Stability, strength and sensitivity:
Converting belief into knowledge

Hans Rott
Department of Philosophy, University of Regensburg
93040 Regensburg, Germany

Abstract. In this paper I discuss the relation between various properties that have been regarded as important for determining whether or not a belief constitutes a piece of knowledge: its stability, strength and sensitivity to truth, as well as the strength of the epistemic position in which the subject is with respect to this belief. Attempts to explicate the relevant concepts more formally with the help of systems of spheres of possible worlds (à la Lewis and Grove) must take care to keep apart the very different roles that systems of spheres can play. Nozick's sensitivity account turns out to be closer to the stability analysis of knowledge (versions of which I identify in Plato, Descartes, Klein and Lehrer) than one might have suspected.

1. Introduction: Grades of knowledge and belief

Gettier has shattered our understanding of knowledge. There is still little agreement among philosophers what knowledge is. Stability theories (also known as defeasibility theories) say that knowledge is belief with a stable (indefeasible) justification. Nozick advanced an influential theory, according to which knowledge is belief that is sensitive to truth (or that "tracks truth"). The contextualist model of Keith DeRose explicates knowledge depends on how strong the subject's epistemic position is with respect to the belief in question.

It is well known that these proposals have difficulties in dealing with certain classes of counterexamples.[1] But my aim in this paper is not to confront the various theories with yet more and yet more complicated examples and counterexamples. I rather take it that they all capture important intuitions that can in some way or other be regarded as relevant to the question whether or not a given belief constitutes a piece of knowledge. This questions I am going to address are the following: Can stability, or more exactly, the stability of beliefs in an interrogation with a truthful critic like Socrates in a Platonic dialogue, be the right basis for the explication of knowledge? Does strength of belief imply stability, or vice versa? If knowledge lies in the stability of a belief, how does knowledge relate to the strength of the belief? How does the strength of the subject's epistemic position with respect to a belief as highlighted in the contextualist literature relate to the strength of the subject's belief? And finally, as the contextualist account is at least in part inspired by Nozick's truth-tracking or sensitivity account of knowledge: How does the *sensitivity of a belief for truth* relate to the strength of the belief?

In my attempt to answer these questions I shall make use of a possible worlds modelling for subjunctive conditionals going back to Lewis (1973) and referred to by Nozick and DeRose. I shall draw attention to the fact that the same formal model can be used for the analysis of the stability approach, but that the interpretation of this model must then be crucially different. It represents the subject's doxastic state and can be used to represent the changes that this state undergoes while a critic tries to undermine the subject's beliefs by advancing potential defeaters. The possible worlds model can thus be used to model both internal (subjective) and external (objective) aspects of knowledge. I close with a short overview of the relation between stability, strength, epistemic position and sensitivity.

2. The stable belief theory of knowledge

The first stability account of knowledge is probably to be found in Plato's *Meno*, where Socrates says that true beliefs convert into knowledge if and only if they become "permanent" after having been "tied down" by giving reasons for them.[2] A less widely known but similar formulation is given by Descartes in his second Replies who claims that true knowledge cannot be "rendered doubtful".[3]

[1] For critical discussions of stability theories, see Shope (1983, pp. 45–74) and of Nozick's theory, see the papers collected in Luper-Foy (1987).

[2] "True opinions too are a fine thing and altogether good in their effects so long as they stay with one, but they won't willingly stay long and instead run away from a person's soul, so they're not worth much until one ties them down by reasoning out the explanation. … And when they've been tied down, then for one thing they become items of knowledge (ἐπιστῆμαι), and for another, permanent (μονιμοι). And that's what makes knowledge more valuable than right opinion, and the way knowledge differs from right opinion is by being tied down." (*Meno* 97e-98a; Plato 1994, p. 69)

[3] "But I maintain that the awareness [*cognitio*] of his [of the atheist, HR] is not true knowledge [*scientia*], since no act of awareness that can be rendered doubtful [*quae dubia reddi potest*] seems fit to be called knowledge." (Adam-Tannery edition, Vol. VII, p. 141; Descartes 1641/1984, p. 101)

After Gettier's seminal 1963 paper, the idea of stability or indefeasibility has always loomed large in epistemological discussions. For instance, Peter Klein (1971, p. 61) suggested the following *Felicitous-coincidence Principle*:

> If *S*'s evidence for α and a description of some of the particular circumstances in which *S* believes that α are such that it would not be reasonable to expect that α is true (based upon *S*'s evidence), even if α is true, *S* does not know α. Consequently, we might tentatively assert that *S*'s evidence for his belief that α is not sufficiently strong to certify his belief as knowledge if there is some fact which, were *S* to become aware of it, ought to cause *S* to retract his knowledge claim.

In the cases described by Klein, the critic just needs to point out to *S* the circumstances that make *S*'s belief that α unreasonable. This should be sufficient to talk *S* out of believing α. Thus α is not a piece of knowledge according to the stability account.

From Lehrer (1965) at least up to Lehrer (1990), Keith Lehrer has been one of the most prominent champions of the stability account of knowledge. I shall concentrate on the version presented in Lehrer (1990, chs. 6–7). Like Plato, Lehrer suggests a dialogical construal of the stabilty idea. The believing subject is imagined as being engaged in a dialogue with a *critic*[4] (a Socratic dialogue partner) who tries to undermine the subject's beliefs. Only if the subject *wins* the dialogue in the sense that he successfully defends his belief against all the critic's objections, can that belief be called *knowledge*.

Two of the essential rules of the *justification game* are that the critic is omniscient and that she confronts the subject only with information that is true. Such a test for knowledge may appear as a purely internal affair, since it seems to involve only the subject's beliefs and changes of belief, that is, only his internal states. But this is not quite true. The assumption that the critic's objections make use only of *true* statements brings in a connection with the actual world. Truth is what binds subjective beliefs to objective facts.[5]

[4] Lehrer (1990) talks of a sceptic, who is renamed into the critic in Lehrer (2000).

[5] I am neglecting here the constraint characteristic of Lehrer that the critic may only advance information about which the subject has had a definite belief to begin with. – In Rott (2003b), I have advocated a "dynamic" interpretation of the account presented in the first edition of Lehrer's *Theory of Knowledge* (1990). Lehrer (2003, p. 344) denies that his theory was ever meant to be dynamic. It seems to me, however, that Lehrer's (1990) continual talk of "moves", "rounds" and "combinations" of eliminations and replacements in the "ultra justification game" between a claimant and his critic clearly suggests an extended conversation with repeated turntaking. It is much harder to find anything dynamic in the substantially revised theory of the second edition of

So beliefs that fall short of knowledge are vulnerable. A point highlighted in many reactions to Gettier's examples is that the *justification for a belief* may be lost if new evidence comes in. Plato's original point, in contrast, was that *the belief itself* may be lost. This is a simpler idea, since it does not depend on the notoriously controversial concept of justification. Let us suppose that the belief changes occasioned by the incoming evidence are rational in some sense. Then, it seems, beliefs that persist enjoy some sort of justification. I want to make this simplifying assumption and base my discusson upon the following explication of knowledge:

> A belief α is a piece of knowledge of the subject *S*, iff α is not given up by *S* on the basis of *any true* information that *S* may receive.

This is what I will call the *stable belief theory* or the, shorter, *stability theory* of knowledge. My avoiding the term *defeasibility theory*[6] is intended to mark terminologically the difference between the loss-of-justification and the-loss-of-belief ideas. None of the approaches I am dealing with is based on the idea of justification. We presuppose that the subject is in some sense rational in accommodating his beliefs to new information, but we do not assume that justification plays a major role in such processes of belief adaptation.

3. Nozick's sensitivity theory as entailing stability

Robert Nozick's (1981) influential *truth-tracking account* or *sensitivity account* of knowledge is usually presented as an important alternative to indefeasibility theories. But it is worth emphasizing this account was devised so as to entail an element of stability as well. According to Nozick, a subject *S* knows that α if and only if (1) α is true, (2) *S* believes that α, and the following subjunctive conditionals are true:

(3) $\neg\alpha \;\Box\!\!\rightarrow \neg(\text{S believes that } \alpha)$

(4) $\alpha \;\Box\!\!\rightarrow S \text{ believes that } \alpha$

the *Theory of Knowledge* (Lehrer 2000). Still, Spohn (2003) offers a sophisticated reconstruction of the new theory in terms of his belief change model.

[6] Usually associated with people like Keith Lehrer and Thomas Paxson jr., Peter Klein, Marshall Swain, David Annis, Gilbert Harman and John Pollock

Nozick's third condition is a *variation condition*, while (4) is an *adherence condition* (Nozick, p. 211). Regarding (3), the question to be answered is this: What would happen if α were false? (– which in fact it is not) The question regarding (4) is a little harder to formulate. Try this: What would happen if α were true? (– which in fact it is) It sounds strange to call (4) a *subjunctive* conditional even though α is known to be true. Like condition (3), condition (4) is supposed to have some modal force: "Not only is α true and *S* believes it, but if it were true he would believe it. … The truth of antecedent and consequent is not alone sufficient for the truth of a subjunctive" (Nozick 1981, p. 176, variable renamed).[7] What is particularly interesting for our topic is that the antecedent of condition (4) is supposed to cover α-worlds in which the subject is interrogated by a critic. Nozick himself relates (4) to the situation of the Socratic dialogues:

> Meno claimed he could speak eloquently about virtue until Socrates, torpedolike, began to question him. He did not know what virtue was, for Socrates' questions uncovered Meno's previously existing confusions. Even if it had been a sophist's questions that bewildered Meno, getting him to believe the opposite, what he previously had would not have been knowledge. Knowledge should be made of sterner stuff.
> Thus, some skeptical arguments play off condition 3, others off condition 4.[8]

It is clear that for Nozick the contingent truth of α and β in the actual world w_a does not suffice to make the subjunctive conditional $\alpha \Box\!\!\rightarrow \beta$ acceptable. But how far must we be ready to deviate from the actual course of events in order to test for the truth of the conditional? How far does adherence to the belief that β have to extend among the α-worlds?

In this attempt to answer this question, Nozick employed a model using *spheres of possible worlds* due to David Lewis (1973). A sphere is the set of possible worlds that are similar to the actual world w_a up to a certain degree. The smallest sphere is the singleton $\{w_a\}$. We already said that this set is not enough for the evaluation of the conditional, we have to consider larger spheres. But is it sufficient to consider the second smallest sphere, the set of possible worlds that are closest to, but not identical with the actual world; or do we have to go

[7] Williams gives good gloss of the conditional (4): "If, in somewhat changed circumstances, it were still the case that α, I should still believe that α." (Williams 2001, p. 30, variable renamed)

[8] Nozick (1981, p. 213). This passage shows, I believe, that classifying Nozick as a pure externalist would miss an important point.

out until we meet the closest $\neg\alpha$-worlds; or is some intermediate sphere adequate? Nozick (1981, pp. 680–681) has a long, complicated and somewhat irresolute footnote about this question, suggesting that we must indeed go out to a level that includes at least the closest (but maybe many more) $\neg\alpha$-worlds.[9] Taken together, it seems that (3) and (4) are meant to imply that the subject's belief-that-α covaries with the fact-that-α "for some distance out in the closest α band to the actual world". Since Nozick thinks that this band contains worlds in which critical conversations with the critic take place, we conclude that meeting a critic does not mean a big deviation from the actual world for Nozick.[10]

So there is a lot of support for the stability analysis in the epistemological literature. Two things remain to be noted. The approach does not take care of the case where the subject is presented with *misinformation*. It is not clear whether knowledge should be robust against local errors of perception or memory, or wrong testimony from the critic. To take up Nozick's phrase, shouldn't knowledge be made of still sterner stuff – stuff that also survives (a modest amount of) misinformation? I just want to raise the question here; I am not going to further pursue it in this paper.

Secondly, even true information may be misleading. Sometimes there is a definite bias in the kind of information that we receive (from a used-cars salesman, for instance). Even if every single piece of information the subject receives is true, the picture that emerges may tempt him to draw the wrong inferences, thereby undermining what he (apparently) knew before. The problem of misleading defeaters and pseudo-defeaters of knowledge has accompanied stability theories form their beginning, and we will return to this point below.

4. An internal affair: Strong beliefs

One may plausibly expect that the stability of a belief derives from its strength. It is instructive to look at the relation between these concepts more closely. We have to account

[9] This is very similar to the *sphere of epistemically relevant worlds* as determined by DeRose's (1995, p. 493) *Rule of Sensitivity*: "When it's asserted that *S* knows (or does not know) that *P*, then, if necessary, enlarge the sphere of epistemically relevant worlds so that it at least includes the closest worlds in which *P* is false." (At this point I neglect that both Nozick and DeRose qualify their definitions by holding fixed a certain method of belief-acquisition; compare footnote 24 below.) Goldman (1987) gives arguments to the effect that the subject need not always go out that far for knowledge.

[10] Nozick's assumption that meeting a critic or a skeptic is a nearby possibility is of course compatible with DeRose's presupposition that the truth of the skeptical hypothesis itself is a remote possibility.

for varying degrees of belief, and we will do that in the simplest possible way, by means of a qualitative modelling.[11] Let us look at two interdependent ideas to represent the idea of *strong belief*:

a) high epistemic entrenchment (high epistemic rank)
b) stability (persistence, tenacity) in certain kinds of belief change

As a model for belief states we take a subjectivist version of the model already appealed to by Nozick. Formally, we replace Lewis's (1973) objectivist conception by Grove's (1988) subjectivist conception of systems of spheres.[12] Let us represent a doxastic state by a system of nested sets of possible worlds, supposing, for the sake of simplicity, that everything is finite. The smallest set is the set of possible worlds which the subject believes to contain the actual world w_a. If the subject receives evidence that the actual world is not contained in this smallest set, he falls back on the next larger superset. And again, should it turn out that the actual world is not to be found in this set either, the subject is prepared to fall back on the next larger set of possible worlds. And so on. The sets or *spheres* of possible worlds correspond to spheres of plausibility, or to put it differently, spheres of deviation from the subject's beliefs. The spheres are the subject's personal spheres of possible worlds as it were, spheres for the first person. The system of spheres taken as a whole represents a mental state (viz., a doxastic state) and must not be expected to be centered on a single world w_a that represents the actual world. If one of the subject's beliefs is wrong, then w_a is not even contained in the innermost sphere, but may occur at any arbitrary position in the sphere system. Let us now see how we can use this modelling to represent the idea of strong belief.

Re a) We can identify the strength of a belief with its degree of *doxastic entrenchment*, where the degree of doxastic entrenchment of a belief α can be measured by the number of spheres that contain exclusively α-worlds. The more spheres (i.e., the more fallback positions) are fully covered by α, the better entrenched α is.[13]

Re b) The entrenchment terminology suggests that we are interested in how hard it is to eliminate a belief. Rather than defining the resistance against elimination with reference to a fixed doxastic state, we can refer directly to the potential developments of that doxastic state.

[11] One can retain the spirit of the possible worlds modelling and in addition take advantage of the structure of ordinal numbers, thereby gaining a lot of additional expressive power. See Spohn (1988).

[12] More generally, one could use non-nested systems à la Lindström and Rabinowicz (1991).

[13] Compare Lindström and Rabinowicz (1991) who incidentally also introduced the fallback terminology.

A belief is *stable* to the extent that it is unlikely that the belief is lost in processes of belief change.

What kinds of belief change should we take into account? Here we return to the dialogue model with the critic, and add that a third rule of a Lehrerian justification game is this: The subject must accept the pieces of true information the critic provides it with (in this sense, the critic's objections must be "successful"). So the subject has to be ready to actually *add* new information. Two questions suggest themselves: Should we be ready to account for the case where what the critic tells the subject is *incompatible* with the latter's beliefs? Should we be ready to account for the case where the critic prompts the subject to *subtract* a belief rather than add a new one? It is important, I am going to argue now, that both questions are answered in the affirmative.

The need for *belief-contravening revisions*, belief changes induced by new information that contradicts the subject's prior beliefs, is obvious if we endorse a simple thesis of *fallibilism*: For all subjects and at all times, some of the subject's beliefs are wrong.[14] That we are all fallible is a basic fact of life. Human beings have a hard time refraining from believing, they tend to be credulous, and many people think: excessively credulous. As a consequence, we always have to face the fact that some the countless beliefs we hold are mistaken. As we gather more evidence and obtain more true information from various sources (from our relentless critic, for example), we will sooner or later encounter conflicts with our previous beliefs. In such cases, we have to perform belief-contravening revisions.

The story of how to base a revision of the subject's beliefs upon the system of spheres representing his doxastic state is easy to tell (Grove 1988, Gärdenfors 1988). If β is the new bit of information, the subject looks for the smallest sphere that contains at least one β-world. The subject believes α after successful performance of a revision by β just in case α is true in all β-worlds that are contained in this smallest β-admitting sphere. This recipe works regardless whether β is or is not consistent with the subject's previous beliefs. Figure 1 may serve as an illustration.

[14] This broadly Peircean or Popperian notion of fallibilism is of course different from Cohen's (1988) and Lewis's (1996) fallibilism which says that there is *fallible knowledge*, knowledge despite *uneliminated* possibilities of error. In contrast to Lewis's contextualist model, the model we are going to talk about in this and the next section cannot be purely "eliminativist" in nature.

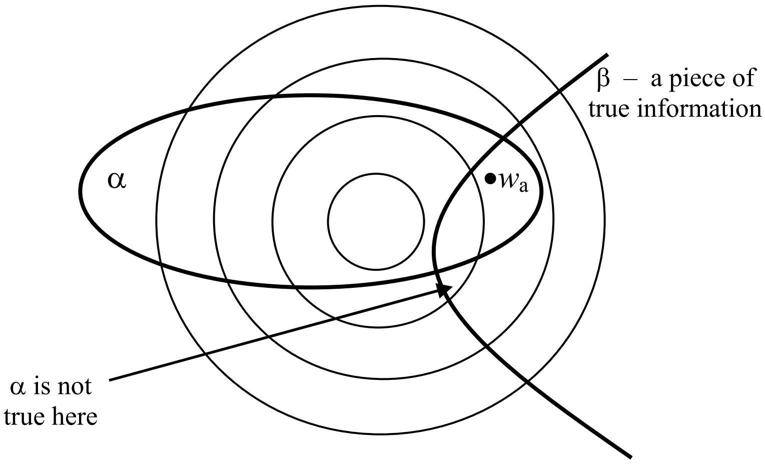

Figure 1. Here, α is a true belief, but α [or more exactly, β⊃α] is not sufficiently well-entrenched to survive the revision by the true information β. So *S* does not know that α.

5. The stability analysis of knowledge and the strength of beliefs

One will be inclined to think that there must be a tight connection between the strength of a (true) belief and its stability under (truthful) criticism. But the two concepts cannot be identical since strength seems to be a purely 'internal' property, whereas stability as just defined imposes 'external' constraints through the requirement that the critic's statements be all true.

The stability account may be formulated in the setting of one-shot belief revision: Subject *S* *knows that* α if and only if *S* believes that α and α is not given up by *S* after receipt any true information (from the critic, say). More precisely, α is a piece of knowledge of *S* if and only if α is not lost when *S*'s set of beliefs is revised by any arbitrary true piece of information. With a little help from belief revision theory, we will now prove that the following result:

Observation. The belief α is stable with respect to the revision of *S*'s belief set by any true piece of information if and only if α is more entrenched in *S*'s belief state than every false belief; or equivalently, in the system of spheres modelling: if and only if α holds not only throughout the innermost sphere but also throughout the smallest sphere containing the actual world w_a.

Proof. We want to show that α is stable under truthful revision iff it is more entrenched than any falsehood, in symbols:

(†) $\forall\beta$ (β is true $\Rightarrow \alpha$ is in $B^*\beta$) iff $\forall\gamma$ (γ is false $\Rightarrow \gamma < \alpha$)

where B denotes the subject's original belief set and $B^*\beta$ denotes the belief set that results from revising B by the sentence β.

First of all we have to connect the notion of entrenchment with the subject's belief change behaviour. A sentence β is at most as *entrenched* as a sentence γ, in symbols $\beta \leq \gamma$, iff β is lost when the subject learns that the conjunction of β and γ is not true, in symbols, iff β is not in $B^*\neg(\beta\&\gamma)$. Call this the definition of entrenchment.[15]

This definition entails the dominance condition which says that $\beta \leq \gamma$ whenever β logically implies γ (in this case β is not in $B^*\neg(\beta\&\gamma) = B^*\neg\beta$).

The left-hand side of (†) implies the right-hand side: Suppose the right-hand side is false, i.e., α is not more entrenched than every falsehood. Then there is a false γ such that $\alpha \leq \gamma$. Now consider $\neg(\alpha\&\gamma)$. This sentence is true, since γ is false. By $\alpha \leq \gamma$ and the definition of entrenchment, it follows that α is not in $B^*\neg(\alpha\&\gamma)$, so α is not stable under truthful revision, i.e., the left-hand side is false.

The right-hand side of (†) implies the left-hand side: Suppose the left-hand side is false, i.e., α is not stable under truthful revision. Then there is a true β such that α is not in $B^*\beta$. Since β is in $B^*\beta$ and this set is logically closed, it follows that $\beta\supset\alpha$ is not in $B^*\beta = B^*\neg((\beta\supset\alpha)\&\neg\beta)$ (notice that $\neg((\beta\supset\alpha)\&\neg\beta)$ is logically equivalent with β). By the definition of entrenchment, this means that $\beta\supset\alpha \leq \neg\beta$. Now by the dominance condition, $\alpha \leq \beta\supset\alpha$, so by the transitivity of entrenchment $\alpha \leq \neg\beta$. Since $\neg\beta$ is false, we have found a falsehood that is at least as entrenched as α, i.e., the right-hand side is false. QED

Using this Observation, we can see that knowledge in the stability interpretation does not require maximal entrenchment,[16] but it is indeed characterized by a certain degree of entrenchment (i.e., by a certain strength of belief). The particular strength of belief that is required depends on the position of the actual world w_a in the system of spheres. If the subject considers w_a to be a fairly plausible world, knowledge does not require very strong belief. If, however, w_a is far out in the subject's system of spheres, knowledge requires very highly

[15] A similar definition in terms of belief contractions was first suggested by Gärdenfors (1988, p. 88).

[16] As suggested ("unofficially") from a belief revision perspective by Segerberg (1998). Also cf. Segerberg (1999, p. 345). That knowledge requires maximal justification or certainty has of course been a central claim in much traditional epistemology.

entrenched belief. Prima facie, it looks like an element of epistemic luck where in the subject's system of spheres the actual world happens to be placed. But perhaps it is not luck after all where w_a is being located, but rather merit – a sign of how good *S*'s doxastic state is. It is certainly a virtue of an epistemic subject to have his beliefs in good accord with the actual world.

Still I think that the Observation discloses a problematic feature of the stability analysis which ties knowledge too tightly to the strength of belief. As a first indication, consider the epistemically ideal case in which *S*'s beliefs are all true. In the sphere model, this means that w_a is contained in the innermost sphere. Then, according to our Observation, a truthful critic can never talk *S* out of believing *any* of his beliefs. True information will only result in a consistent addition of beliefs (i.e., in the elimination of possible worlds from the innermost sphere). According to this analysis, if all of the subject's beliefs are true, each and every belief of his constitutes a piece of knowledge. This, however, is counterintuitive. Intuitively, having only true beliefs does not protect *S* against being dissuaded from believing a particular one of his beliefs. Problems for the more realistic case where *S* has some false beliefs will be discussed in Section 8.

6. Critics, skeptics and the meaning of *might* sentences

The skeptic is not so much a provider of new evidence as someone who *raises doubts* and *calls* beliefs *into question*. The critic, we said, supplies the subject with new, truthful information. The skeptic, in contrast, does not furnish positive information. Her mission is a negative one, it typically leads to the subject's relinquishing some information without getting anything new. That is, she instigates processes of belief *elimination* or *contractions* of belief sets rather than their revisions. The skeptic does not positively claim that *S is* a brain in a vat, she rather points out that *S might be*, for all he knows, a brain in a vat. She does not assert that those animals in the zoo of Berlin *are* cleverly painted mules, she only says it is *possible* that they are.

This leaves us with the question of how to deal with such modalized statements. Assuming again that the subject *S* has to accept what the skeptic is saying, we need to specify the sort of belief change that goes on in *S* after accepting the skeptic's *might* sentence. So suppose the skeptic says *might*-α. What the subject does first, I suggest, is try out what his beliefs would look like after accepting α. But then, since he has no positive evidence that α is actually true, he settles for what is common to his current belief set *B* and the result of revising *B* by α. This procedure can also be reinterpreted as a process of withdrawing $\neg\alpha$ from the subject's belief set *B*; in this reading "revise *B* by *might*-α" means "withdraw $\neg\alpha$ from *B*".[17]

If we admit *might* sentences as skeptical objections, we can frame an argument to the effect that Nozick's positive conditional (4) implies his negative conditional (3), given that α is true. We show this by contraposition. So suppose that α is true, but that *not*

(3) $\neg\alpha \;\Box\!\!\rightarrow \neg(S$ believes that $\alpha)$

According to the semantics for subjunctive conditionals, this means that there is a plausible/relevant possible world such that

(‡) $\neg\alpha$ and *S* believes that α

is true in that world. Now assume that the critic tells the subject about this possibility by uttering the sentence

might-($\neg\alpha$ and *S* believes that α)

According to the rules of the justification game, *S* accepts this sentence. We said that this means that *S* checks, for the sake of argument, what his beliefs would look like after revising them by (‡). Since consistency is to be respected, the subject loses his prior belief α in the revised belief set, and α remains lost of course if this set is intersected with the original belief set. We have now described a scenario in which α is true and '*S* believes that α' is false. Let us assume (with Nozick) that such a scenario is plausible and relevant, and thus close to the actual world. Then it follows that the positive conditional

(4) $\alpha \;\Box\!\!\rightarrow S$ believes that α.

does not hold. This completes the proof that (4) implies (3), provided that α is true, the conversation with the skeptic is close to the actual world and the skeptic is allowed to put forward *might* sentences.

What difference does it make whether the "information" supplied by the critic comes in the form of a categorical or in the form of a modalized sentence? It makes a big difference, since the rules of the game constrain her to give true information only. If the sentence α is false, then *might*-α may still be true. So the critic – or rather: the skeptic – has a lot more

[17] In symbols: $B^*(\textit{might}\text{-}\alpha) = B \cap B^*\alpha$. Likewise, a belief contraction with respect to $\neg\alpha$ can be defined by the equation $B{-}\neg\alpha = B \cap B^*\alpha$ which is known in the belief revision literature as the Harper identity (Gärdenfors 1988, p. 70).

possibilities to talk *S* out of believing a proposition if she is allowed to use *might* sentences. By assumption, she is omniscient, she knows the whole truth, and she speaks nothing but the truth. But when is a *might* sentence true? This, of course, depends on the meaning of the modality. The most common reading of skeptical objections is to understand *might* epistemically: *S* cannot exclude, for all he knows, that he is deceived by an evil demon, that he has been envatted by evil scientists, that this animal in the zoo is a cleverly painted mule etc. But it is doubtful that the epistemic understanding of *might* is the right one to plug in into Nozick's conditions (3) and (4). We are stepping into deep waters here, waters that we cannot even begin to fathom out here. For the rest of this paper, I shall assume that the objections that the critic raises can be expressed in non-modal terms. I will not deal with *might* sentences any more. My critic is not supposed to be a skeptic.

7. More on internal affairs: Dialogues and piecemeal evidence

Plato's Socrates liked to stretch his teaching out in long dialogues. Lehrer, too, used dialogues to illustrate his concept of knowledge. It has been argued that there is not only a heuristic, but an epistemologically significant difference between presenting corrective evidence all at once and presenting it *seriatim*.[18] In order to account for this, the classical model of one-shot belief revision must be extended to a more elaborate one that a conversation with the critic typically consists of several rounds, in each of which she would release new information. A good model of the belief change that the subject is experiencing in such a conversation must be able to describe *iterated belief changes*.

If we want to stick to the simple systems of spheres modelling, there is a rather limited number of methods for iterated belief change, and is not quite clear which (if any) of these models can adequately capture the kind of process that we need for the conversation with the critic. Consider for illustration a slightly modified variant of an example of Lehrer (1965). Let *p* stand for the sentence "Jones owns a Ferrari", *q*, *r*, *s* and *t* for corresponding sentences about other colleagues of the subject owning a Ferrari. Let us suppose that the doxastic state of Gettier's subject regarding this matter is represented by the system of spheres in Figure 2:

[18] Fogelin (1994, ch. 2) and Williams (2001, ch. 4).

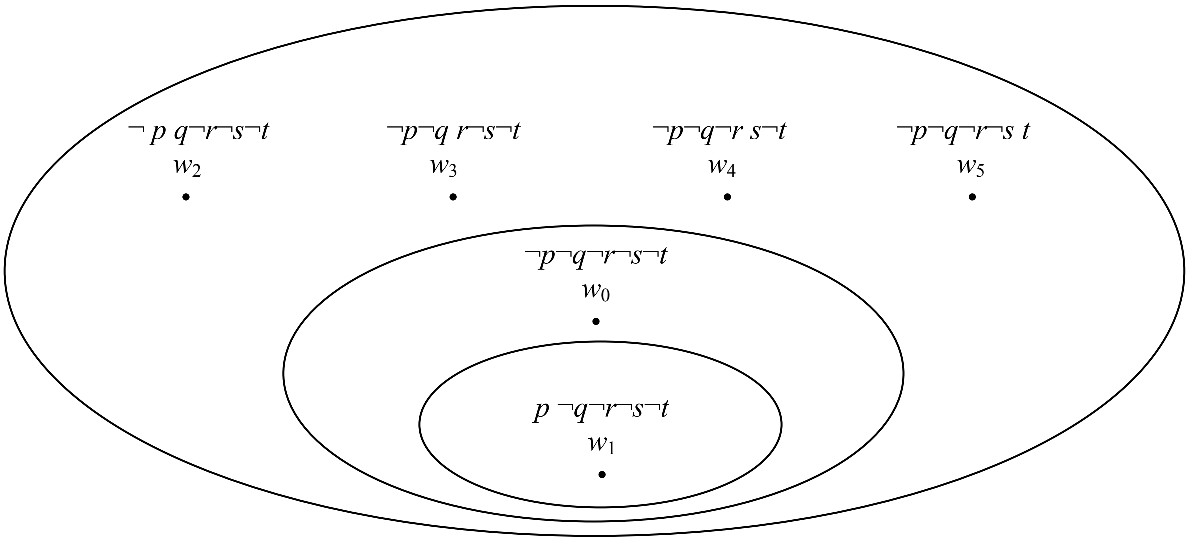

Figure 2. Gettier case, with *S* believing that Jones owns a Ferrari (*p*).

The subject's initial beliefs include that Jones has got a Ferrari, while the others have not. *S* thinks that w_1 is the actual world. The second most plausible situation is the one in which none of his friends owns a Ferrari, i.e., w_0. Only at the next level are there worlds in which some of his other colleagues owns a Ferrari: worlds w_2 through w_5. Figure 2 does not show the still more far-fetched situations in which more than one of his friends owns a Ferrari. After all, Ferraris are not meant to be everybody's cars.

Now assume that it is in fact Brown who owns a Ferrari ($w_a = w_5$, say), and imagine the critic beginning to tell *S* the truth about the situation. Her first hint is

(1) "Jones has not got a Ferrari." $(\neg p)$

The subject's straightforward reaction is, on any of the standard accounts of belief revision, to proceed to a belief state that takes w_0 to be the true world. With this, *S* is still wrong. Imagine the critic passing on a second piece of information to the subject

(2) "Someone has got a Ferrari." $(p \vee q \vee r \vee s \vee t)$

Now the classical one-shot belief revision theory of the 1980s (Gärdenfors 1988) was at a loss about how to revise the subject's beliefs in the second step. In the 1990s, however, a number of techniques were developed to deal with iterated changes in the simple possible worlds setting that we are using in this paper.[19] Different methods lead to different reactions to the second of the critic's hints. If *S* chooses to apply the method of *conservative belief revision*, he returns to w_1 as the most plausible world, and again believes that Jones owns the Ferrari. However, since our critic invariably tells the truth, forgetting about (1) is not the type of reaction that we would like to see. If *S* applies the method of *moderate belief revision*, he reaches the conclusion that the true world is among $w_2, \ldots, w_6$, and thus believes that the owner of the Ferrari is one of the persons in question, with the exception of Jones. This conclusion is what we expect of a rational person.

If knowledge is stable belief, Jones cannot be said to know that someone in his class owns a Ferrari – which is in accordance with our intuitions. If the revision method employed is the conservative one, however, then Jones may be said to know that *if* someone owns a Ferrari, then it is Jones. This is too conservative. The subject should be able to learn more from the critic's information, he should not revive in the second step his false initial belief that Jones is the owner of the Ferrari. The method of moderate belief change is just what we need for the dialogue with the critic. It consistently accords incoming information priority over old beliefs. In fact, since everything the critic says is true (by the rules of the justification game), the conjunction of her statements is consistent. Moderate belief change is such that iterated changes by a sequence of jointly consistent bits of information $\alpha_1, \alpha_2, \ldots, \alpha_n$ always result in the same belief set as a single change effected by the conjunction α_1 & α_2 & … & α_n.[20] All evidence supplied by the critic *seriatim* can be collected and has the same effect as if the evidence were presented all at once.

We can conclude that belief revision theory has the resources appropriate to deal with a stepwise correction of the subject through an extended dialogue with the critic. As long as each piece of input is true, the stability of a belief under sequences of revisions is reducible to its stability under various one-shot revisions. The Observation of Section 5 linking the stability account to strengths of belief transfers to the iterated case without modification.

[19] For a general survey of these developments and for a discussion of the methods of conservative and moderate belief change, see Rott (2003a).

[20] Rott (2003a, pp. 131–136). This reduction of iterated revisions of belief sets to one-shot revisions crucially depends on the critic's being consistent – which is guaranteed because she only speaks the truth. For the same reason, I think that the order-independence of the revisions in question is intuitively desirable. If we look at the level of systems of spheres rather than at the level of belief sets, then no reduction of iterated to one-shot revisions is possible; at this level, it becomes manifest that the method of moderate revision invariably gives

8. A problem for the stability account of knowledge

We have mentioned in the introduction that defeasibility theories of knowledge were diagnosed as problematic soon after their invention. We shall now show that our move to the stability theory (that substitutes loss-of-belief for loss-of-justification) does not get round the problems. The point is that it is fairly easy for the critic to talk the subject out of a belief, even if intuitively the belief constitutes genuine knowledge. Consider the following abstract argument due to Jacob Rosenthal which can actually be seen as an illustration of our Observation in Section 5.[21] Suppose that *S* knows that α, but that α is not maximally entrenched in *S*'s belief state. Suppose further that *S* has a very well-entrenched belief β that happens to be false. Let us assume that α is not more entrenched than β. Then *S* can be talked out of believing α in the following way. The critic correctly points out that α&β is false. By the rules of the justification game, *S* recognizes that what the critic says is right, and he accepts ¬(α&β). In order to maintain the consistency of his beliefs, *S* has to remove α&β. Being logically competent, *S* realizes that he has to remove either α or β. By our hypothesis that α is not more entrenched than β, the belief α has to go (this is what the term 'entrenchment' means). So *S* has been talked out of believing α by the critic. – Hence, if the stability analysis of knowledge is correct, *S* has not known that α to begin with. Contradiction. Hence *S* can know α only if α is more entrenched in *S*'s belief state than every other belief that happens to be false. One well-entrenched false belief erases as it were a lot of putative knowledge that has not got anything to do with it.

Now the obvious question is: Doesn't this show that the stability analysis is fundamentally flawed? Tentative answer: No, but we have to refine it. Intuitively, it seems the critic should only question statements that are somehow 'basic', statements on which α depends rather than statements that depend on α themselves. And in the argument just sketched, the criticized proposition α&β was presented as parasitic on the (more) basic beliefs α and β.

priority to more recent over less recent information. For conservative belief change, a reduction of iterated revisions to revisions by conjunctions is impossible even at the level of belief sets.

[21] See Rosenthal (2001, pp. 546–547; 2003, pp. 254–255). His discussion is inspired by Lehrer's (1990, pp. 137–140; 2000, pp. 156–160) recent discussions of examples for misleading evidence, viz., the Grabit and the newspaper examples. Lehrer takes the Grabit example to refute Klein's (1971) proposal. The criterion Lehrer takes as decisive in this context is the "dependence on a false belief", but I doubt that this notion can carry the theoretical weight necessary for the separation of knowledge from mere belief.

But we cannot get rid of the problem that easily, as is shown by the following more concrete example. Suppose I think I observed that Grabit stole a book from the library at 3 p.m. Suppose further that I had forgotten my glasses that afternoon. So, being short-sighted, I am not absolutely sure that it was Grabit who stole the book (*p*), although for all practical purposes I would not hesitate to rule out the possibility that it was someone else. When making this observation, I looked at my very reliable Rolex watch, so I am very sure that it was 3 p.m. when the book was stolen (*q*). I have an excellent reason to believe *q*, a better reason anyway than I have for my believing that *p* is true. As a matter of fact, however, Grabit did steal the book, but it was already 3:30 when that happened. (My reliable Rolex had stopped working for a while, a fact that escaped my attention because it later reset itself with the help of a radio signal.) By everyday standards, I may truly be said to know that Grabit stole the book. But of course I cannot be ascribed knowledge that this event took place at 3 p.m. At that time Grabit was still having lunch with some of his colleagues, all respectable people who make for irreproachable witnesses. Now a critic may rightfully point out to me that my original belief that Grabit stole the book at 3 p.m. is not true. Being forced to retract this belief, I conclude, on the basis of the quality of the evidence that I possess, that *p* must be false and *q* must be true.

This is certainly a rational reaction. The critic, however, has managed to talk me out of believing something that I seem to have known before, viz., that Grabit stole the book (*p*). What are we to say now? Was *p* unstable knowledge, or was it no knowledge at all? The stability theorist is committed to saying it wasn't knowledge to begin with, but this seems counterintuitive. The mere fact that the subject has a false belief *q* that is sufficiently well-entrenched to drive out the true belief *p* should not in itself be sufficient to discredit *p*'s claim to the status of knowledge. But the tentative answer to the abstract case described before does not seem to be available any more. There is no reason to deny that my "basic" belief was precisely that Grabit stole the book at 3 p.m. Doesn't it look artificial to formalize this original belief by *p&q*? There is no good motivation for splitting this belief up, as we just did for the sake of exposition, into the two halves "Someone stole the book at 3 p.m." and "Grabit stole the book some time". And it can hardly be claimed that the critic's clue was misleading.

As a side remark, the following observation may be interesting: It can be shown that if we assume that the subject's belief set is logically closed,[22] then *no* belief-contravening revision that does not result in an omniscient belief set strictly enlarges the set of true beliefs. The subject is bound to lose some true information in his conversation with the critic, notwithstanding the fact that the latter's clues may improve the subject's belief set in any intuitive sense. Even if the subject's prior beliefs contain some falsehoods while the new piece of information as well as all posterior beliefs are true, the revision prompted by the critic will make the subject lose some true beliefs.[23] To be sure, this is only an observation about *belief*. But we begin to form an idea that it is much easier for the critic talk the subject out of his *knowledge* than we might have suspected.

It still seems to me that the notion of stability captures an important aspect of knowledge, but I confess that I do not know how to repair the stability account so as to avoid the problems we have identified. We now leave the criteria that are internal in the sense that they refer to the development subject's mental state (without referring to processes of justification), and turn to the external notions of the strength of the subject's epistemic position.

9. External affairs

In the last two sections, I have used systems of spheres as representations of belief states, as structures that determine the strength of a subject's belief and help him to revise his beliefs not only once, but several times. Systems of spheres of possible worlds were also appealed to by epistemologists like Nozick (1981) and DeRose (1995). It is important, however, to keep distinct in the formal modelling what is distinct in substance. In Nozick's sensitivity account of knowledge, subjunctive conditionals are evaluated with the help of systems of spheres. These spheres are not those of the epistemic subject, but those of a third person that ascribes

[22] Most contextualists assume that a subject's *knowledge* set is closed under *known* logical implication. What I am assuming here is different: That the subject's *belief* set is closed under logical implication. Of course this assumption is not realistic for explicit beliefs. But it makes good sense for implicit beliefs (beliefs ascribed or beliefs the subject is committed to). Advocates of the assumption include Daniel Dennett, Isaac Levi and Robert Stalnaker.

[23] Proof: Suppose *S* is provided with new belief-contravening information, α, and *S* believes $\neg\alpha$ before the revision. Suppose further that β is some truth that is not believed after the revision has taken place. Then, by the deductive closure of *S*'s prior and posterior belief sets, $\alpha\supset\beta$ is a true proposition that is believed before, but not after the revision (for more details, compare Observation 6 of Rott 2000). If we wanted to show that $\alpha\supset\beta$ is a piece of *knowledge* that is being lost, we would have to flesh the story out in such a way that the original belief $\alpha\supset\beta$ was not dependent on the false belief $\neg\alpha$.

knowledge to the subject. They do not, however, represent the ascriber's belief state. They are meant to represent objective similarities between possible worlds. The location of a possible world in such a system does not represent the world's plausibility, but its distance (however conceived) from the actual world w_a. Nozick and DeRose appeal to Lewis-style systems centered on the actual world w_a. In contrast, the Grove-style systems used for the stability analysis are not centered on any world, and the position of the actual world w_a cannot be determined on a priori grounds (we said before that its placement may just be a matter of luck, but it may also be a sign for the aptness of the subject's doxastic state). Systems of spheres representing the subject's belief state are obviously subjective. Systems of spheres representing similarities are subjective in a much less evident way; similarity is always similarity in certain interesting or salient respects.[24]

Systems of spheres may thus play very different roles in the analysis of belief and knowledge. There are also two different notions of *strength* that must not be confused. We have already noted that the *strength of a belief* may be identified with its degree of epistemic entrenchment in the subject's doxastic state. Everything about this concept is internal, and the strength of a belief may be assumed to be completely transparent to the subject.

This is all very different from the *strength of the epistemic position* in which a subject is with respect to a certain belief. An interesting interpretation of this concept is suggested by Keith DeRose (1995, 490–492). A subject is in a *strong epistemic position* with respect to α if and only if his belief that α covaries with the truth of α not only in the actual world, but also in all worlds that deviate from the actual world to a quite significant degree. The more deviation from the actual world is tolerated without destroying the covariance of truth and belief, the stronger the epistemic position of the subject is. How strong or weak the subject's epistemic position actually is need not be transparent to him. The strength of the epistemic position is partly an objective matter (after all, it is the *actual world* that forms the center of the relevant system of spheres) and partly something to be judged by the third person's subjective standards (it is *her* similarity relation that serves as a measure of the deviation).

[24] Motivated by some recalcitrant examples, both Nozick and DeRose instruct us that in judging similarities we need to give a lot of weight to the subject's method or way of coming to believe α. This is not an aspect that would normally be regarded as important for determining a possible world's overall similarity with the actual world.

Might it be that a system of spheres is subjective and objective at the same time? Well, in cases of self-attributions of knowledge, the first person appears to take the role of the third person and many of the distinctions we have just made seem to collapse. But we should be aware of the fact that even when strength of belief (first person perspective) and strength of epistemic position (third person perspective) pull in the same direction, this is not sufficient to decide the case for knowledge. Let us have a look at the system of spheres depicted in Figure 3, which we now assume to represent both the subject's belief state and the attributor's similarity relation at the same time.[25] Let *H* stand for a skeptical hypothesis (like "I am a brain in a vat" or "This is a painted mule") and *O* stand for an ordinary hypothesis (like "I have hands" or "This is a zebra"). In the actual world, *S* believes that O, but does not believe that *H* is true. More importantly, since *H* and *O* are conceptually incompatible, the ordinary belief can only be true if the skeptical hypothesis is false.

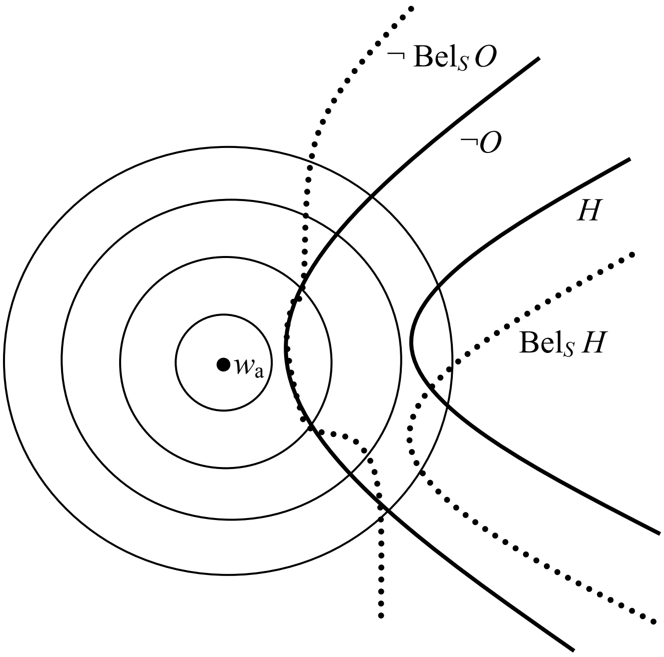

Figure 3. Hypotheses and beliefs in hypotheses.

In the situation depicted in Figure 3, *S believes more firmly* that $\neg H$ than that *O*, since $\neg H$ is true throughout four spheres and *O* is true only throughout two spheres (notice that the smallest sphere is the singleton set $\{w_a\}$). Similarly, the covariance of belief and truth extends

[25] This assumption is made just for the sake of argument. It is dubious even if first and third persons coincide. There are at least three serious problems: (i) Plausibility for the subject is conceptually different from similarity with the actual world; (ii) a sphere system with $\{w_a\}$ as its innermost sphere, epistemically interpreted, represents a situation in which the subject is both infallible and omniscient; (iii) since belief is transparent to the agent, one would probably expect that Bel$_S$ *O* and Bel$_S$ $\neg H$ hold throughout all spheres from the first person perspective.

farther with respect to ¬*H* than with respect to *O*, which means that the *S*'s *epistemic position* with respect to ¬*H* is *stronger* than with respect to *O*. Nevertheless, according to the sensitivity model of Nozick, the subject *knows* that *O* but does *not know* that ¬*H*. In each of the closest (most plausible) worlds where *O* is false, *S* would cease to believe that *O*, but there are some closest (most plausible) worlds where *H* is true and yet *S* would not believe it. Since the skeptical hypothesis is very far-fetched (as skeptical hypotheses typically are[26]), the subject needs to be in a *very* strong epistemic position with respect to *H* in order to *know* whether it is true or false. Since on the other hand the ordinary hypothesis is much more mundane, the epistemic position with respect to *O* need not be very strong in order to know whether *O* is true or false. So even strength of belief and strength of epistemic position taken together provide no reliable indication of a belief's claim to the status of knowledge – if knowledge is understood as characterized by Nozickian truth tracking.

10. Conclusion

There are various routes to explicating knowledge without reference to the concept of justification. In this paper I have had a first look at the relationship between some other properties that have been thought to contribute to converting true belief into knowledge: stability, sensitivity to truth, strength of belief and strength of epistemic position. I have tried to make clear the different roles that can be played by a model using systems of spheres of possible worlds, and to sort out some subjective and objective factors involved.

While it is fairly obvious that strength of a belief cannot in itself be a criterion for knowledge, the other properties have indeed been held to be good criteria of knowledge, at least by some authors. It is time to summarize what we have found out about their mutual relationship.

The relation between sensitivity and strength of epistemic position has been set out nicely by Keith DeRose (1995, pp. 491–492), and there is little to add to that. For ordinary beliefs, a *good* epistemic position is sufficient for knowledge according to the sensitivity analysis. For extraordinary beliefs (like the belief that a skeptical hypothesis is false) a *good* epistemic position is not normally sufficient; it has to be excellent. Conversely, if one knows that α

[26] Skeptical hypothesis are far-fetched in the sense that the first worlds satisfying them can be found only in the periphery of the system of spheres. This is an important presupposition of DeRose's account. Michael Williams has made it clear **(in discussion)** that he strongly disagrees with this view.

according to the sensitivity analysis, this implies that the epistemic position is fairly good for ordinary beliefs, and it implies that it is excellent for extraordinary beliefs. Figure 3 shows how one's belief in *O* can be sensitive and one's belief in $\neg H$ can be insentitive while at the same time one's epistemic position is stronger with respect to $\neg H$ than it is with respect to *O*.

Now we turn to the stability of a belief which we saw to be intimately linked to its strength in Section 5. Our discussion has suggested that stability is related to sensitivity, but that this relation is far from perfect. Nozick himself pointed out that his fourth condition is meant to imply persistence of the belief under Socratic criticism. If the subject's belief cannot survive such a procedure, we have to deny the subjunctive conditional $\alpha \mathbin{\Box\!\!\rightarrow} (S \text{ believes that } \alpha)$. Nozick's argument depends crucially on the idea that such a critical conversation may take place in worlds that are close to the actual world.

For the converse direction, it seems that knowledge according to the stability analysis implies knowledge according to the sensitivity analysis only in one of two possible cases. As we have seen in Section 6, if the negative conditional

$$\neg\alpha \mathbin{\Box\!\!\rightarrow} \neg(S \text{ believes that } \alpha)$$

is wrong, then the critic can feed the subject with the modalized information that *S may* be in a situation in which *S* believes that α is true, even though $\neg\alpha$ is actually true. The subject's belief in α would then appear to be shaken. If, on the other hand, the positive conditional

$$\alpha \mathbin{\Box\!\!\rightarrow} (S \text{ believes that } \alpha)$$

is wrong, then I cannot see how this could be detected by the critic's attempt to talk *S* out of believing α.

Regarding the relation between stability and strength of epistemic position (in the DeRose's sense), the latter implies the former in so far as α is true, given that we endorse Nozick's assumption that the conversation with the critic takes place in the vinicinty of the actual world. Stability, on the other hand, does not seem to imply strength of epistemic position.

The traditional notion of justification plays no role in the accounts that we have discussed. In fact, it is controversial even between the founders of the standard belief revision paradigm to what extent this paradigm can account for the justificatory structure of beliefs. Gärdenfors (1990) argues that foundationalist intuitions can be captured in the AGM model at least by reconstruction, while Makinson (1997) emphasizes how important it is to realize that this model does *not* come equipped with any justificatory struture. I tend to think that Makinson's picture better captures the nature of belief revision theory.

Belief revision theory thus seems orthogonal to the traditional concerns of mainstream epistemology. Perhaps the best account of how to understand their relation is still to be found in Harman (1986).[27] But there is also a fully developed alternative philosophical theory of knowledge that is not only congenial with belief revision theory, but has to some extent even motivated it, namely the work of Isaac Levi (1980, 2004). Levi's pragmatist attitude is stongly opposed to any kind of 'pedigree epistemology' and importantly characterized by the thesis that it is not *beliefs* but *changes of belief* that are in need of justification, and that such justification has to be given in decision-theoretic terms. It seems to me, however, that Levi's account is still only loosely connected with mainstream epistemology. This is a regrettable state of affairs, and one that should be finished soon.

ACKNOWLEDGEMENTS

This paper is a substantially revised version of a talk given at the Mainz conference on contextualism in September 2003. I want to thank Elke Brendel, Christoph Jäger, Michael Williams, and many other members of the audience for a stimulating discussion. Special thanks go to my commentator Lydia Mechtenberg and an anonymous referee of this journal for their valuable comments.

REFERENCES

Cohen, Stewart: 1988, "How to Be a Fallibilist", *Philosophical Perspectives* 2, 91–123.

DeRose, Keith: 1995, "Solving the Skeptical Problem", *Philosophical Review* 104, 1–52.
(Page references to the reprint in Sosa and Kim eds., pp. 482–502)

[27] A recent paper on 'Belief revision and epistemology' by Pollock and Gillies (2000) does not address the same problem. It rather compares the very special system of nonmonotonic reasoning invented by Pollock with standard belief revision theory, notes that the two do not fit together and puts all the blame on the latter theory. I think the divergence is easily explained using the terminology suggested in Rott (2001, chapter 3): Pollock takes the "vertical perspective" while standard belief revision theory takes the "horizontal perspective", and one cannot take two perspectives at the same time.

Descartes, René: 1641/1984, "Author's Replies to the Second Set of Objections", in John Cottingham, Robert Stoothoff and Dugald Murdoch (eds.), *The Philosophical Writings of Descartes*, Vol. 2, Cambridge University Press, Cambridge, pp. 93–120.

Fogelin, Robert: 1994, *Pyrrhonian Reflections on Knowledge and Justification*, Oxford University Press, Oxford.

Gärdenfors, Peter: 1988, *Knowledge in Flux*, MIT Press, Cambridge, Mass.

Gärdenfors, Peter: 1990, "The Dynamics of Knowledge and Belief: Foundational vs. coherence theories", *Revue Internationale de Philosophie* 44, pp. 24–46.

Gettier, Edmund: 1963, "Is Justified True Belief Knowledge?", *Analysis* 23, 121–123.

Goldman, Alan H.: 1987, "Nozick on Knowledge: Finding the Right Connection", in Luper-Foy (ed.), pp. 182–196.

Grove, Adam: 1988, "Two Modellings for Theory Change", *Journal of Philosophical Logic* 17, 157–170.

Klein, Peter: 1971, "A Proposed Definition of Propositional Knowledge", *Journal of Philosophy* 68, 471–482. (Page reference to the reprint in Sosa and Kim eds., pp. 60–66.)

Lehrer, Keith: 1965, "Knowledge, Truth, and Evidence", *Analysis* 25, 168–175.

Lehrer, Keith: 1990, *Theory of Knowledge*, Routledge, London.

Lehrer, Keith: 2000, *Theory of Knowledge*, second, substantially revised edition, Westview Press, Boulder.

Lehrer, Keith: 2003, "Coherence, Circularity and Consistency: Lehrer Replies", in Erik Olsson (ed.), *The Epistemology of Keith Lehrer*, Kluwer, Dordrecht, pp. 309–356.

Lehrer, Keith, and Thomas Paxson, Jr.: 1969, "Knowledge: Undefeated Justified True Belief", *Journal of Philosophy* 66, 225–237.

Levi, Isaac (1980), *The Enterprise of Knowledge*, MIT Press, Cambridge, Mass.

Levi, Isaac (2004), *Mild Contraction: Evaluating Loss of Information Due to Loss of Belief*, Oxford University Press, Oxford, forthcoming.

Lewis, David: 1973, *Counterfactuals*, Blackwell, Oxford.

Lewis, David: 1996, "Elusive Knowledge", *Australasian Journal of Philosophy* 74, 549–567. (Page reference to the reprint in D.L., *Papers in Metaphysics and Epistemology*, Cambridge University Press, Cambridge 1999, pp. 418–445.)

Lindström, Sten, and Wlodzimierz Rabinowicz: 1991, "Epistemic Entrenchment with Incomparabilities and Relational Belief Revision", in André Fuhrmann and Michael Morreau (eds.), *The Logic of Theory Change*, Springer LNAI 465, Berlin etc., pp. 93–126.

Luper-Foy, Steven (ed.): 1987, *The Possibility of Knowledge: Nozick and His Critics*, Rowman and Littlefield, Totowa, N.J.

Makinson, David: 1997, 'On the force of some apparent counterexamples to recovery', in E. Garzón Valdés et al. (eds.), *Normative Systems in Legal and Moral Theory: Festschrift for Carlos Alchourrón and Eugenio Bulygin*, Berlin: Duncker and Humblot.

Nozick, Robert: 1981, *Philosophical Explanations*, Belknap Press, Harvard University Press, Cambridge, Mass.

Plato: 1994, *Meno*, in *Plato's Meno in Focus*, transl. and ed. Jane M. Day, Routledge, London.

Pollock, John L., and Anthony S. Gillies: 2000, "Belief Revision And Epistemology", *Synthese* 122, 69–92.

Rosenthal, Jacob: 2001, "Einige Bemerkungen zum Gettier-Problem", *Zeitschrift für philosophische Forschung* 55, 540–555.

Rosenthal, Jacob: 2003, "On Lehrer's Solution to the Gettier Problem", in Erik Olsson (ed.), *The Epistemology of Keith Lehrer*, Kluwer, Dordrecht, pp. 253–259.

Rott, Hans: 2000, "Two Dogmas of Belief Revision", *Journal of Philosophy* 97, 503–522.

Rott, Hans: 2001: *Change, Choice and Inference*, Oxford University Press, Oxford.

Rott, Hans: 2003a, "Coherence and Conservatism in the Dynamics of Belief. Part II: Iterated Belief Change Without Dispositional Coherence", *Journal of Logic and Computation* 13, 111–145.

Rott, Hans: 2003b, "Lehrer's Dynamic Theory of Knowledge", in Erik Olsson (ed.), *The Epistemology of Keith Lehrer*, Kluwer, Dordrecht, pp. 219–242.

Segerberg, Krister: 1998, "Irrevocable Belief Revision in Dynamic Doxastic Logic", *Notre Dame Journal of Formal Logic* 39, 287–306.

Segerberg, Krister: 1999, "Default Logic as Dynamic Doxastic Logic", *Erkenntnis* 5, 333–352.

Shope, Robert K.: 1983, *The Analysis of Knowing – A Decade of Research*, Princeton University Press, Princeton N.J.

Sosa, Ernest, and Jaegwon Kim (eds.): 2000, *Epistemology – An Anthology*, Blackwell, Oxford.

Spohn, Wolfgang: 1988, "Ordinal Conditional Functions", in William L. Harper and Brian Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, Vol. II, Reidel, Dordrecht, pp. 105–134.

Spohn, Wolfgang: 2003, "Lehrer Meets Ranking Theory", in Erik Olsson (ed.), *The Epistemology of Keith Lehrer*, Kluwer, Dordrecht, pp. 129–142.

Williams, Michael: 2001, *Problems of Knowledge*, Oxford University Press, Oxford.