to be (I may have forgotten about them since I took the path last week), my belief that this is the way to Larissa is not defeated.

Or take the central case of knowledge by perception.³⁰ Basically, the causal theory of perception applies here. That is, a subject x sees or perceives that p if the fact p itself causes x 's belief in p in normal, non-deviant perceptual ways. Thus, x knows that p . The knowledge background then consists in the conditions required for this non-deviant causal relation: fair perceptual conditions, an undisturbed mind of the perceiver, the absence of fancy stories invented as examples of deviant causation, and so forth. The point then is that all further (negative) evidence about p is powerless against the belief in this background and the conditional belief in p given any possible conditions in this background. This is the kind of guarantee and the epistemological gain acquired by perception.

This is my 'proof' of my 'equation' Nozick + Spohn Williamson, that is, of my explication of the initial quote from Williamson (2000): 'Present knowledge is less vulnerable than mere present true belief to *rational* undermining by future evidence.' As intended, this account of the surplus value of knowledge concerns only our epistemic dealings with the world. Usually, though, it will be accompanied by a practical advantage. If I am less vulnerable having to revise my beliefs, I am also less vulnerable having to revise my decisions based on those beliefs. However, I shall not elaborate on this suggestion.³¹

Should we be impressed by this account? I am not sure. One may have hoped for more, for some stronger truth guarantees. However, this is what results when we translate Nozick's sensitivity analysis of knowledge (or other variants of the modal account of knowledge) into precise epistemological terms along the lines of the fact asserting epistemic interpretation of conditionals. If this result is found unsatisfactory, one may seek fault either with my epistemological reinterpretation or with the modal accounts themselves. But then a formal account of the surplus value of knowledge would still be wanting.

³⁰This is also the first core example of Goldman (1967, pp. 358f.).

³¹Another way of summarising this paper in one sentence would be to say that it is simply the application of my ideas about the connection between causal explanations and the stability of reasons in Spohn (1991) to the causal theory of knowledge employed here. However, there was no point in reintroducing these general ideas.

6 | SOME COMPARATIVE REMARKS

Let me add a few comparative remarks. Perhaps it is worth noting that the stability theory of belief of Leitgeb (2014) shares only the label. What Leitgeb calls stability of belief is rational monotony, that is, the preservation of beliefs under not disbelieved conditions and assumptions, and Leitgeb presents a solution to the problem how this stability goes together with a high (conditional) probability notion of (conditional) belief. Knowledge is not at issue there (and not claimed to be).

Lehrer (1965, 2000) may rightly be attributed to have developed a stability theory of knowledge. Knowledge must be more stable than belief, because it must survive the justification game for personal justification and the ultra justification game for undefeated justification. This might be taken as another explication of the quote from Williamson (2000). Lehrer presents his account in a semi-formal way. Rott (2004) reconstructs this account within AGM belief revision theory and employs this reconstruction for a response to the Meno problem. We cannot enter now a fuller discussion of this proposal. Let me only point to what I take to be the crucial difference to the present account.

In the ultra justification game an omniscient opponent is allowed to raise any *true* objections to my beliefs, and those of my beliefs that survive this game (and hence are stable in this sense) are my items of knowledge. This may be a severe test of my knowledge. It differs, though, from the account in Section 5, since there is no explicit reference to what I called the knowledge ascription background. And it distinguishes knowledge from belief from the perspective of a third person who already knows the truth. Such a person knows what she can object to and may thus be able to tell my knowledge from my mere beliefs. Perhaps it is soothing to know that some of my beliefs, those which are knowledge, are stable in this sense and will never be lost (because I will only get true information).³² However, from inside I cannot tell which ones these are. From inside, the distinction collapses. If asked, which of my beliefs I believe to survive the ultra justification game, I can only respond: all of them. If the opponent only raises objections which I take to be true, I will never have to give up any of my beliefs; and I certainly believe that he will not raise objections which I take to be false. This is just the spurious appendant stability of beliefs.

The stability explicated here is a different one. It is a stability of (conditional) belief under all possible information about the conditions within the knowledge background, whether true or false. This stability is provided precisely in view of my partial ignorance of the conditions, and this is why it confers value to those of my beliefs which I take to be stable in this more pertinent sense and thus to be knowledge.

ACKNOWLEDGEMENTS

I am indebted to two reviewers whose comments helped considerably improve the paper. I gratefully acknowledge support from the German Research Foundation under Project No. EXC 2064/1 (Grant No. 390727645) and Project No. SP279/21-1 (Grant No. 20094936). Open Access funding enabled and organized by Projekt DEAL.

ORCID

Wolfgang Spohn  <https://orcid.org/0000-0002-3213-8907>

REFERENCES

- Adams, E.W. (1965) The logic of conditionals. *Inquiry*, 8, 166–197.
Adams, E.W. (1975) *The logic of conditionals*. Dordrecht: Reidel.
Edgington, D. (1995) On conditionals. *Mind*, 104, 235–327.

³²Note, however, that Rott (2004, sect. 8) himself sees problems with his explication for which he can't offer a solution.

- Freitag, W. (2013) *I know. Modal epistemology and scepticism*. Münster: Mentis.
- Gärdenfors, P. (1988) *Knowledge in flux. Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.
- Gärdenfors, P. & Rott, H. (1995) Belief revision. In: Gabbay, D.M., Hogger, C.J. & Robinson, J.A. (Eds.) *Handbook of logic in artificial intelligence and logic programming, vol. 4, Epistemic and temporal reasoning*. Oxford: Oxford University Press, pp. 35–132.
- Gettier, E.L. (1963) Is justified true belief knowledge? *Analysis*, 23, 121–123.
- Goldman, A.I. (1967) A causal theory of knowing. *Journal of Philosophy*, 64, 357–372.
- Goldman, A.I. & Olsson, E. (2009) Reliabilism and the value of knowledge. In: Haddock, A., Millar, A. & Pritchard, D.H. (Eds.) *Epistemic value*. Oxford: Oxford University Press, pp. 19–41.
- Good, I.J. (1966) On the principle of total evidence. *British Journal for the Philosophy of Science*, 17, 319–321.
- Goodman, N. (1947) The problem of counterfactual conditionals. *Journal of Philosophy*, 44, 113–128.
- Greco, J. (2011) The value problem. In: Bernecker, S. & Pritchard, D. (Eds.) *The Routledge companion to epistemology*. New York: Routledge, pp. 219–231.
- Hawthorne, J. & Stanley, J. (2008) Knowledge and action. *Journal of Philosophy*, 105, 571–590.
- Koschke, J. (2021) A new argument for Goldman and Olsson's solution to the extra-value-of-knowledge problem. *Theoria*, 87, 799–812.
- Lehrer, K. (1965) Knowledge, truth, and evidence. *Analysis*, 25, 168–175.
- Lehrer, K. (2000) *Theory of knowledge*, 2nd revised edition. Boulder, CO: Westview Press.
- Leitgeb, H. (2014) The stability theory of belief. *Philosophical Review*, 123, 131–171.
- Lewis, D. (1973) *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1979) Counterfactual dependence and time's arrow. *Noûs*, 13, 455–476.
- Nozick, R. (1981) *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Plantinga, A. (1993) *Warrant: the current debate*. Oxford: Oxford University Press.
- Plato, *Meno*. Transl by C. Woods (2012) Available at <https://doi.org/10.2139/ssrn.1910945>
- Pritchard, D. (2005) *Epistemic luck*. Oxford: Oxford University Press.
- Pritchard, D., Turri, J. & Carter, J.A. (2018) The value of knowledge. In: Zalta, E.N. (Ed.) *The Stanford encyclopedia of philosophy*, URL = <https://plato.stanford.edu/archives/spr2018/entries/knowledge-value/>
- Raidl, E. (2021) Definable conditionals. *Topoi*, 40, 87–105.
- Rott, H. (2004) Stability, strength and sensitivity: converting belief into knowledge. *Erkenntnis*, 61, 469–493.
- Rott, H. (2011) Reapproaching Ramsey: conditionals and iterated belief change in the spirit of AGM. *Journal of Philosophical Logic*, 40, 155–191.
- Savage, L.J. (1954) *The foundations of statistics*, 2nd edition. New York: Wiley. Dover, 1972.
- Schulz, M. (2017) Decisions and higher-order knowledge. *Noûs*, 51, 463–483.
- Skyrms, B. (1981) The prior propensity account of subjunctive conditionals. In: Harper, W.L., Stalnaker, R.C. & Pearce, G. (Eds.) *Ifs*. Dordrecht: Reidel, pp. 259–265.
- Skyrms, B. (1990) *The dynamics of rational deliberation*. Cambridge, MA: MIT Press.
- Sosa, E. (2000) Scepticism and contextualism. *Philosophical Issues*, 10, 1–18.
- Spohn, W. (1991) A reason for explanation: explanations provide stable reasons. In: Spohn, W., van Fraassen, B.C. & Skyrms, B. (Eds.) *Existence and explanation*. Dordrecht: Kluwer, pp. 165–196.
- Spohn, W. (2015) Conditionals: a unified ranking-theoretic perspective. *Philosophers' Imprints*, 15(1), 1–30.
- Stalnaker, R.C. (1984) *Inquiry*. Cambridge, MA: MIT Press.
- Williamson, T. (2000) *Knowledge and its limits*. Oxford: Oxford University Press.
- Williamson, T. (2005) Contextualism, subject-sensitive invariantism and knowledge of knowledge. *The Philosophical Quarterly*, 55, 213–235.

How to cite this article: Spohn, W. (2024) The surplus value of knowledge. *Theoria*, 90(2), 208–224. Available from: <https://doi.org/10.1111/theo.12521>