Croatian Journal of Philosophy

Articles

Vol. XV · No. 45 · 2015

Croatian Journal of Philosophy

1333-1108 (Print)
1847-6139 (Online)

Co-published by
"Kruzak d.o.o."
Naserov trg 6, 10020 Zagreb, Croatia
fax: + 385 1 65 90 416, e-mail: kruzak@kruzak.hr
www.kruzak.hr

and

Institute of Philosophy
Ulica grada Vukovara 54/IV, 10000 Zagreb, Croatia
fax: + 385 1 61 50 338, e-mail: filozof@ifzg.hr
www.ifzg.hr

Available online at http://www.ceeol.com and www.pdcnet.org

CROATIAN
JOURNAL
OF PHILOSOPHY

Vol. XV · No. 45 · 2015

Articles

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

Knowledge, Reflection, and Action[1]

ERNEST SOSA
Rutgers University, New Brunswick, USA

Our main topic is epistemic agency, which can be either free or unfree. This aligns with a distinction between two sorts of knowledge, the reflective and the animal. We first take up the nature and significance of these two sorts of knowledge, starting with the reflective. In a second section we then consider the nature of suspension and how that relates suspension to higher orders of meta-belief. Finally, we consider a distinction in epistemology between animal competence and reflective justification. All of these topics and distinctions are important for virtue epistemology, in ways to be considered.

Keywords: Epistemic agency, reflective knowledge, animal knowledge, suspension of judgment, reflective justification.

A. *Introduction: Agency and Reflection*

Our main topic is epistemic agency, which can be either free or unfree. This aligns with a distinction between two sorts of knowledge, the reflective and the animal. We first take up the nature and significance of these two sorts of knowledge, starting with the reflective.

Reflection has two aspects: first, reflection as careful, conscious thought, as meditation; second, reflection as thought that turns back on itself, as higher-order thought. The two aspects come together in Descartes's *Meditations*. And both are present also in British Empiricism, where reflection is the operation of the mind by which it is consciously aware of its own contents.

The two aspects are separable, since higher-order reference to one's own mind can be *subconscious*. So we can distinguish two degrees of reflection. A first degree involves the mind's turning back on itself, whether consciously or subconsciously. A second degree also requires the higher-order thought to be conscious.

[1] What follows will lay out and extend key aspects of the virtue epistemology expounded most recently in my *Judgment and Agency* (Oxford University Press, 2015).

Both aspects of reflection are important in epistemology—both its conscious and its higher-order character—each in its own way.

The importance of the *higher-order* emerges, first, with epistemic suspension of judgment. Suspension is constitutively a second-order mental phenomenon, or so I will argue shortly, in section B. The higher-order is also important, second, because of our aim to keep epistemic risk within proper bounds in our search for truth. This risk-assessment is inherently second-order. It is an assessment of how risky it would be to judge that p. Does the risk permit judgment, or does it require suspension instead?

So much for the *higher-order*. The importance of *consciousness* emerges in contexts of critical assessment, whether in private thought or in public dialectic, when one must weigh all pertinent reasons, while judging in light of the total evidence. Among the relevant reasons as one ponders a question is the fact that one already holds a certain belief on that question, if one does, even if it is just stored in memory. Take any context of critical assessment: that of legislation, for example, or of the courtroom, or the doctor's office, or the lab, or the criminal investigation, or the philosophy seminar. Questions in such contexts must be addressed by weighing all the reasons in view. Yes, the fact that one already believes a given answer is among the reasons in view. But it cannot be allowed to trump automatically any contrary reasons that may also come into view. Nor will the stored answer necessarily outweigh the newly available reasons *simply because of how very reliable is the perception-plus-memory that diachronically delivers that stored answer*. A believer who has forgotten just how reliably he acquired and retained his belief cannot now draw the belief from storage *with a weight determined simply by its diachronic reliability*. The believer now needs reason to self-attribute such reliability. And this self-attribution will be on the second order and also conscious.

Reflective knowledge of the highest degree involves reflective thought that is both conscious and higher-order. This is the *scientia* that Descartes takes as his epistemic aim. It is knowledge consciously endorsed on the second order as reliable enough, as belief whose correctness manifests superlative competence. Here we have both components required for appropriate reflection of the higher sort. A first-order belief is endorsed *consciously* on the *second order* as one whose correctness manifests superlative competence.

Should we also allow a lesser degree of reflective knowledge requiring no *conscious* awareness? Nearly all one's knowledge remains implicit at any given time, and not all implicit beliefs are epistemically on a par. Only some derive from proper risk assessment, for one thing, and from a proper grasp of one's relevant competence. Some are on an animal level, unaccompanied by sufficient assessment of risk or grasp of competence, whether conscious or subconscious. For example, the beliefs of a blindsighter ignorant of his competence fall short epistemical-

ly, despite their animal epistemic standing. We should thus recognize a distinction between knowledge that is merely animal and knowledge that is reflective even subconsciously. This distinction is epistemically worth drawing not just because the knowledge that is thus modestly reflective already seems superior *in that respect* to the knowledge that falls short. Our distinction is epistemologically significant also for a further reason: because main traditional arguments for skepticism threaten reflective knowledge irrespective of whether the second-order endorsement is conscious or subconscious.

What is that threat? We attain the reflective level of knowledge only when we self-attribute the competence manifest in the correctness of a first order belief. It is the ostensibly vicious circularity involved in seeking knowledge of such competence that poses the threat. The targets include first order sources like perception, testimony, and introspection, but also the faculties of armchair thought, such as rational intuition and deduction. Reflective knowledge would allegedly involve a vicious circle or regress because it requires second-order endorsement of the reliability of one's first-order sources. This gives us reason to distinguish between animal knowledge on one side, and reflective knowledge on the other, whether this latter derives from subconscious or from conscious endorsement. Either sort of reflective ascent would be blocked by the skeptic's argument.

B. *Suspension and Reflection*

1. *What is suspension?*

Often enough the right choice when we consider a belief is not to endorse it but to suspend judgment. Sometimes that is *epistemically* the right thing to do. What is involved in such suspension?

What *is* it to suspend judgment on a given question? Is it just consciously neither to believe nor to disbelieve while consciously considering the question? There are reasons to doubt that answer.
For one thing, that is what one does while still deliberating, undecided whether to suspend.

Secondly, what if one decides to suspend until further consideration? What about the suspending one does when the question has faded from conscious view? What constitutes one's suspending at that point, when one is neither believing nor disbelieving, nor even consciously considering the question?

Objection: "One *does* already suspend while still deliberating, so one does not really deliberate on whether to suspend. One deliberates rather on whether to *continue* suspending. As for suspension after one stops considering the question, this can just be a dispositional suspension, the disposition to forbear both affirming and denying upon considering the question."

Reply: Fair enough. But consider what happens when one *concludes* inquiry. This might happen when one has "conclusive" evidence to affirm, or to deny. When one concludes inquiry, as when one concludes armchair pondering, this might be because one settles into an attitude of positive affirmation, occurrent or dispositional, or of positive denial. One reaches a point where one finds it appropriate to settle into such an attitude indefinitely. If one does not reach that point, this might be for either of two reasons. One might be in a position to settle indefinitely into an attitude of double-omission, and this is the true suspending, the settled, conclusive suspending that might conclude inquiry in a way analogous to how affirmation or denial might do so. Alternatively, one might stop inquiry by deferring it, not by concluding it. Here one continues either active or dispositional *provisional* suspension. This is not the settled suspension that is one of the three ways of truly concluding inquiry, along with settled affirmation and settled denial.

2. *Is suspension always second-order, irrespective of whether it is provisional or settled? Suspension is plausibly a second-order phenomenon, moreover, whether the suspension is provisional or settled. Let us look into this.*

Compare first one's actions as one drives home while engrossed in conversation. There are many things one does intentionally while unaware that one is doing them. There are even choices that one freely determines without doing so consciously. When one puts on the brakes in response to a red light, one does so intentionally. One intends to do so, and one's intention *to put on the brakes* derives from the combination of a certain policy and a certain perceptual belief. It might be thought that one intentionally does so with no prior or concurrent intention to do so. Even when thwarted by stuck brakes, however, one at least *tried* to stop. And what is it to try if not to act on a present-directed intention?

One might thus view suspending as *intentionally neither-affirming-nor-denying*. Moreover, this can be similar to the actions we perform automatically and subconsciously while driving. Such actions answer to intentions that derive by some sort of reasoning from policies that implicitly guide us.

Epistemic policies can concern, among other things, the evidential requirements for proper cognitive attitudes. And the attitude of main interest here is not just confidence above a certain threshold. Nor is it just a stored state that can play its role unattended. Our focus now is rather on judgment, or on a disposition to judge, where to judge affirmatively is a distinctive all-or-nothing conscious mental act. Note well the distinction introduced. On one side is (a) the act of judgment, of conscious episodic affirmation, of affirming to oneself. On the other side is (b) a disposition to so judge. This latter is what we appeal to when we say of someone sleeping that "in his judgment" we should follow a certain course of action. We are not saying that he is at that moment performing the relevant mental act. We are saying rather that he would

perform that act if he were then to entertain the relevant question with the aim of answering correctly.

Action-guiding belief can remain on an animal level when, being inexpressible, it is inapt for proper reflective endorsement. It would be wrong to denigrate such deeply animal belief. Nevertheless, articulable judgment is obviously essential for a social, linguistic, and rational species. By judgment I mean, again, either the act of affirmation, in thought or speech, or the disposition to so affirm when sincerely aiming to affirm with truth. Something important happens when we conclude pondering and opt to judge (rather than suspend, or deny). Even if one has not yet voiced this judgment it is now ready for voicing, so that the information stored is suitable for sharing.

Suppose one *could* suspend without ascent to the second order while intentionally omitting both affirmation and denial. Even so, *proper* suspension is still plausibly second-order, since it cannot amount just to omission oblivious to the relevant risks. On the contrary, in order to be epistemically proper it must be properly responsive to such risks, which must be perceived adequately.

If proper suspension *is* plausibly second-order, since it requires intentions that target first-order conduct, also plausibly second order is then the judgment assessable epistemically *in the same sort of way as suspending is assessable epistemically*. Accordingly, such judgment would be *first-order judgment in obedience to second-order intentions that guide the believer implicitly*.[2]

Granted, such judgment does differ significantly from its paired suspending. On the first order there is not much, if anything at all, that constitutes the suspending. Of main relevance on the first order are simply *absences*: the absence of affirming/believing and the absence of denying/disbelieving. What positively constitutes suspending lies on the second order. It is the intending to not affirm/believe and not deny/disbelieve, or to persist in doing so. Perhaps the forbearing that derives from that intending is on the first order. I mean the forbearing constituted by (*intentionally) omitting affirming/believing and denying/disbelieving*. But the being intentional of this double-omission derives from something constitutively positive and on the second order: namely, the subject's *intention* (conscious or subconscious) to *not affirm/believe* and *not deny/disbelieve*. And this intention may implement an evidential policy (where the policy and the implied choice, the coming to intend, can be either conscious or subconscious).

[2] This makes problematic the well-known view that epistemic reasons for believing that p are just reasons that bear positively on the truth of p. We now see why the pertinent epistemic reasons must likely bear (also?) on the epistemic risk undertaken by one's own believing that p. As to whether *the* aim of belief is truth, a second-order aim here (also?) comes into view: The rational believer aims to take appropriate epistemic risk, no less and no more, in opting on the relevant threefold choice: affirming, denying, suspending.

3. *How general is the interest of second-order assessment?*

The interest of assessment on the second order goes beyond Pyrrhonian or Cartesian epistemology. Even in the most ordinary assessment of someone's judgment as epistemically justified or rational, the sort of rational justification involved is often on the second order, or is at least dependent on the second order, since it is justification of that judgment as superior to suspension, and concerns therefore a performance that is an alternative to suspension (or dependent essentially on an alternative to suspension). Such an alternative would be one among mutually exclusive and exhaustive options: affirming, denying, suspending. If suspending constitutively involves the intention to *not* first-order-believe and *not* first-order-disbelieve, then plausibly the relevant alternative affirming will constitutively involve the intention *to* first-order affirm, and the relevant alternative denying will constitutively involve the intention *to* first-order deny.

The believing (and disbelieving) of interest now can be either occurrent or dispositional. That is to say, it can be constituted either by an act of judgment or by a disposition to so judge. Either way, it is an *intentional* act or state. We have reached this result through the parity of such belief with its correlated suspending, and through the reasoning that supports our conception of such suspending as intentional double-omission, whether the intention is consciously episodic or subconsciously ongoing.

That is so even if ordinarily one's animal knowledge needs no reflective endorsement. Animal beliefs *can* often come under rational scrutiny, after all, in one or another setting where we focus on certain questions and beliefs, as in the law court, or the criminal investigation, or the doctor's office. And they can come under rational scrutiny more generally, as in the philosopher's reflection or seminar discussion. A belief under such scrutiny is not properly affirmed merely on the basis of its diachronic standing, no matter how excellent this standing may be. Synchronic justification is now required, which imports second-order assessment of first-order beliefs. Reflective assessment need not but often does rely on such second-order inquiry, properly so. We must rely on it, of course, when our *first*-order *competence* is itself explicitly under attack. But we can rely on it also when we more directly scrutinize the first-order *question* rather than the first-order competence. We may need to base our continuing first-order judgment on a positive view of our relevant first-order competences.

4. *What determines whether, on a certain question, suspending is epistemically justified?*

If the foregoing analysis is correct, what justifies our (intentionally) suspending is what justifies our intending to withhold belief and disbelief. What justifies our so intending? What might one endeavor to accomplish thereby? One cannot attain truth by *forbearing* from af-

firming/denying and from believing/disbelieving. One's objective in so forbearing is rather to avoid falsehood. One might conceivably pursue that objective by arbitrary suspending on the specific question. But the relevant objective is not *just* to avoid falsehood, but to do so properly, wisely, which requires attention also to the pursuit of truth. Cost/benefit analysis is required.

What are the costs and benefits relevant to epistemic choices? On a given question that one takes up, a main cost is false judgment or belief, a main benefit true judgment or belief. What is the likelihood that one will attain truth and avoid falsehood by affirming? What is the risk that one will fail instead? A justified attitude responsive to these queries, be it judgment or suspension, needs to manifest epistemic competence.

In arriving at the correct attitude one must assess one's level of *complete* competence with respect to the question addressed. This includes three components. The first is one's basic constitutional competence, one's skill in answering such questions. The second is one's current shape for employing that skill. Is one awake, alert, sober, etc.? Third and last is one's situation, including any relevant external relations. Is the light adequate? How far is the object? And so on. All three of these— skill, shape, situation—are constitutively involved in one's complete competence. Only such complete SSS-assessment (however quick and subconscious) can properly determine whether one is likely enough to answer the question correctly. A negative conclusion would require one intentionally to forbear from answering. Instead one would need to suspend.

One affirms with full epistemic competence, by contrast, only if the epistemic risk is competently assessed as low enough. One then affirms on a basis shared with the intention implemented: the basis provided by the favorable risk assessment. One hence falls short in so affirming unless that basis amounts to knowledge. It follows that the affirmation will itself fall short unless it amounts to reflective knowledge. In order to affirm properly, one must answer the first-order question correctly, manifesting thereby one's relevant first-order competence. Moreover, the exercise of that competence must itself be intentional, based on the second-order assessment of the relevant risk. Of course *this* assessment must in turn manifest sufficient competence.

Reflective epistemic status is therefore a status above animal-level fixation of belief, no matter how reliable the latter may be. This calls for an epistemology with both animal and reflective components. Reflective competence is required for the higher epistemic status. We need not always be seeking that status, nor is it a status required for proper trust in our first-order beliefs. Even when we need not, however, we often do seek that level of scrutiny and endorsement, as we consider a question in a setting that requires reflection. Plausibly enough, moreover, a belief would always attain a higher epistemic status if it *did* gain proper endorsement through such scrutiny.

C. *Animal Competence and Reflective Justification: the Interest of the Synchronic*

1. Much of our knowledge serves us well with no need of reflection. It would be practically inadvisable to scrutinize the trust we place on our own stored beliefs, or on the testimony of others, as we go through an ordinary day. Such quotidian trust is appropriately blind, unaided by reflection. Not so in the law court, or the legislature, or a detective's investigation, or the scientific lab or philosophy seminar. These settings call for critical scrutiny; it will not do just to voice our stored animal beliefs or to take on trust the say-so of others. Questions here present three options: affirming, denying, suspending. Opting properly on a question under reflective scrutiny requires a synchronic rational basis.

The fact remains that much everyday knowledge is not consciously reflective. Such animal knowledge permits—may even *require*—blind trust rather than consciously deliberate choice. This knowledge will often have been acquired competently, reliably enough, and will have been stored through competent retentive memory. Once stored, it will have done its work unseen, with no need of conscious attention. Much of our animal knowledge is acquired through normal childhood development, much absorbed from the culture. By contrast, the reflective knowledge of interest to us is normally attained through the conscious weighing of reasons.[3] Judgment-involving knowledge requires us to opt among affirming (or reaffirming), denying, and suspending. And this choice must be made in the light of the reasons available at that time. Suppose we can draw from storage an answer superbly acquired and retained. Even so, it would seem stubbornly irrational to just voice our belief despite the weight of synchronic reasons tilted against it.[4]

Judgment should be based on the total evidence available at the time. It must be so based in the law-court, the lab, the seminar room, and the criminal investigation. It takes priority when we must reason consciously to an answer for a question posed explicitly. The premises adduced in such reasoning ought not to be retrieved dogmatically from storage, not based just on the epistemic quality of the storage and retention. The first-order judgments that provide premises for critical reasoning require the rational support of any evidence synchronically available. Such rational support is required both for private thoughts in conscious reasoning, and also for public assertions whereby we convey information to others by speaking on our own behalf.

True, we can often endorse what memory delivers if nothing in view tells against it. "Methodological conservatism" is thus right to bestow squatter's rights on beliefs already in storage. Nevertheless, counter-evi-

[3] Through "ratiocination"—to use Wittgenstein's term in *On Certainty*.

[4] Note well: "reasons tilted *against* it." Again, this is supposed to be so despite whatever reason the believer may have—and it may be quite considerable—for conservatively trusting his own belief on the subject matter involved.

dence synchronically in view might still properly trump the conservative claim of the belief in storage *even when such synchronic evidence is far less reliable than the diachronic process that lies behind the stored belief.*

2. Again, plenty of beliefs are initially acquired through competent introspection, perception, or reasoning, and then stored in memory. Such a belief will often linger even after you forget how it was initially acquired. You might then be able to say little more than "I just remember." What then is the later standing of that belief?

We are focused on a time late in the life of the belief. No-one can now detail how it was acquired and retained. Suppose only slight direct evidence is now available for its content. If we go by this evidence, the belief no longer counts as justified. How competently can you now retain it? Its epistemic standing will now depend essentially on two things about your memory: first how good it is for that sort of belief, second how well qualified you are to assess it on that occasion. You must now assess how well your belief is likely to have been acquired and sustained. But your full competence for this second-order assessment might be inferior to the competence that yields the belief itself on the first order. First-order competence will often combine excellent perceptual acquisition with excellent mnemonic retention.

What then is the believer to do as time passes? Should confidence dwindle in tandem with reduced qualification to endorse? Consider the steady decay of the information required for endorsement. Despite such second-order weakening, the believer's retentive memory can remain strong indeed. The retained belief is very probably true, given the perception that originally produced it and the memory that has kept it securely stored. That belief may thus constitute first-order, animal knowledge of the highest quality. Quite often what decays over time is just the reflective, second-order perspective.

Here is an example. At noon on a certain date you are mistreated as a child. You know extremely well that it is noon on that date. You store that belief for years, retaining it through excellent memory. In general people would not remember so well. In general your own memory may not work so well. But it does in this case, on this sort of subject matter. That event stands out in your mind, and your memory of it is outstanding. The perception-plus-memory manifest in your continuing belief is of the highest quality. Compatibly with that, your second-order competence can decay. Just based on common sense, you may come to doubt your memory of that event. You may even learn that ostensible memory of such mistreatment is far less reliable than common sense had supposed. Human beings in general do not recollect as reliably as had been thought, especially not on *such* subject matter. By hypothesis, however, *your* memory *is* in this case extremely reliable.

That is one example of the phenomenon I wish to highlight. But we need not invoke abnormal powers. Another example might involve just normal human perception and memory. In combination these might

lead reliably to a correct present belief, even if the believer is now unable to detail how he acquired and retained his belief. He knows who directed a certain film but cannot detail how he acquired that information, nor how reliable his source may have been.

There can thus be a clash between the diachronic and the synchronic, either of which can be excellent while the other is poor.

3. Our puzzle does not arise merely from a clash between externalist reliabilism and internalist evidentialism. The important clash is between two epistemic statuses that a belief can have:

> First, there is the status a belief gains *diachronically* through the subject's thinking and memory, no matter how internal such thinking and memory may be over time, nor how internal the subject's initial data may have been.
>
> Second, there is the status a belief gains *synchronically* through the support of reasons present to the thinker's consciousness at that very moment.

The problem thus transcends two familiar divides: one between externalism and internalism; another between evidentialism and reliabilism.

Such disparity between animal quality and reflective quality would involve a divergence between

> first, the high status a belief derives diachronically from a retention-involving first-order competence, and
>
> second, the lower status that same belief might have synchronically, due to the diminished epistemic quality of the believer's second-order competence, because it is less reliable or anyhow less productive of justification.

Earlier we considered examples of long-term memory. A similar example involves arithmetical calculation. You may doubt your ability to perform a complex addition without flaw (despite performing it flawlessly). Although initially you may have believed the result without a second thought, doubt sets in when you recall how unsure you are of your competence.

Suppose your reflective capacity to endorse a given first-order belief is thus diminished. What about your *judgment itself,* on the first order? Here are questions on the first order as you view a hand, or a fire: Is this a hand? Is that a fire? What attitude should you adopt on such questions within the privacy of your own thought, and what can you properly assert to others? Judgment, rather than suspension or inattention, is required for conscious reasoning, and for proper assertion when speaking in your own person.

Two issues thus arise concerning a pondered first-order question. First, how if at all should it be answered? What attitude should you adopt from among the relevant three: judgment, denial, suspension? Second, how should you assess epistemically whatever attitude you do

adopt? For example, what determines the epistemic standing of your first-order affirmation/belief or denial/disbelief? Is it the quality of your total first-order competence, including its diachronic components? Or is it rather the quality of your synchronic rationale, including the contribution of your second-order competence to exploit what is synchronically available to it? These two ways of assessing a first-order judgment might differ dramatically, since the two sets of factors can differ greatly in epistemic quality.

Again, it is not just the *assessment* of a first-order judgment that may derive either from a first-order animal competence or from a second-order reflective competence. There is also this question: *Which perspective should have priority in determining how to judge on the first order?* Should you trust your excellent first-order competence, or should you trump that competence once your belief is under scrutiny, with the inevitable bearing of the second-order perspective? Should you now decide whether to trump based on *all* the reasons *presently* available to you for conscious consideration?

Suppose we give priority to the reasons presently available. This in effect recognizes the bearing of a kind of reflective knowledge, involving a second-order judgment (or disposition to judge). This second-order judgment itself depends for *its* standing on the quality of the competence that *it* manifests. Reflective knowledge will thus enable your conscious reasoning, and sustain your place in the community as testimonial transmitter. Such reflective knowledge is constituted by a judgment (or disposition to judge). Indeed it is regularly constituted by two such attitudes: one an endorsing judgment on the second order, and the other a judgment on the first order—whether disposition or act. These attitudes often figure in our conscious reasoning, and in our sincere assertion when we inform others.

We have been considering this question: What should determine one's act of judgment, and one's disposition to so judge? Is it *diachronic* competence, even if its initial inputs have long receded, or is it rather the reasons *synchronically* available and operative? If we opt for the present-time-slice, we upgrade the second-order perspective. This is because synchronic reasons for stored beliefs are so often to be found within that perspective.[5]

[5] Recall the examples wielded by internalists against reliabilist externalism, such as BonJour's Norman, the clairvoyant-out-of-the blue, and Lehrer's Truetemp, unaware of the thermometer embedded in his brain. In these cases too a belief derives with high reliability from some process or faculty relevantly beyond the subject's awareness. Here again reliability clashes with rationality. But there is a significant difference between those cases and our case of diachronic/synchronic clash. Our clash does not occur in remote, contrived examples. It is rather a familiar and pervasive feature of everyday cognition. Moreover, this diachronic/synchronic clash does not reveal a deep, unbridgeable chasm. On the contrary, consider the knowledge of the blindsighter, and our knowledge of simple math or logic. These cases plausibly suggest that rationality itself is to be explained at fundamental levels by appeal to relevant, reliable competence. Compare even a familiar bit of knowledge that

We have focused mainly on an important sort of belief that is not just a degree of confidence above a certain threshold, nor just a stored state that subconsciously guides behavior, as when one conducts everyday business on automatic pilot. Instead, the belief of interest to us is a judgment. It need not be an episodic conscious affirmation. It might be the sort of judgment that one attributes to someone when one says "In his judgment, p." This attribution can be correct even if the subject is not at that moment affirming that p. A judgment can be constituted rather by a *disposition* to affirm if sincere, to oneself or to others. This is a disposition to affirm if under the influence of no conscious aim beyond answering the question correctly. Important synchronic reasons for or against such belief will often reside on the second order. What you can consciously affirm depends on your synchronic rational basis. Endorsement of your stored beliefs may now turn on how well you can defend the quality of your acquisition-plus-storage.

4. Justified judgment will thus involve your second-order competence to assess your first-order competence. As memory dims on how you initially acquired your first-order belief, you must increasingly rely on your epistemic self-trust. Suppose the first-order belief to be put in doubt, either through overt disagreement, or through a challenge to your relevant competence. In responding you need to defend your competence. You must now defend your belief from a second-order perspective on your relevant first-order competence. After all, how properly you endorse that first-order belief is determined by the reasons you may now have in view. A major portion if not the whole of this rationale will include whatever you can adduce in favor of your relevant first-order competences, and will reside on the second order.

you might have some morning: namely, that more than two seconds have elapsed since you awoke. You can have this knowledge even without having looked at any timepiece. Your belief is surely rational, moreover, even with no rational basis on the first order. The corresponding seeming, the inclination to believe, is itself rational and justified despite its lack of rational basis. What could possibly give it this status (an epistemic status withheld from a bigot when someone seems dumb to him based just on facial appearance; and withheld also from a gambler to whom it seems that 7 or 11 will come up next).

What matters for human rationality is whether the relevant competence is a fundamental component of the human cognitive structure. And this reveals a further reason why human diachronic competence is epistemically important. Human diachronic competence differs importantly from the competences distinctive of Truetemp or of clairvoyant Norman. Human diachronic competence *is* after all a fundamental component of the human cognitive structure. For example, it can simply involve a familiar combination of basic perception with retentive memory. Yet it can clash with the rationale synchronically available to the subject at some later time. In order to proceed rationally, the subject must favor what is then, at that later time, available to his synchronic consciousness. The defense against diachronic reliability must in this way go beyond the defense against clairvoyant or Truetemp reliability. It must now appeal not only to what is fundamental to the human cognitive structure. It must *also* appeal to *synchronic, conscious* factors.

A belief that is apt through diachronic competence falls short if it is *not* endorsable synchronically through the balance of available reasons. In that case you cannot rely blindly on your stored belief and on the diachronic competence that sustains it. Often enough only reflective knowledge can fully serve our needs as conscious reasoners and speakers.

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

Knowledge and Conditionals of (Dis)connection[1]

DANILO ŠUSTER
Department of Philosophy, Faculty of Arts,
University of Maribor, Slovenia

The gist of modal epistemology is expressed in the idea that you fail to know if you do believe truly but it is seriously possible for you to believe falsely. According to subjunctivism, this idea is captured by certain subjunctive conditionals. One formulation invokes a safety condition—"If S had believed P, then P would have been the case," while the other invokes a sensitivity condition—"If P had been false, S would not have believed that P." According to simple subjunctivism, such conditionals do not contrapose and Sosa derives important epistemological consequences which favor safety from this difference. However, simple subjunctivism is inadequate. I return to Goodman and his analysis of factuals and propose modal stability, which is restricted sensitivity or enhanced safety as a proper epistemic condition for the non-accidental connection between the basis for the belief and the relevant facts of the matter. The idea of modal stability combines robustness (benefits of safety) with responsiveness to facts (benefits of sensitivity) and recovers the original motivation for the relevant alternatives theory—when testing for claims of knowledge that p we ask what might be the case if not-p, but we ignore irrelevant possibilities. Epistemic modal conditions should be expressed in terms of conditionals of connection which contrapose within the limits of relevance.

Keywords: Modal epistemology, safety, sensitivity, Sosa, conditionals, Goodman, contraposition.

1.

"Even if p had been false, you would have believed it anyway," looks like a good reason for denying that one knows that p. Formulated in

[1] The first (but very distant) version of this paper was presented at the June 2007 Bled (Slovenia) philosophical conference "Epistemology." An expanded version was presented at the November 2009 workshop "Intentionality" (Philosophy Department of the University Graz, Austria). I would like to thank the participants for their comments.

positive terms, one could paraphrase a famous line from Shakespeare: belief counts as knowledge only if it "alters when it alteration (in facts) finds."[2] Belief (as opposed to, perhaps, love), if it is to be promoted to knowledge, should be finely dependent on facts. However, how fine is fine enough and what if one did not easily believe that p without being right? It is tempting to say that belief can still be a good candidate for knowledge, even if it does not *always* alter when it alteration finds. Only when it is seriously possible for you to falsely believe that p, do we have a good enough reason for denying your knowledge that p. Alternation within limits looks like a plausible reconciliation of these two intuitions. What counts as *serious* is usually specified in terms of modal distance—how far is a certain possibility from the actual world. Modal distance maps the strength of the connection between one's belief that p and the fact p—if the possibility of believing falsely is close, then there is no connection, or a very unstable one. Within a space of nearby possibilities, modal connection requires that a belief should "alter when it alteration finds," as for very distant possibilities, well, even love will shatter should your beloved one be transformed into a monstrous vermin. The old Platonic insight that knowledge is at least a non-accidentally true belief is then interpreted as the idea that knowledge requires a modally stable connection between the belief and the relevant facts of the matter.

This picture is rough but familiar. Many versions of contemporary reliabilism impose modal conditions on knowledge, understood as a certain necessity, non-accidentality, or anti-luck condition. According to Unger (1968), S knows that P just in case it is "not at all accidental that S is right about its being the case that P." It not being accidental that one is right about P amounts to there being something in one's situation that guarantees, or makes it highly probable, that one were not wrong. *Externalism* in general is sometimes expressed in these terms: you can know something noninferentially, without reasoning from prior knowledge, so long as it is no accident or coincidence that you are right (Sosa 1997: 419). The notion of *serious* possibility comes from Goldman's development of reliabilism with respect to perceptual knowledge (1976: 775). The position I develop in this paper is, in a certain sense, simply a generalization of Goldman's (1986: 46): "A true belief fails to be knowledge if there are any relevant alternative situations in which the proposition p would be false, but the process used will cause S to believe p anyway."

According to Zalabardo (2012: 4) "Reliabilism ...is the view that whether a true belief has the status of knowledge depends on how the natural order connects the state of affairs the belief consists in with the state of affairs whose obtaining determines the truth value of the belief—that is, S's belief that p with p." I agree with the centrality of

[2] Love is not love / Which alters when it alteration finds (Shakespeare, Sonnet 116)

connection—knowledge must be underwritten by some kind of stable modal connection. What is the profile of this connection? First, what are the relata? The state of affairs or the fact that *p* is more or less uncontroversial, but the epistemic *relatum* is not a simple belief that *p*. Various proposals include: evidence for *p*; (conclusive) reason for the belief that *p* (Dretske 1971); experiential reasons for the belief (Sosa 2004a); a basis for the belief (Sosa 2002); a belief based on a method which indicates that *p*; or simply a method of belief formation (Luper 2012: 210). Relativisation is necessary for various reasons. Here is simple a case based on Luper (based on Nozick). I believe that Mary, my daughter, is well because I see her playing tennis, but if she weren't I'd believe she was through wishful thinking. Suppose she just made a dangerous move, overstretching her arm, which might easily have led to a ruptured biceps. My true belief about her well-being is based on perception, but (in the case of her injury) I might easily believe falsely, by way of wishful thinking. Still, the possibility of my using a bad method does not discredit my perceptual knowledge of her well-being.

The possibility of applying a different method of belief formation suggests that non-accidentality might be understood as having multiple dimensions (cf. Yamada 2010).[3] Bogardus (2012) presents a nice analogy with bridges—methods of acquiring knowledge are like bridges to one's destination. If Godzilla is rampaging through the area, even the world's sturdiest bridge might be in danger. It may be false that, were one to take the bridge, one would arrive at one's destination. Yet, if Godzilla has not yet hit the bridge, it remains as sturdy as you like. In one dimension it is no accident that you safely crossed the river using this bridge; in another dimension you were simply lucky. Various cases of accidental knowledge which are supposed to show that modal condition is not necessary for knowledge might in the end just show that the one-dimensional space of possible worlds (measuring the proximity of a simple belief-fact mismatch) is just not pliable enough to map all of our epistemic intuitions.

I will avoid complications, however, and mainly work with the simple idea that knowledge must be underwritten by some kind of stable modal connection between the truth of the belief and the epistemic basis for the belief. The modal nature of the connection indicates that actual true belief is insufficient for knowledge; true belief in some range of counterfactual situations is also required. Which range? Two different ways to characterize them have been extensively discussed. Here is Black (2011: 189):

> First, to say that Smith's belief is true simply as a matter of luck might be to say that there is nothing about Smith's circumstances, in which his belief happens to be true, that ensures that he will believe that C—even if C had been false, Smith might nonetheless have believed that C. This way of giving expression to our anti-luck intuition corresponds to epistemologies known as *sensitivity* theories...

[3] Engel (1992) already introduces different dimensions of epistemic luck.

> Next, to say that Smith's belief is true simply as a matter of luck might be to say that there is nothing about that which led Smith to believe that C that ensures that C will be true—it might have been that Smith's circumstances are just as they actually are, but that his belief that C is false. This way of giving expression to our anti-luck intuition corresponds to modal epistemologies known as *safety* theories...

Well, what is the *difference*? Lucky belief, only accidentally connected with facts of the matter, in terms of safety, is a belief held by Smith such that the basis for Smith's belief that C does not ensure the truth of C. It is possible that Smith's circumstances are just as they actually are, but that his belief that C is false. Not quite so. Smith's circumstances are not exactly as they actually are; in this possible situation the facts about C must have changed for his belief to be false. Which is to say: if C had been false, Smith might nonetheless have believed that C on the same basis (in *roughly* the same circumstances). So his belief is not sensitive. This looks very much like a difference without a difference.

I do not deny that it is possible to understand the passages quoted in a way which does not minimize the difference. Still, I will argue that the two characterizations of one's failure to know are just two perspectives on a broken *liaison*: (i) Smith's basis for his belief that C does not *depend* on the fact that C; (ii) Smith's basis for his belief that C does not *indicate* the truth of C. A standard way to spell out the difference is to express the two conditions in terms of subjunctive conditionals. According to *subjunctivism*,[4] a true belief rises to knowledge just in case certain subjunctives about the truth of that belief hold. A belief that p is safe iff

SAF S would believe that p only if p were true. (Alternatively: S would have believed that p only if p had been true.)

And a belief that p is sensitive iff

SEN If p were false then S would not believe that p. (Alternatively: If p had been false then S would not have believed that p.)

SEN is almost universally interpreted in accordance with the Lewis-Stalnaker account of such conditionals: in the closest (sometimes the nearest) world in which p is false, S does not believe that p.[5] I will call a combination of the SAF, SEN and standard possible worlds interpretation of the subjunctives involved *simple subjunctivism*. Simple subjunctivism cannot be quite right—when p and q are both true, it is an artifact of these kind of semantics that "if p were the case, then q would be the case" is automatically true, since the actual world is more similar to itself than any other world (and q is true in the actual world). True beliefs will then automatically be safe—not what we want, no luck eliminated, so a different interpretation has to be used

[4] A term used by Fogelin (1994) and Vogel (2007).

[5] Cf. Black (2011: 189), Alspector-Kelly (2011: 129); Vogel (2012: 122) as a sample list.

for safety (true/true) conditionals. However, I think one should play the game of epistemic modal conditions consistently and with the same set of rules, motivated by the idea of modal connection. Although a good deal of thought has gone into the details concerning how the safety and sensitivity conditions should be interpreted, this point has not been appreciated sufficiently.

Consider the notion of the closest possible world(s). What exactly counts as *the closest* is open to discussion, but some have taken close possible worlds to be *nearby* worlds (Rysiew 2006: 275). This may lead to confusion—the closest need not be nearby. The closest non-human inhabited planet might be in the Gliese 1 system, so far, far away, but still closer than, say the Gliese 876 star system. However, there are no *nearby* non-human inhabited planets: there are none in our Solar system and its vicinity, Alpha Centauri, for instance. When speaking about sensitivity in terms of simple subjunctivism, I will use the notion of *close* worlds (which need not be nearby), and I will reserve the notion of being nearby for those worlds which are "really" near (in roughly the way that Venus is near to Earth and Gliese 1 is not; the question of how to measure distances in modal space is a vexed one).

Let me now state the problem for sensitivity understood in terms of simple subjunctivism: in some cases the closest non-p possibility is not a serious option at all and in some cases the closest non-p possibility is not the *only* serious option. Radical skeptical scenarios belong to the first type—we are told, for instance, to imagine the remote possibility that at this very moment we are a brain hooked up to a sophisticated computer program that can perfectly simulate experiences of the outside world. Is this possibility nevertheless relevant for my actual knowledge that I am not a victim of a brain-in-a-vat illusion? I do not think so, although this is controversial. I will touch upon this again later.

Goldman (1983: 84) illustrates the second case with the following scenario:

> Sam correctly believes that Judy is before him, but if it were Judy's twin sister instead, he would mistake her for Judy. Then, as long as the twin sister's being there is a serious possibility, Sam doesn't know that Judy is before him. Suppose that what would be the case if Judy weren't before Sam is that nobody would be there, and if nobody were there, Sam wouldn't believe that Judy is there. Then Sam's bid for knowledge survives Nozick's condition (3), and nothing else in the analysis is able to defeat it.

Nozick's condition (3) is just SEN. It is clear that *more* no-Judy possibilities should be considered, not just the closest one and only. It might, after all, be merely a happy accident that Sam sees Judy—the connection between the basis of his belief and Judy's presence is not stable. The profile of modal connection is not adequately captured by the sensitivity condition as stated by simple subjunctivism. Goldman objects that Nozick's analysis does not make reference to serious possibilities; it talks about what would be the case if p weren't true. It is not so

clear that conditionals expressing modal connection should be about the closest possibility only—sensitivity is a technical notion, after all. But I agree that an overly simplistic understanding of the subjunctives involved obscures the nature of modal connection. A more sophisticated semantics is usually offered,[6] but then simple subjunctivism faces the problem that *explanans,* an adequate account of subjunctive conditionals, might turn out to be more complicated than *explanandum*, the analysis of epistemic modal connection.

Let me try a different approach. I started with the intuition that a proper modal connection is a necessary condition for knowledge and this necessity was formulated in terms of a conditional of *dis*connection (let us ignore, for the time being, the problem of the epistemic *relatum*):

SEM Even if *p* were false, S would believe it anyway.

SEM is a semifactual denying that S knows that *p*. Goodman (1991: 11) introduced the phrase 'semifactual' for a conditional with a false antecedent and true consequent in contrast to the 'counterfactual' in which both the antecedent and consequent are false. For Goodman, in practice, full counterfactuals affirm while semifactuals deny that a certain connection obtains between antecedent and consequent. Literally, however, a semifactual and the corresponding counterfactual are not contradictories but contraries, and both may be false. SEM has the force of denying "if *p* had been false, you would not have believed it," but it is actually stronger than required for a denial of SEN. Both Nozick (1981: 199) and Dretske (1971: 9–10) were aware of the fact that a proper denial of SEN is:

> If *p* were false, then S might still believe that *p*. (Alternatively: If *p* had been false, then S might have believed that *p*.)

So,

D1 S might have believed that *p*, (even) if *p* were false.

is enough to deny knowledge. How about:

D2 *p* might be false, even though S were to believe that *p*?

I think that the denials are equivalent within the same range of nearby worlds. Let me use the following example as an analogy for an accidental correlation (*Wikipedia*): "Since the 1950's, both the atmospheric CO_2 level and obesity levels have increased sharply." But atmospheric CO_2 does not cause obesity, rather, richer populations tend to eat more food and consume more energy. So we have:

D1' Obesity might increase even though the level of atmospheric CO_2 would not.

But also:

[6] Nozick (1981: 680–681) avoids the trivial truth of factuals by adopting a rather complicated semantics for true-true conditionals.

D2' The level of atmospheric CO_2 might not increase even though obesity would increase.

Both conditionals are true in the same possible situations, those in which obesity increases but the level of atmospheric CO_2 does not. Still, the perspective is different: we tend to understand the first conditional as stating that the level of obesity does not track (is independent of) the level of atmospheric CO_2. We *could* understand the second conditional as stating that the level of atmospheric CO_2 does not track (is independent of) the level of obesity (just think of cows and methane they produce). However, given certain background assumptions about the direction of causation (from CO_2 to obesity) it seems more plausible to suppose that D2' states that the level of obesity is not a good *indication* of the level of atmospheric CO_2.

In terms of epistemic conditions, D1 is officially a denial of "sensitivity"—our bases for beliefs do not track (are independent of) the relevant facts. D2 is a denial of "safety," our bases for beliefs do not indicate the relevant facts. According to one of Sosa's formulations (Sosa 2002), for a belief to be safe it must be based on a reliable indication. Or, to use the formulation of safety by Black, there is nothing about that which leads us to have reasons for the belief that *p* that ensures that *p* will be true. Still, the "truthmaker" for D1 and D2 is the same—a (relevantly similar) possible situation in which one falsely believes that *p*. D1 (dependency) and D2 (indication) are two perspectives on a modal disconnection representing two aspects of a proper epistemic connection: stability and responsiveness to facts. Safety corresponds to stability and robustness—modal states that concern what could not *easily* have happened (cf. Williamson 2000: 123). The aspect of sensitivity is the aspect of responsiveness to changes and dependency on facts—knowledge attribution depends on whether S's belief that *p* (or, the cognitive processes responsible for the production of that belief) is responsive to whether *p*, whether S would believe that *p* even if some not-*p* alternative were in fact the case (Goldman 1976: 85). Both are conditions of reliability and they should extensionally co-vary—within reasonable *limits*!

Responsiveness to changes is often explained in terms of the subject's possession of certain discriminative capacities (cf. Rysiew 2006). Let us therefore take an underwater camera as our analogy for a cognitive, belief forming mechanism. There are two aspects of a good picture taking mechanism: (i) responsiveness to environment (making good and recognizable pictures in variable conditions); (ii) good housing, robustness, stability—the camera should not break easily. The first requirement corresponds to sensitivity and the second to safety. A camera which easily breaks is unsafe and not a good camera and neither is a camera with poor sensitivity that produces blurred pictures. It is true that 11.000 m under the water the camera might become dysfunctional. It is also true that our belief forming mechanisms become

dysfunctional when subjected to radical skeptical scenarios. However, these are extreme conditions, as a camera that does not work 11.000 m under the water can still be a good underwater camera. Discrimination under not too-farfetched conditions makes for a solid camera, while discrimination within reasonable limits of relevant possibilities is good enough for knowledge (or better, necessary, as there might be other conditions for knowledge).

Note that the failure of stability ("unsafety") and the failure of responsiveness ("insensitivity") are usually *explained* in the same way. Consider the familiar Russell-Gettier case in which Mrs. Smith forms a true belief about what the time is ("It is 5 o'clock in the afternoon") by looking at a stopped clock, one that just happens to be showing the right time. In the closest not-5-o'clock world Mrs. Smith would still believe it is 5-o'clock; her belief is not sensitive. Why? Because she is unable to discriminate the "5 o'clock in the afternoon—world" from the relevant but incompatible hypothesis. Also, Mrs. Smith could easily believe it is 5-o'clock even if the time were different; her belief is not safe. Why? Well, based on the evidence she has in these circumstances, Mrs. Smith is unable to discriminate between 5 o'clock and the relevant not-5 o'clock worlds. She is both easily prone to error and her belief is not responsive to facts because of her inability to tell apart the relevant alternatives.

In making these judgments, we should take into account those seriously possible situations in which she falsely believes that *p*, but *not* those in which she fails to believe truly. Many have noticed that it is the modal proximity of a false belief that matters for the epistemic (dis) connection, not the proximity of undetected truths. Sosa (2004b: 280) gives the following example:

> If I see a large pelican alight on my garden lawn in plain view, I will know that there is a bird in my garden. And this is not affected by the fact that a small robin sits in the garden in its nest out of view. In such circumstances, there might very easily have been a bird in the garden without my believing it.

One might object that my belief that there is a bird is based on my pelican-experience, and there could not easily have been a bird in the garden without my believing it on this *basis*. How, exactly, should we individuate bases (reasons, methods)? This is a difficult problem that I will avoid; let me just say that the "pelican-experience" seems to be too fine-grained. Yet, there are many other cases. Luper (2012: 212) argues convincingly that we can know things based on reasons (methods) that miss instances (*p* is true but our method of knowing fails to indicate *p*), counter-instances (*p* is false but our method of knowing fails to indicate the falsity of *p*), or both. He gives the case of a gappy thermometer: if a person's temperature is over 101°F, then the thermometer indicates a fever. People are feverish when their temperature is not over 101°F, but the thermometer will not indicate that such persons have fevers.

Suppose Frieda's temperature is slightly over 101°F and the thermometer indicates this fact. She might easily have 100.8°F degrees and be feverish without the thermometer indicating this. Still, using this thermometer, we know she has a fever.

What counts as a denial of knowledge is the possibility of believing falsely, not the possibility of ignorance. We should avoid *errors* in similar cases as Williamson would say, but our potential failures to form beliefs do not discredit our claims of knowledge. This observation unites a theory of modal (dis)connection with its predecessor, the original relevant alternatives theory. If you know that p, then, in some sense, you "can't be wrong" about p; there are no *relevant* possible situations in which you have the same basis for your belief, but p is *false*. Or, in other words, if it is seriously possible for S to falsely believe that p, then S fails to know that p.

Serious possibility is determined by modal distance, and it is hoped that modal distance captures relevance (modally far-away is irrelevant). Ideally, serious possibility, relevance and modal proximity should match. When in an anti-skeptical mood, philosophers tend to agree that irrelevant (radical skeptical) scenarios are far-away and not seriously possible, so they do not have an adverse impact on everyday knowledge. However, a theory of relevance is everybody's problem. We might end up with a cheese like topology, a set of relevant possibilities that cannot be naturally seen as a "sphere" of nearby possible worlds but rather as a set of unconnected "islands" (Schaffer 2005: 125). Let's hope that this is not the case and work with the usual, not entirely precise notion of relevance (for Goldman 2012: 69, for instance, a situation is relevant only if it is "realistic," fairly likely to occur, or does occur in a nearby possible world).

3.

Simple subjunctivism is not an adequate formulation of epistemic modal connection. We could avoid conditionals altogether and state modal conditions directly, by means of a possible-worlds heuristic.[7] Or, we might try with a semantics that allows for a uniform treatment of modal conditions and interpret epistemic modal conditions in terms of a special type of conditional. I am sympathetic to all of these approaches. Ideally, they should converge—I will try to rehabilitate a conditional of connection introduced by Goodman which allows for a uniform treatment of factuals and counterfactuals and has truth-conditions which respect our epistemic modal intuitions when expressed in terms of possible worlds.

We ascribe a lack of knowledge that p to S when p is true and S believes that p on a basis b, but continues to believe that p (on this basis) in one of the seriously possible non-p worlds. In this case a *semi-*

[7] This is proposed by Greco (2012: 194), who refers to Hawthorne.

factual, a conditional of disconnection, is true: "Even if p were false, S might still believe that p on basis b". For Goodman, in practice, full counterfactuals affirm while semifactuals deny that a certain connection obtains between antecedent and consequent. However, the same connection can be expressed in terms of a *factual*. Goodman considers the case when we say of a piece of butter that was eaten yesterday and that had never been heated: "If that piece of butter had been heated to 150°F, it would have melted." He then remarks (Goodman 1991: 10):

> The problem of counterfactuals is equally a problem of factual conditionals, for any counterfactual can be transposed into a conditional with a true antecedent and consequent; e.g., "Since that butter did not melt, it wasn't heated to 150°F." That "since" occurs in the contrapositive shows that what is in question is a certain kind of connection between the two component sentences; …

This may or may not be the proper approach to analyzing counterfactuals in natural language, but it certainly looks like a good starting point for analyzing conditionals of *connection*. "Since X, Y" captures Sosa's idea of safety, "Not easily X without Y." In the same way as not-melting of the butter indicates its not-being heated, modal connection requires that the existence of S's reasons (evidence, basis) for the belief that p —the way an object appears to S, for example—is a reliable *indicator* of the truth of p. For Sosa an agent S counts as knowing p only if S believes p by way of a safe "indication," where indications are deliverances of epistemic sources such as perception, memory, inference, etc. (Sosa 1999: 149 and Sosa 2002). Occasionally, he uses a different formulation: a belief that p is basis-relative safe, if and only if it has a basis that it would (likely) have only if true. The idea, however, is the same: the basis for p indicates the truth of p, or, in terms of a conditional: "Since S believes that p on basis b, it is true that p." Dretske's (1971) notion of a conclusive reason R (S knows that P on the basis of R) could also be understood as "since R is the case, so is P."

Not-melting of the butter does not cause its not-being heated, and reasons do not cause facts, so I will understand Goodman's "since" conditionals neutrally as expressing a connection without direction. They can be used causally. McCall (1983) uses "since" and "because" interchangeably to indicate a connection between the antecedent and consequent and analyses a factual: "Since the butter was heated it melted," which we naturally read as stating that the antecedent causes the consequent. Let "A > B" stand for the conditional such that the proposition or state of affairs expressed by A bears a "connection" of a logical or nomological nature to that expressed by B. When A and B are both true, there are two conditionals of connection—a factual: "Since A, B," and a corresponding contraposing counterfactual: "If B were not the case, then A would not be the case." They are extensionally equivalent according to Goodman and McCall. *Epistemic* conditionals connect a basis for a belief that p ('$B_b p$' is a proposition stating that S believes that p on a basis b) with the fact that p and I will understand 'Since $B_b p$,

p' as an *indication* ($B_b p$ indicates *p*) and 'If *p* were not the case, then $B_b p$ would not be the case' as *dependency* ($B_b p$ depends on *p*). The first conditional expresses the aspect of safety, and the second the aspect of sensitivity—but these are, following Goodman, just two aspects of one connection, which is to say that indication conditionals and dependency conditionals validly *contrapose*.

This is not a standard view, however. So, let us examine more closely the *rationale* against contraposition. I will examine Sosa's argument against contraposition in a special section. Suppose we understand safety in terms of simple subjunctivism as "If S were to believe *p*, *p* would be true," or "$B_S p > p$" and sensitivity as its contrapose: "$\neg p > \neg B_S$ p." Let us also adopt a slightly modified, Nozickian version of Stalnaker/Lewis semantics for true/true subjunctives, so that "If S were to believe *p*, *p* would be true" is true if and only if *p* is true at all close worlds at which S believes *p* (not just the actual one). These conditionals do not contrapose (Williamson 2000: 149):

> $B_S p > p$ can be true and $\neg p > \neg B_S p$ false if *p* is true at every close world but S believes *p* at the closest (but not close) world at which *p* is false. Equally, $\neg p > \neg B_S p$ can be true and $B_S p > p$ false if S believes *p* at some close but not closest worlds at which *p* is false.

Is this really decisive? S actually truly believes that *p*, but the closest world where S believes falsely is *not* nearby, so S's belief is safe but not sensitive? True, this constitutes a problem for simple subjunctivism—but we saw above how to reply: sensitivity has to be restricted and the closest worlds which are not nearby will then be irrelevant. This objection then loses its bite. Note also that Williamson ascribes to Nozick *different* semantical criteria for factuals (safety) and counterfactuals (sensitivity). Counterfactuals are interpreted according to standard Lewis semantics—"$\neg p > \neg B_S p$" is true at a possible world *w* if and only if either $\neg p$ is true at no possible world (the vacuous case) or, for at least one possible world *x*, $\neg p$ is true at *x* and $\neg B_S p$ is true at every possible world at least as close in the relevant respects as *x* is to *w*. Setting aside various qualifications (the possibility that there might be worlds that get closer and closer to the actual world without limit), we can say that this condition is true if and only if, in the closest non-*p* worlds, S does not believe that *p*. So we use close (*nearby* in my terminology) worlds for safety and the closest one(s) for sensitivity. But why the different standards?[8]

Sensitive but not safe? S actually truly believes that *p* and there is a world nearby where S believes falsely, but this is not the closest world in the set of nearby worlds (at the closest world where *p* is false S does not believe that *p*). Goldman's twins are an example—Sam cor-

[8] To be fair, Williamson (2000: 152) considers the option of interpreting sensitivity "$\neg p > \neg B_S$ p" as requiring the truth of its consequent at all contextually relevant worlds at which the antecedent is true. This view makes "B_S p > p" and "$\neg p > \neg B_S$ p" equivalent in any given context and is similar to the position defended in this paper.

rectly believes that Judy is before him, but if it were Judy's twin sister instead, he would mistake her for Judy. The possibility of Judy's twin sister is nearby but it is not realized in the closest world. But is it not in this situation just a happy *coincidence* that Sam sees Judy? To repeat the moral—for a connection to be stable, it has to hold within a set of nearby worlds, not just the closest one(s). It would be unwise to permit the *dependency* conditional "$\neg p > \neg B_S\, p$" to be true given that there is a world nearby where S believes falsely that *p*. In that case the connection is unstable; it might be only a happy accident that the closest not-*p* world is a world where it is not the case that S believes that *p* rather than a world where S believes that *p*.

I will adopt McCall's proposal (1983) to model the idea of modal connection (as before, '$B_b p$' is a proposition stating that S believes that *p* on a basis *b*). According to McCall, a *counterfactual* "not-*p* > not-$B_b p$" would be true iff some not-*p* & not-$B_b p$ is closer to the actual world than any not-*p* & $B_b p$ world. The key to the notion of connection lies in the requirement that not-*p* & $B_b p$ worlds must in all cases lie outside the set of not-*p* & not-$B_b p$ worlds, *p* & $B_b p$ worlds and possibly also *p* & not-$B_b p$ worlds centered around the actual world (McCall 1983: 312). The last requirement does not hold in the case of epistemic connections—possibilities where one fails to believe truly do not refute knowledge claims. However, when *p* and $B_b p$ are both true the idea of a connection requires that for a *factual* $B_b p > p$ to be true, some not-$B_b p$ and not-*p* world must be closer to the actual world than any not-*p* & $B_b p$ world. Given these semantics, factuals and counterfactuals of connection are extensionally equivalent within a set of nearby worlds.

To see this, suppose that the actual world is a "*p* & $B_b p$" world, that a counterfactual "not-*p* > not-$B_b p$" is true but a factual "$B_b p > p$" is false. In this case some *nearby* not *p*-world is a $B_b p$-world. Consequently, it might be only a happy accident that the *closest* not-$B_b p$ world *x* is a not-*p* world: there is a world which is both near to *this* world *x* and to the *actual* world where not-$B_b p$ & *p*. Next, suppose that the actual world is a "*p* & $B_b p$" world, that a factual "$B_b p > p$" is true, but a counterfactual "not-*p* > not-$B_b p$" is false. According to McCall (1983: 314), it would seem unwise to permit a factual "$B_b p > p$" to be true given that the closest not-*p* worlds (nearby) were $B_b p$ worlds, for in that case it might also be only a happy accident that the actual world was $B_b p$ & *p* rather than a $B_b p$ & not-*p* world. And if the closest not-*p* worlds are not nearby then we can exclude them on the grounds of irrelevance.

McCall develops a branched possible world model structure to develop a semantics which corresponds with Goodman's view—in cases where there exists a connection between the antecedent and the consequent, contraposition holds. His model presupposes indeterminism and the details might be problematic, but I think that the general idea holds water. I am inclined to accept what Cogburn and Roland (2013: 10) call a "Linguists' Version of Lewis's Semantics for Counterfactuals" for conditionals of *connection*:

> A conditional of the form, "if it were the case that *p*, it would be the case that *q*," is true in context C and world w just in case either (i) there is no *p*-world or (ii) if there is a *p*-world, then all (C, w)-relevantly similar *p*-worlds are q-worlds.

According to this condition indication conditionals ("safety") and dependency conditionals ("sensitivity") *contrapose* within the context C (which we take to be the set of relevant worlds).[9] By way of example, let us examine a paradigmatic case of failure of contraposition for counterfactuals given by Nute in the reference book on conditionals (Nute 1984: 394–395, his numbering):

24. If it were to rain heavily at noon, the farmer would not irrigate his field at noon.
25. If the farmer were to irrigate his field at noon, it would not rain heavily at noon.

I assume that the contrapose (25) is supposed to be "obviously" false because we tend to read (25) as if the farmer's actions brought about or made it the case that it did not rain heavily, which is absurd. However, what if this suggestion is removed and we simply concentrate on a connection without any particular direction? For instance:

25' If the farmer were to have irrigated his field at noon, it would not be the case that it has rained heavily at noon.

Or even better, to use one of Sosa's formulations of safety conditionals (2004a: 40):

25" It would not be so that the farmer irrigated his field at noon without it being so that it had not rained heavily at noon.

This sounds awkward, but it seems to me that (25") now really follows from (24). Someone might truly say in the evening:

> *Since* the farmer irrigated his field at noon, it was not the case that it had rained heavily at noon.

We may conjecture that in the case of a "connection" (understood very broadly) between the antecedent and the consequent, subjunctive conditionals contrapose.

4.

The modal camp in general accepts the idea that if you believed that *p* even if *p* might (seriously) not be the case, then you fail to know that *p*. However, principles of safety and sensitivity are supposed to be two inequivalent ways to express our modal intuition regarding how our beliefs in genuine cases of knowledge should be connected with facts not just in the actual world, but also in a relevant range of possible worlds. Wrongly so, as I think that a proper modal connection

[9] The idea of equivalence within limits, but on different grounds, is defended by Ichikawa (2011: 311), Leplin (2009: 154), Luper (2012: 221–22) and Yamada (2010: 78).

requires reliability throughout a space of relevant counterfactual situations which encompasses both aspects. This position has been hinted at approvingly—Pritchard (2008: 453) says that a rendering of safety and sensitivity in which they are extensionally equivalent would be "a fascinating development." On the other side such a proposal was criticized (Greco 2012) but never fully developed.

Safety is a later arrival on the scene, developed as a criticism and improvement on sensitivity. I will address some of the criticism, but note, I have to address arguments for a *distinction* of the two conditions and the superiority of safety, not the arguments against modal conditions in general (very often both conditions are in the same boat, e.g. with respect to the problems of closure). A typical argument has already been addressed—the alleged cases of sensitive but unsafe beliefs which intuitively do not count as knowledge. Sam truly believes that Judy is before him, but if it were her twin sister Trudy instead, he would mistake her for Judy. The closest possible world where his belief is false is a world where nobody stands before him (his belief is sensitive), but the possibility of Trudy being there is nearby (his belief is unsafe). The reply should be obvious by now: why consider just the closest (one and only) possible world as your interpretation of sensitivity? You have to consider *more* non-*p* worlds (not just the closest one), but not those which are irrelevant. Is this still *sensitivity* (properly understood) at all? Well, "what's in a name?" Pritchard calls the proposed modal condition super-safety, while Greco calls it restricted sensitivity; I have used the name modal stability (Šuster 2013). Never mind the name—considerations about relevance, non-accidentality of a proper modal connection and the equal treatment of conditionals involved motivate the extensional equivalence of safety *enhanced* by sensitivity within a set of nearby worlds and sensitivity *restricted* to a set of nearby worlds.

One of the main arguments for the difference between the two conditions has been the argument from contraposition, mainly developed by Sosa. Basically, safety and sensitivity expressed in terms of subjunctive conditionals are contrapositives, but since subjunctive conditionals do not contrapose, the conditions are different. Although long a proponent of safety, Sosa now thinks that though more adequate than the sensitivity requirement, this requirement of safety is still inadequate, as knowledge requires *aptness*: for a true belief to be knowledge it must be apt—accurate because it is adroit and grounded in a broader intellectual virtue or ability (Sosa 2007: 98). Safety still figures in his account of knowledge (e.g. an exercise of a competence is safe relative to its normal conditions if it would not easily have issued a false belief if exercised in those conditions) and some have argued that this condition is equivalent to basis-relative safety (Fernandez 2010: 44). However, I agree with Comesana (2013) that modal conditions are no longer central to Sosa's epistemology. Nevertheless, Sosa has done a great deal to develop the safety condition in its ability to give a better account of anti-skeptical

knowledge than other modal conditions. There are also other proponents of safety and I think that the issue is still interesting by itself, so in the rest of the paper I will be concerned with Sosa-as-a-safety-advocate and his formulation of the argument from contraposition.

Here is Sosa's reasoning in a nutshell: (i) sensitivity and safety expressed in terms of subjunctives (he calls them 'strong' conditionals) are inequivalent contrapositives; (ii) safety is more plausible than sensitivity; (iii) the plausibility of the sensitivity requirements derives from the corresponding safety requirements so easily confused with them through the failure to appreciate that strong conditionals do not contrapose. Sosa gives various reasons why safety is more defensible than sensitivity as a requirement for knowledge (inductive knowledge, knowledge of necessities, reflective knowledge, Sosa 1999: 145–146). However, superiority is most striking in a Moorean response to radical skepticism.

Consider a typical skeptical scenario—in the closest brain-in-a-vat world (which is, we assume, far, far away) humans do not believe that they are envatted brains. This belief is insensitive but it remains safe, however, as I would not easily believe I was not a brain-in-a-vat without it being the case that I was not a brain-in-a-vat. There are no *nearby* worlds where humans are envatted brains. If sensitivity is required for knowledge then we do not know we are not radically deceived; If safety is required for knowledge, then we know that we are not radically deceived. This is the corner-stone of a (neo)-Moorean response to radical skepticism as defended by Sosa (repeatedly provided in a number of places, even in the book where safety is superseded by aptness, Sosa 2007: 27):

A1. S's true belief that she is not radically deceived is safe but insensitive.
A2. When expressed in terms of subjunctive conditionals, safety is a contrapositive of sensitivity.
A3. Subjunctive conditionals do not contrapose.
A4. S can know that she is not radically deceived.
A5. Safety is a necessary condition for knowledge, but sensitivity is not.
A6. We easily confuse safety with sensitivity or at least think they are equivalent, because it is easy to assume that subjunctive conditionals contrapose.
A7. S's belief that that she is not radically deceived only appears not to be knowledge because it is insensitive, which explains the attraction of skepticism.

If sensitivity is restricted to a set of (relevant) nearby worlds and radical deception worlds count as irrelevant, the difference with respect to radical skepticism vanishes. Still, there is something attractive about this argument, and A3 and A6 have to be addressed from the perspective of conditionals of (dis)connection.

Subjunctive conditionals are commonly understood as counterfactuals (corner conditionals 'A > B', box-arrow conditionals 'A $\Box\!\!\rightarrow$ B'—'If it had been so that A, then it would have been so that B'). Sosa is somehow idiosyncratic. He speaks about subjunctive conditionals, the arrow conditionals and strong conditionals:

> My rough-and-ready conception of a (*strong*) *conditional* is this: *sentence that expressively conditions something x on something y, either as a necessary or as a sufficient condition.* If a sentence expresses <*p*> as sufficient for <q>, then its *contrapositive* is the same except only for negating each of '*p*' and 'q' and inverting their positions (Sosa 2004b: 279).

Sosa avoids the counterfactual reading (of the type "If the subject had believed that *p*, then it would have been so that *p*") and uses an arrow conditional '$p \rightarrow$ q', a definition of which can be extracted from various places:

> …'x → y' is short for 'It would be so that x only if it were so that y'. This is to be distinguished from 'If it had been so that x, then it would have been so that y', which has unfortunate implicatures or worse (Sosa 2004a: 40).

> 'r → *p*'…, (in an inequivalent but closely related alternative reading) as "Not easily would it be so that *r* without it being so that *p*." (Sosa 2004a: 54).

> And there are also more idiomatic variants, such as: <*p*> *would be false only if* <*B(p)*> *were false* (Sosa 2004b: 279).

> It would not be so that *p* unless it were so that *q* (Sosa 2004b: 322, footnote 7).

> Here '*p* → q' will be short for 'it would not be so that *p* without it being so that *q*'; or we might stipulate that in our usage it amounts to 'that *p* subjunctively implies that *q*'; the idea is that its being so that *p* offers some guarantee, even if not an absolute guarantee, that it is also the case that *q*. The guarantee is as weak as that offered by the truth of "If I should next release this pencil (held aloft and unsupported, etc., in an actual speech context), then it would fall"(Sosa 2002: footnote 4, 284).

> As a matter of fact, though perhaps not as a matter of strict necessity, not easily would *p* be the case without it being the case that *q* (Sosa 1999: 142).

One could roughly characterize Sosa's conditionals as expressing a connection without direction. All the formulations avoid the suggestion (common in the case of counterfactuals) that the antecedent of the conditional makes it the case, brings about, contributes causally to … , the occurrence of the consequent. I take this to be the "unfortunate implicatures" of the "If it had been so that *p*, then it would have been so that *q*" formulation feared by Sosa. When saying "I would believe that *p* on basis *b* only if *p*," we want to avoid the suggestion that the truth of (the basis for) our belief somehow makes it the case, brings about, contributes causally to … , the fact that *p*. Modally stable connection is perhaps best expressed in terms of a negative formulation: "It would not be so that S believes that *p* on basis *b* without it being so that *p*," and I have proposed a paraphrase: "Since S believes that *p* on basis *b*, it is true that *p*."

Consider next Sosa's typical formulation of a difference between safety and sensitivity:

> S's belief that *p* is *safe* iff it would not be true that S believes *p* without it being true that *p*, whereas it is *sensitive* iff it would not be true that not-*p* without it being true that S does not believe *p*. (More idiomatically a belief is *safe* iff it would be true if held, and *sensitive* iff it would not be held if false.) These being contrapositives, they are easily confused, or at least thought

> equivalent; but contraposition is invalid for such conditionals (Sosa 2004b: 276).

It is taken more or less for granted that subjunctive conditionals do not contrapose, but why, exactly? Here are some stock counter-examples from the literature:

> "If Boris had gone to the party, Olga would still have gone. So if Olga had not gone, Boris would still not have gone." Suppose that Boris wanted to go, but stayed away solely in order to avoid Olga, so the conclusion is false; but Olga would have gone all the more willingly if Boris had been there, so the premise is true (Lewis 1973: 35).

> The following story is the background for a counterexample to the argument form of contraposition or transposition *(if A then B;* therefore *if not-B then not-A.):* My dog is a mutt. His paternity is in some doubt, but even if his father were a purebred dog, my dog would still be a mutt since his mother was one. Now consider the contrapositive of the conditional claim made in this remark: if my dog were a purebred, his father would be a mutt. (I assume that *mutt* and *purebred* are contradictory properties, as applied to dogs.) This conditional is not only false, but impossible, and so cannot be a consequence of the true conditional claim made in the story (Stalnaker 1987: 124).

> It can be true that A > C and ¬A > C: if I were to sna*p* my fingers, the truck would go on rolling; if I were not to sna*p* my fingers, the truck would go on rolling—yet if Contraposition were valid, this would entail both ¬C > ¬A and ¬C > A, which is contradictory. It might be true that (even) if the British and Israelis had not attacked the Suez Canal in 1956, the Soviets would (still) have invaded Hungary later in the year, without its being true that if the Soviets had not invaded Hungary when they did it must have been the case that the British and Israelis had earlier attacked Suez (Bennett 2003: 172).
> If water now flowed from your kitchen faucet, for example, it would then be false that water so flowed while your main house valve was closed. But the contrapositive of his true conditional is false (Sosa 2007: 25).

The list could be extended, but there is a common feature that all the counter-examples have in common—they have premises which we hear as containing a tacit 'even' or 'still.' Let us check the examples listed. Lewis' conditional "If Boris had gone to the party, Olga would still have gone," was redescribed as an *even if* by Hunter (1993: 285):

> It seems to me (1) that the premise is equivalent to an 'even if'; (2) that the premise (which suggests that Boris' presence would have been something of a deterrent to Olga) does not fit the situation described (that Olga would have gone all the more willingly if Boris had been there); and (3) that the 'still' in the conclusion is in the wrong place for a strict contrapositive.

Here are Stalnaker's remarks about his own example:

> One might reject the counterexample on the grounds that the conditional contraposed is an "even if" conditional—a semifactual which should receive an analysis different from the one given to ordinary counterfactual conditionals (Stalnaker 1987: 124).

Bennett's conditionals are clear cases of *even ifs*. How about Sosa's housework problems? Contraposition would yield the absurd:

> If water flowed from your kitchen faucet while your main house valve was closed, it would then be false that water now flowed from your kitchen faucet.

The structure of this conditional is not entirely perspicuous. An analogy with an old joke might help.

> A pious man is praying to God to help him in his financial troubles and make him a winner of a fair lottery. Year after year he prays for the win and then he dies, disappointedly, without ever winning the lottery. He complains bitterly to God and receives a reply: "But why did you not buy a ticket, you miserable man!"

Here is the structure:

> Even if S had won the lottery, it would still be false that S won while not buying a ticket (even God cannot make S a winner in that case!).

And the structure of Sosa's example is the same: "Even if R, it would still not be the case that R without Q":

> Even if water now flowed from your kitchen faucet, it would still be false that water so flowed while your main house valve was closed.

Many took the failure of contraposition for subjunctives as an established fact, but there were always opposing voices. According to Hunter (1993), contraposition is not valid for "even ifs" but "even ifs" differ from ordinary "ifs." Goodman interpreted even-ifs as semifactuals which deny that a certain connection obtains between the antecedent and the consequent. Chisholm (1946: 298) suggested paraphrasing semifactuals before analyzing them. "Even if you were to sleep all morning, you would be tired" is to be read as a denial of the counterfactual connection: "It is false that if you were to sleep all morning, you would not be tired." Semifactuals admit 'even' in the antecedent and 'still' in the consequent and, according to the mainstream view, assert (or imply) their consequents. For Pollock (1976), the subjunctive "Even if A, C" is true only if C is true. However, not all "even ifs" seem to have a true consequent. Suppose I win in a game of chess and we afterwards analyze a pawn c5 variation. "If I had sacrificed the bishop (in this variation), I would have lost," I say, truly. But then I realize: "Even if I had not sacrificed the bishop, I would still have lost (in that variation)." Luckily for me, I did not choose the pawn c5 variation and won the game. Both conditionals seem to be true but they have a false consequent. Perhaps it is better to say that "P even if Q" implies Q no matter which of the relevant P-alternatives occurs, as suggested by Sanford (Sanford 1989: 216). In our case—no matter what I do with my bishop, there was no *connection* between the moves of my bishop and my winning the game (in *that* c5 variation). In the case we are ultimately interested in—no matter what I do would make any difference to whether I know that the skeptical hypothesis is false, so even if I were to be radically deceived, I would still believe I was not.

There is a faint glimmer of hope for a consensus on the semantic classification of even ifs. The field has been dominated by unifiers who

opt for a uniform treatment of *all* subjunctive conditionals (Stalnaker-Lewis style), but there have always been dissenting voices arguing that "even ifs" belong to a special class with different properties. The view that "P even if Q" implies Q (no matter what) is not quite accurate, but it is nevertheless closely related to failures of contraposition. In the cases where "P > Q" is true, but "not-Q > not-P" is not, the probability of the consequent 'Q' is close to 1 (Adams 1998, 143) or, in different terms (Lowe 1995: 51), the negation of Q is contextually impossible. "If it rains tomorrow there will not be a terrific cloud burst," does not contrapose to "if there is a terrific cloudburst tomorrow it will not rain," because, in this context, apparently, a terrific cloudburst tomorrow is not a salient possibility and "Even if it rains tomorrow, there will still not be a terrific cloudburst" looks appropriate. This is an indicative, but most theoreticians would agree that inferences involving counterfactual conditionals conform to many of the same laws as those involving indicatives.

We could also include "even ifs" in a broader category. Davis (1983) introduced a distinction between *weak* conditionals and *strong* conditionals, arguing that any good conditional containing 'then' between its antecedent and its consequent remains good if 'then' is dropped; however, the converse does not hold, because some good conditionals lacking 'then' turn bad when 'then' is inserted into them (Lycan 2006: 19). Here is Lycan's example: (i) If you open the refrigerator, it will not explode; (ii) If you open the refrigerator, *then* it will not explode. The *weak* (i) would normally be used merely to reassure the hearer that there is nothing about opening the refrigerator that is connected with an explosion ("do not worry, even if you open it, …") but the *strong* (ii) suggests that opening the refrigerator would keep it from exploding, perhaps because the refrigerator has been rigged to explode unless its door is opened in time. According to Lycan, weak conditionals are like semifactuals in that they readily take 'even' and they do not contrapose, but they do not so clearly assert their consequents. Furthermore, he adds that he knows of *no* purported counterexample to a contraposition that does not have such a weak conditional or a semifactual as a premise (Lycan 2006: 35)!

I will avoid a detailed analysis of the interactions between "even", "if," "still" and "then." I will speak in general about *weak* conditionals ('even ifs') and *robust* (connection) conditionals. The latter contrapose validly, while the non-contraposing, weak conditionals contain the structure: "(Even) if …, (still) …" which signals a *lack* of connection.

5.

Let us return to matters epistemic. "Even if *p* were false, S would still believe that *p* on basis *b*,"[10] is a non-contraposing weak conditional

[10] A weaker "Even if *p* were false, S might still believe that *p* on basis *b*,"

which expresses a lack of epistemic connection. Yet we expect robust conditionals, conditionals of connection, to contrapose within the limits of relevance. "Sensitive" beliefs (within limits) have their "safe" counterparts (with sensitivity extended to *all* nearby worlds), but one should also be able to infer dependency from indication: "Since S believes that p on basis b, it is true that p," therefore "If p were not the case, then S would not believe that p on basis b." Normally this is the case: if a reliable thermometer indicates my having a fever, then, if I were not feverish, the thermometer would not indicate fever and I would not form this belief. Now, take the critical case of S's true belief that she is not radically deceived. According to Sosa, this belief is safe but insensitive, but we easily confuse safety with sensitivity because it is easy to assume that subjunctive conditionals contrapose. "Even if she were radically deceived, she would still believe that she is not," seems to be true. Still, S's actually true belief indicates that she is not radically deceived: "Since S believes that she is not radically deceived, she is not radically deceived," does not look like a non-contraposing "even if." So, dependency is violated but indication is not and Sosa's diagnosis looks appropriate.

I think, however, that the proper diagnosis depends on (dis)connections between the basis for the belief and the relevant facts of the matter. One *could* read "Since S believes that she is not radically deceived, she is not radically deceived," as a *weak* and non-contraposing conditional, on the model of even-ifs, where "P even if Q" implies "Q no matter what." If S is not radically deceived "no matter what," the probability of the consequent is close to 1, or, the realization of a skeptical scenario is contextually impossible or a remote possibility that does not count. Contraposition is then really invalid and dependency (sensitivity) fails, since it is not the case that if S were radically deceived, she would believe so. However, one *could* also read "Since S believes that she is not radically deceived, she is not radically deceived" as a *robust* conditional of connection and in this case its contrapose: "If she were radically deceived, she would not believe that she is not," is also true. Let me explain these two options.

Alspector-Kelly (2011: 129–130) introduces an instructive distinction between "near-safety" and "far-safety." If there are no nearby worlds in which a proposition believed is false, then the belief is, according to Alspector-Kelly, automatically safe: it is not easily wrong simply because the proposition believed is not easily false. Such a belief is "far-safe," with far-safe beliefs to believe is to know, no matter why you believe it. In a certain sense, the structure of modal space, the location of the actual world within the set of possible worlds, does "all the work." Sosa sometimes speaks that way: "The possibility of radical

when 'might' is restricted to the space of serious possibilities is enough to deny a connection but the stronger formulation is more common in discussions about the lack of knowledge and skepticism.

deception is so outlandish that one's belief to the contrary would tend to be correct" (Sosa 2010: 79). And "In the actual world, and for quite a distance away from the actual world, up to quite remote possible worlds, our belief that we are not radically deceived matches the fact as to whether we are or are not radically deceived" (Sosa 1999: 147).

On this reading S's true belief that she is not radically deceived is really safe and insensitive, but satisfying standards that low should not really count in S's epistemic credit. True, not easily would one falsely believe that one is not radically deceived without that actually being the case. Yet, one would also not easily believe almost anything without it being the case that one is not radically deceived! No wonder that friends of sensitivity protest—this "knowledge" looks vacuous; S knows that *p* without being able to rule out any not-*p* worlds (Heller 1999: 207; Dretske 2005: 22–23). The appeal to modal remoteness does not *explain* my knowledge that I am not radically deceived, it only analyzes it according to Becker (2007: 162, fn 24). Discrimination or the ability to tell apart would provide an explanation, but on this terrain the skeptic apparently wins—if I have no evidence that tells me I am not radically deceived, how can I know it? The conditional "Since S believes that she is not radically deceived, she is not radically deceived" is true but far-safety does not succeed in establishing the connection which explains the falsity of its contrapose, the insensitivity of this belief.

According to Alspector-Kelly (2011: 130) the knowledge of far-safe propositions in general and of the negation of the skeptical hypothesis ("I am not radically deceived") in particular "imposes no conditions on the agent or her environment whatsoever (beyond, of course, those facts which ensure that the proposition's negation is remote)." If there are worlds in which a proposition that *p* is false, which lie within the boundary of nearby worlds, then safety requires that the agent not believe that *p* in any such world. When this requirement is realized, some feature of the way things actually are with the agent and/or her environment rules out the existence of such worlds. These beliefs are "near-safe." However, does it really follow that "... if the negation of the putatively known proposition is far, mere belief is enough, whereas if it is near, much more is required"? I do not think so. If modal distance is a function of the actual facts, if it is, e.g. because of the actual molecular structure of glass that it is fragile, then the actual facts about the cognizer and her environment sometimes ground epistemic connections, even if the negation of the putatively known proposition is far away.

Far-safety is safety on (a)steroids, safety gone wild, which establishes a connection between A and B on the basis of the following formula: "Not easily A without B because not easily not-B at all." This looks like a spurious connection, close to *dis*connection, expressed as "even if A, still not-B." However, there are types of genuine connections even when not-B is (in a certain contextual sense) out of the question: "Not easily A without B and not easily not B, but A is connected with

B." Take, for example, a connection between cooling a glass of water and its freezing at zero degrees Celsius. The freezing point of water depends on pressure, but the connection is stable; the freezing temperature of water would change by less than a degree if you increase the pressure by a factor of 100. Not an easy task; you would not easily cool the glass of water so that it froze at *one* degree Celsius, because you could not easily produce conditions which would realize this higher freezing point. Yet, if you change the environment radically (increase the pressure enormously), the connection is gone.

This, of course, invokes an externalistic perspective on epistemic connections. Remoteness of skeptical scenarios is established on the basis of actual categorical properties on the assumption that we are safely "placed" in a normal world and our beliefs meet appropriate external conditions. In a hostile environment, normal connections break down, but that is only to be expected. Consider, for example, externalism with regard to mental content. In order to have beliefs, it is necessary to be related to the environment in the right way. Given the radically different environment in skeptical scenarios, we would likely have access to few if any of the contents we actually have, and the connection is gone. However, it is then true, after all, that if I were a brain in a vat I would not believe I was not one. I would lack the conceptual resources required in order to believe anything about brains or vats (Sainsbury 1997: 918–919). The connection is externalistically grounded, so indication (safety) now contraposes dependency (sensitivity)!

Friends of safety and friends of sensitivity *alike* make this externalistic assumption when explaining the possibility of ordinary knowledge (and they face the same charge of begging the question against the skeptic). However, this externalistic move can be extended to our knowledge that we are not radically deceived. Epistemologists typically invoke a difference in the method of belief acquisition in order to make skeptical scenarios irrelevant. For Pritchard (2005: 156), safety requires that the agent's belief be true in the actual world and in a wide class of nearby possible worlds in which the relevant initial conditions are the same as in the actual world— "and this will mean, in the basic case, that the agent at the very least forms the same belief in the same way as in the actual world." Since, by definition, the brain in the vat is not using the same method as the agent in the good case, the skeptical scenario is irrelevant (far-away). Of course, this move can be and has been mimicked by a friend of sensitivity. The modal condition that the belief in question, along with its truth, be replicated in relevantly similar possible worlds—worlds where the belief's basis or its method of formation remain present, does not distinguish safety from sensitivity. Black (2002) employs the fact that Nozickian sensitivity must take explicit account of the methods of arriving at belief. The only worlds that are relevant to whether or not I know that p are those in which my belief is produced by the method that *actually* produces it. Skeptical

scenarios are supposed to be realized in possible worlds in which my belief is produced by different methods, so they are irrelevant.

More can be added to this external core; the belief that we are illusion-free is rationally coherent within our overall view (Sosa 1999: 147). To use Sosa's well-known distinction, the "animal" externalistic perspective can be enhanced by reflective knowledge. Ampliative inferences like induction and inference to the best explanation provide a reflective basis for a belief in the denial of skeptical hypothesis, and if plausibility is an acceptable standard then we are justified in rejecting skeptical scenarios (Leplin 2009: 141).

If skeptical scenarios are irrelevant and the basis for our belief that we are not radically deceived is appropriately connected with the facts of the matter then *indication*: "Since S believes that she is not radically deceived, she is not radically deceived," is a robust conditional of connection. In this case its contrapose, *dependency*: "If she were radically deceived, she would not believe that she is not," is also true, within the limits of relevance. There are various strategies to implement his result. The simplest option is to interpret the irrelevance of skeptical scenarios as (perhaps contextual) impossibility and adopt the Lewisian strategy of interpreting counterfactuals with impossible antecedents as trivially true (on the model of "If pigs had wings …"). Recall (Cogburn and Roland 2013: 10):

> A conditional of the form, "if it were the case that p, it would be the case that q," is true in context C and world w just in case either (i) there is no p-world or (ii) if there is a p-world, then all (C, w)-relevantly similar p-worlds are q-worlds.

In the case of "If she were radically deceived, she would not believe that she is not," the condition (i) is fulfilled. Two other options were also mentioned—a radically deceived person would not believe that she is not because she would lack the conceptual resources to form beliefs (Sainsbury 1997), or because her belief that she is not would be produced by methods other than actual, perceptual ones (Black 2002: 157).

6.

According to Sosa, we find the position of the skeptic plausible because we confuse the sensitivity condition, which is incorrect, with the safety condition, which is correct, and we invalidly contrapose the latter and confuse it with the former. I have argued that the issue is more complicated, as robust conditionals do contrapose; it is weak conditionals that do not. "Since S believes that she is not radically deceived, she is not radically deceived" is ambiguous between the weak reading (not radically deceived no matter what) and the robust—connection—reading. Its contrapose, "Even if she were radically deceived, she would still believe that she is not," is true, given the disconnection reading, but false given the connection reading. Indication based on connection is extensionally equivalent to dependency (within the limits of relevance).

Does the insensitivity of one's belief that one is not radically deceived *explain* the attraction of skepticism? Insensitivity reflects our intuitions about our inability to know the denial of a skeptical hypothesis, but it does not really explain it. McGinn poses a legitimate question when he asks what facts about the believer and her relation to the world make it true that if it weren't the case that p she would not believe that p. He offers a capacity to tell the difference between true propositions and false ones within some given class of propositions as an explanation ("global reliability," McGinn 1999: 17). On the face of it, dependency (sensitivity) more readily meets this explanatory requirement because we usually relate the agent's responsiveness to changes to her capacity to discriminate. However, the same relation holds between robustness of the basis for the agent's belief (safety) and her ability to tell apart. If S is unable to discriminate between relevant alternatives then her true belief that p is not responsive to the falsity of p (it is not sensitive). Suppose that the falsity of p is a serious possibility: in this case her true belief that p might easily be false (it is unsafe).

The critical question is always: when you actually believe that p on basis b, is it seriously possible for p to be false, but you continue to believe that p on the same basis? If so, then the connection is fragile and in this case even if not p, you might still believe that p. Yet "even if" is silent about the sources of disconnection. A reliabilist has a ready reply—one's (un)reliability grounds the connection and the capacity to discriminate is the mark of reliability (Goldman 1976). In this spirit Sosa (2000: 42) agrees with Lehrer that the real source of skepticism is precisely the indescernibility condition:

> If there is no discernible difference between that evidence I have for believing p if p were true and the evidence I would have for believing p if the denial of p were true, then I do not know that p on the evidence I have for believing that p (Lehrer 2000: 35).

The skeptic wants us to consider remote worlds in which the skeptical hypotheses are true and the evidence we have there is not reliable to conclude that it is not reliable in the actual world. However, of course, we have no means of distinguishing being in that condition from not being in that condition. This is the source of our oscillation between the connection and the disconnection interpretation of our belief that we are not radically deceived and the relevant facts of the matter. In the actual world, things being what they are, evidence is reliable, even if, in other worlds and other situations, it would not be. On the other hand, it seems that nothing we could do would make any difference to whether we know that the skeptical hypothesis is false, and that if it isn't false. Connection seems to be purely a function of our standards of remoteness, not of our justification. We are back to far-safety and its explanatory deficiencies.

I think that considerations about the explanatory lacunae of modal epistemology led Sosa to a new requirement of aptness. Necessary true

propositions are the most striking examples of far-safety—there are no worlds in which such a proposition is false. Take a simple case of two mistakes in a mathematical proof that cancel each other out, resulting in the correct conclusion C (Miščević 2007: 49). Suppose that mistakes are extremely hard to detect, so that the thinker, call her Jane, is justified in trusting her calculation. Jane's true belief in C is safe (necessarily true, so there are no worlds where she believes falsely) but still lucky, so not knowledge. Miščević proposes that with respect to *apriori* ("armchair") propositions we focus on the truth value of the belief that is formed in nearby possible worlds on the same basis as in the actual world, even when the resulting belief is not of the *same* proposition. Had Jane's ways of thinking been slightly different, she would not have managed to arrive at the same true belief as in the actual world, she might have ended up with believing the negation of the target proposition. The agent's belief is true in the actual world, but her cognitive structure and/or functioning might have differed in a minimal way from the actual one, and her beliefs would be false. However, once we see that characterization of luck should have a strong agent-concerning component, "we understand that exclusive focus upon modal instability of truth(s) is unwarranted even in the a posteriori cases" (Miščević 2007: 64).

The very idea of modal (or anti-luck) epistemology is now threatened. The requirement of cognitive stability of the cognizer (Miščević 2007: 67) corresponds to *aptness*. Sosa earlier (1999: 146) objected that sensitivity cannot deal with the problem of necessary truths (one cannot make the supposition that they are false). He now recognizes that safety faces the same problem, given that belief of any necessary truth is automatically safe. Aptness, the manifestation of epistemic competence, is now required to account for the divide between beliefs in necessary truths which are knowledge and those which are not (Sosa 2011: 85). According to Sosa, aptness may perhaps suffice with no need of safety as a separate condition at all. One may wonder, however, whether with the decline of safety the Moorean anti-skeptical strategy is still an option.

In any case, modal epistemology in general is endangered. I do not have much to say about these objections in this paper, as I still believe that the main insight of modal epistemology will remain: if it is seriously possible for S to falsely believe that p, then S fails to know that p. So let me summarize: the idea that what is distinctive about knowledge is captured by certain conditionals should be explained in terms of conditionals of (dis)connection and not in terms of simple subjunctives. The epistemic modal connection between one's basis for believing and the truth of one's belief should be stable (it should hold throughout the space of nearby worlds) and responsive to changes—those which are relevant. The fact that the epistemic connection breaks down in an unfriendly environment is the lesson from externalism. Safety (stabil-

ity) and sensitivity (responsiveness) are then two extensionally equivalent aspects of one modal relation (within a restricted space of possible worlds) or one modal connection viewed from two perspectives.

Once we drop the requirements of sensitivity *unlimited* (we ignore far-away worlds) and sensitivity *unique* (we do not consider just the closest, one and only world), safety is no longer superior but equivalent to sensitivity. Sosa often remarks that the requirements of safety and sensitivity are very similar and easily confused. It now seems that they go together through the good times (the heydays of modal epistemology) and the bad times (decline raised by problems of explanation) alike. What if appearances are not misleading? To amalgamate Groucho Marx and David Lewis, one could equate them with a *sotto voce* proviso: Safety may look like Sensitivity and talk like Sensitivity, but don't let that fool you—it really is Sensitivity! Psst!—within limits (which are difficult to specify).

References

Adams, E. W. 1998. *A Primer of Probabiliy Logic*. Stanford, CA: CSLI Publications.

Alspector-Kelly, M. 2011. "Why safety doesn't save closure." *Synthese* 183: 127–142.

Becker, K., 2007. *Epistemology Modalized*. Springer.

Becker, K., Black, T. (eds.) 2012. *The Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press.

Bennett, J. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Black, T. 2002. "A Moorean Response to Brain-in-a-Vat Scepticism." *Australasian Journal of Philosophy* 80: 148–163.

_____ 2011. "Modal and Anti-Luck Epistemology." In Bernecker, S., Pritchard, D. eds. 2011. *The Routledge Companion to Epistemology*. London and New York: Routledge.

Bogardus, T. 2012. "Knowledge under Threat." *Philosophy and Phenomenological Research*. DOI: 10.1111/j.1933–1592.2011.00564.x

Chisholm, R. 1946. "The Contrary-to-Fact Conditional." *Mind* 55: 289–307.

Cogburn, J., Roland, J.W. 2013. "Safety and the True–True Problem." *Pacific Philosophical Quarterly* 94: 246–267.

Comesana, J. 2013. "Safety and Epistemic Frankfurt Cases." In Turri, J. ed. 2013. *Virtuous Thoughts: The Philosophy of Ernest Sosa*. Springer: 165–178.

Davis, W. 1983. "Weak and Strong Conditionals." *Pacific Philosophical Quarterly* 64: 57–71.

Dretske, F. 1971. "Conclusive Reasons." *Australasian Journal of Philosophy* 49: 1–22.

_____ 2005. "The Case against Closure." In Sosa, E., Steup, M. (eds). *Contemporary Debates in Philosophy*. Oxford: Blackwell: 13–25.

Engel, M. 1992. "Is Epistemic Luck Compatible With Knowledge," *Southern Journal of Philosophy*, XXX/2: 59–75.

Fernández, M. A. 2010. "Aptness and safety: how are they related?" *Crítica* 42: 27–46.

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

Goldman, A. 1976. "Discrimination and Perceptual Knowledge." *The Journal of Philosophy* 73: 771–791.

_____ 1983. "Philosophical Explanations by Robert Nozick." *The Philosophical Review* 92: 81–88.

_____ 1986. *Epistemology and Cognition*. Massachusetts: Harvard University Press.

_____ 2012. *Reliabilism and Contemporary Epistemology*. Oxford: Oxford University Press.

Goodman, N. 1991. "The Problem of Counterfactual Conditionals." First published in *The Journal of Philosophy*, 44 (1947): 113–28. Reprinted in Jackson F. (ed.) 1991.: *Conditionals*. Oxford: Basil Blackwell: 9–27.

Greco, J. 2012. "Better safe than sensitive." In Becker, K. and Black, T. (eds.). *Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press. 193–206.

Heller, M. 1999. "Relevant Alternatives and Closure." *Australasian Journal of Philosophy* 77. 196–208.

Hunter G. 1993. "The Meaning of 'If' in Conditional Propositions." *The Philosophical Quarterly* 43: 279–297.

Ichikawa, J. 2011. "Quantifiers, Knowledge, and Counterfactuals." *Philosophy and Phenomenological Research*, Vol. LXXXII, 2: 287–313.

Lehrer, K. 2000. "Sensitivity, Indescernibility and Knowledge." *Philosophical Issues* 10: 33–37.

Leplin, J. 2009. *A Theory of Epistemic Justification*. Springer.

Lewis, D. 1973. *Counterfactuals*, Cambridge Mass.: Harvard University Press.

Lowe, E. J. 1995. "The Truth about counterfactuals." *The Philosophical Quarterly* 45: 42–59.

Luper, S. 2012. "False negatives." In Becker, K., Black, T. (eds.) 2012: 207–226.

_____ 2003. "Indiscernability Skepticism." in Luper, S. (ed.) *The Skeptics: Contemporary Essays*. Aldershot: Ashgate Publishing: 183–202.

Lycan, W. 2006. *Real Conditionals*. Oxford: Oxford University Press.

McCall, S. 1983. "If, Since and Because: A Study in Conditional Connection." *Logique et Analyse* 26: 309–321.

McGinn, C. 1999. "The Concept of Knowledge." In *Knowledge and Reality*. Oxford: Clarendon Press, 7–35.

Miščević, N. 2007. "Armchair Luck: Apriority, Intellection and Epistemic Luck." *Acta Analytica* 22: 48–73.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge: Cambridge University Press.

Nute, D. 1984. "Conditional Logic." In Gabbay D., Guenthner F. (eds.) *Handbook of Philosophical Logic*, vol. II. Dordrecht: Reidel, 387–439.

Pollock L. J. 1976. *Subjunctive Reasoning*. Dordrecht: Reidel.

Pritchard, D. 2005. *Epistemic Luck*. Oxford: Oxford University Press.

_____ 2008. "Sensitivity, Safety and Antiluck Epistemology." In Greco, J. (ed.) *Handbook of Skepticism*. Oxford: Oxford University Press: 437–455.

Roush, S. 2005. *Tracking Truth*. Oxford: Oxford University Press.

Rysiew, P. 2006. "Motivating the Relevant Alternatives Approach." *Canadian Journal of Philosophy* 259: 259–280.

Sainsbury, R. M. 1997. "Easy Possibilities." *Philosophy and Phenomenological Research* 57: 907–19.

Sanford, D.H. 1989. *If P, then Q: Conditionals and the foundations of reasoning*. London: Routledge.

Schaffer, J. 2005. "What shifts?: Thresholds, standards, or alternatives?" In Preyer, G. and Peter, G. (eds.) *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Oxford University Press: 115–130.

Sosa, E. (1997). "Reflective Knowledge in the Best Circles." *The Journal of Philosophy* 94: 410–430.

_____ (1999). "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13: 141–54.

_____ (2000). "Replies." *Philosophical Issues* 10: 38–42.

_____ (2002). "Tracking, Competence, and Knowledge." In Moser, P. K. (ed.) *The Oxford Handbook of Epistemology*. Oxford: Oxford University Press.

_____ (2004a). "Relevant Alternatives, Contextualism Included." *Philosophical Studies* 119: 35–65.

_____ (2004b). "Replies." In Greco, J. (ed.) *Sosa and His Critics*. Oxford: Blackwell Publishers.

_____ (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

_____ (2010). "Replies to my critics." *Crítica* 42: 77–93.

_____ (2011). *Knowing Full Well*. Princeton and Oxford: Princeton University Press.

Stalnaker, R. 1987. *Inquiry*. Cambridge, Mass.: The MIT Press.

Šuster, D. 2013. "Tracking without Concessions?" *Prolegomena* 12: 337–352.

Unger, P. 1968. "An Analysis of Factual Knowledge." *Journal of Philosophy* 65: 157–70.

Vogel, J. 2007. "Subjunctivitism," *Philosophical Studies* 134:73–88.

_____ 2012. "The enduring troubles with tracking." In Becker, K., Black, T. (eds.) 2012: 122–151.

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

Yamada, M. 2010. "Getting It Right By Accident." *Philosophy and Phenomenological Research* LXXXIII: 72–105.

Zalabardo, J. L. 2012. *Scepticism and Reliable Belief*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

Safety, Virtue, Scepticism: Remarks on Sosa

PETER BAUMANN
Swarthmore College, Swarthmore, USA

Ernest Sosa has made and continues to make major contributions to a wide variety of topics in epistemology. In this paper I discuss some of his core ideas about the nature of knowledge and scepticism. I start with a discussion of and objections against the safety account of knowledge—a view he has championed and further developed over the years. I continue with some questions concerning the role of the concept of an epistemic virtue for our understanding of knowledge. Safety and virtue hang very closely together for Sosa. All this easily leads to some thoughts on epistemic scepticism and on Sosa's stance on this.

Keywords: Knowledge, safety, epistemic virtue, scepticism, Sosa.

Ernest Sosa has made and continues to make major contributions to a wide variety of topics in epistemology. In this paper I discuss some of his core ideas about the nature of knowledge and scepticism. I start with a discussion of the safety account of knowledge—a view he has championed and further developed over the years. I continue with some questions concerning the role of the concept of an epistemic virtue for our understanding of knowledge. Safety and virtue hang very closely together for Sosa. All this easily leads to some thoughts on epistemic scepticism and on Sosa's stance on this.

1. *Safety*

It is 0.05 am on January 1, 2001. Jack just finished his first letter ever to his old friend Jill. Jack knows the time and date and comes to believe that it took him until the 21st Century to finish his first letter to Jill. There seems no reason to doubt that Jack knows this. Does it matter that Jack is confused about when the 21st Century begins? He thinks it began on January 1, 2000. If that would speak against his knowledge claim, then at least for some time at the beginning of the 21st Century many people, perhaps the majority, did not know that they were in the

21st Century. This seems false. We want to grant all those people this kind of knowledge even if they're confused about a particular year in the past.

However, the calendaric confusion matters with respect to something else. Jack's belief that he finished his first letter to Jill in the 21st Century could have been easily false. Not much would have had to be different for his belief to be false: He would just have had to finish the letter 6 minutes earlier. He would still have believed that he finished the letter in the 21st Century but that belief would have been false. Jack's belief is thus not "safe", as this is called these days. Knowledge, according to some and especially to Sosa, requires a safe true belief. S's belief that p is safe just in case

> (S) S believes that p ⇒ p (with "⇒" for the subjunctive conditional; cf., e.g., Sosa 1999, Williamson 2000).

Safety theorists are, of course, realistic enough to restrict the subjunctive conditional to close possible worlds, and not to include all possible worlds. We do not have to deal with further details of the account or with additional clauses here.

The example above suggests that safety is not necessary for knowledge: Jack's belief constitutes knowledge but is not safe. This is an interesting result, given that many epistemologists nowadays adhere to the view that knowledge requires safety.[1] But couldn't the safety theorist say something in reply?

They could point out that Jack's belief is not safe but also deny that Jack's belief constitutes knowledge. One reason could be that his true belief was based on a false assumption. However, if we add a no-false-belief condition for knowledge (cf., e.g., Clark 1963), then we're excluding many clear cases of knowledge. Someone who knows that they are living in the year 2888 by deducing it from the false belief that the millennium started on January 1, 2000 would thus come out as not knowing that they live in the 29th Century. This seems false. It will also not help much to try to argue that Jack's justified true belief is gettierized. This would seem ad hoc and, at least, in need of support by further arguments.

A more promising route might be to accept that Jack has knowledge and to deny that Jack's belief is unsafe and rather to affirm it is safe because a world in which he finished the letter 6 minutes earlier is not close enough to actuality to matter. But why is it not close enough? Not simply because that would save the safety account. If there is a fact of the matter which determines whether a possible world is close to the actual world, then our alternative world would rather come out as very close. Or so it seems. If there is no such fact of the matter, then all the worse for the safety account. Calling a possible world "close" to the

[1] Baumann forthcoming discusses all this in detail. Cf. also Neta/ Rohrbaugh 2004: 399–400, Roush 2005: 118–126, Comesaña 2005: 397, Goldberg 2015, sec. 1 and Sosa 2003: 159.

actual world would simply be arbitrary. Nobody has so far shown that there is a non-arbitrary closeness metric for possible worlds (cf. Lewis, a, b). Whether there is such a metric or not, in both cases the safety theorist cannot get what he wants.

Should one, perhaps, say that Jack's belief is safe but just not safely safe (cf. Sainsbury 1997; Williamson 2000: 123–130)? Have I mistaken the lack of second-order safety for the lack of first-order safety? Roughly speaking, a belief is safe but not safely safe just in case it could not have been easily been false, given conditions C, but conditions C could easily not have been met. The problem with this view is that there does not seem to be a way to determine such unique conditions C. Depending on which conditions we choose, we get different answers to the question whether a belief was safe. Jack's belief was safe, given that he was in the year 2001 when he finished the letter but it was not safely safe because he could have easily been in the year 2000. However, Jack's belief would come out as unsafe, given that he wrote the letter at some point during the winter 2000/2001. Nothing seems to determine one right description or choice of condition C.

If we do not have a plausible closeness metric for possible worlds, then the safety account hangs in the air because judgments about closeness of possible worlds are arbitrary. But even if we have a non-arbitrary ordering of possible worlds according to closeness to the actual world, we do get results in cases like Jack's which speak against the safety view. One should therefore conclude that safety is not necessary for knowledge.

Sosa has recently proposed to relativize safety to the "basis" of the belief (cf. 2004: 322, fn.3 and 2007: 25–28): "A belief that p is *basis-relative safe*, then, if and only if it has a basis that it would (likely) have only if true" (2007: 26). However, it is hard to see how this could help the safety theorist in our counter-example above. Why should one say that the basis of Jack's belief would be different had he finished his letter just a couple of minutes ago? Certainly, to say that the basis would be different because then his belief would be false would trivialize the safety account. As long as no account is given of the criteria for the identity of a basis in general and how this helps against our counter-example (as well as against similar examples), the doubts about the safety account remain.

2. *Closure*

There are more problems for the safety account. One has to do with a violation of closure:

(C) If S knows that p and also knows that $(p \rightarrow q)$, then S knows that q.

More conditions would have to be added but this rough, basic form of a closure principle should be sufficient for our purposes here. Take the following version of Kripke's red barn objection against Nozick's sensitivity account of knowledge (cf. also Goldman's dachshund example in

1983: 84). Mary is in fake barn county; only red barns are exempt—they are never fake. Mary is aware that red barns are never fake but she is not aware that she is in fake barn county. She finds herself in front of a red barn and comes to believe and know that there is a red barn in front of her. Her belief is safe. She can infer from it that there is a barn in front of her. However, since her belief that there is a barn in front of her is not safe (it could have easily been wrong in fake barn county), the safety theorist would have to deny that she knows there is a barn in front of her. This, however, is not plausible. Apart from that, (C) would thus be violated (cf. Kvanvig 2004: 209).[2] Sosa (2004: 292–294) concedes that there is a problem and proposes to relativize safety to a basis (see above). Again, it is hard to see how this could help the safety theorist. What is the difference as to the basis? It cannot be the truth of Mary's belief that there is a red barn nor that her belief constitutes knowledge—because that would trivialize the safety account. As long as no more promising version of the safety account is available we have good reason to remain sceptical.

3. *Probabilistic Safety*

Perhaps one should express the intuition behind the safety account not in modal but in probabilistic terms (cf. for a similar move for the sensitivity account Roush 2005). Here is the rough idea: Knowledge that p requires a true belief that p which also meets the following condition (with "P (p/ Bp)" referring to the conditional probability of p, given the subject's belief that p):

(S*) P (p/ Bp) > some suitable value m

or, alternatively,

(S**) P (not-p and Bp) = some low value m << P (p and Bp).

Kvart (2006) has proposed such an account for the case of perceptual and memorial knowledge. The basic, rough idea is that

(K) P (p/ Bp) >> P (p) (cf. Kvart 2006: 7).

And the value of the left hand side has to be high (10). Kvart has much more to say about this but we can leave it at this basic level here.

Unfortunately, there seem to be insuperable problems with such an account, too. First, it does not work for beliefs in necessary truths. P (p/ Bp) is always maximal (= 1), in such cases and the belief that p has nothing to do with P (p): P (p/ Bp) = P (p). Hence, (S*) or (S**) do not seem to add anything relevant and interesting to the true belief condition for knowledge in cases of necessary truths. Kvart's condition (K)

[2] Pritchard (2005: 167–168) objects that one has to consider a wider range of possible worlds here which would make Mary's belief that there is a red barn unsafe. However, this move seems ad hoc. One should also add that if Kripke's example is a problem for sensitivity accounts, then it is also one for safety accounts. It is often assumed that it is only a problem for sensitivity accounts.

would be false in such cases—as he himself recognizes—because P (p/ Bp) = P (p).

But even if we restrict the above probabilistic conditions to empirical knowledge of contingent propositions (as Kvart explicitly does), problems remain. Couldn't the probability that Bill is in the 21st Century, given that he believes it be high (=1 or close to 1) just because he spends all (or almost all) of his life in that century? If the identity of a human being (like Bill) is determined by the identity of the parents' sperm and egg and if there was no way Bill's parents could have saved the sperm and egg for a different century, then the probability that Bill could have lived in a different century equals or approximates 0. Again, it seems that neither (S*) nor (S**) can do any work additional to the true belief condition for knowledge: P (p/ Bp) = P (p). And again, Kvart's condition (K) turns out to be false. It is not clear how one could further restrict (S*), (S**), (K) or any other principle of that kind in order to deliver the right results. Apart from all that, the probabilistic version of the safety account would have problems with the closure principle, too, as a probabilistic version of the red barn objection shows (I spare the reader the repetition in this case; but see some ideas in Roush 2005: 41–47).

4. *Virtues*

Perhaps we should then move the focus away from modal or probabilistic safety to virtues? Is knowledge true belief which results from the exercise of an epistemic virtue, as Sosa has been arguing for quite some time? Sosa (2007: 42) proposes the following definition of knowledge: Knowledge is justified true belief which is also apt (that is, the result of the exercise of an epistemic virtue). One could call this "the JTAB account of knowledge" (cf. also Sosa 1988: 174–184; 1991a: 138–145; 1991b; 1991c; 1992: 85–89; 1994: 29–33; 1997: 419–420; 2007: ch. 2).[3]

A being has a virtue only if there is the possibility of falling short of the virtue or of the exercise of the virtue. This implies that no being which cannot but behave in a certain way can be virtuous (or not virtuous) with respect to that kind of behaviour. A being which simply cannot help but perceive their immediate environment realistically and correctly would not count as epistemically virtuous (nor as epistemically deficient or "vicious"). To the degree that human perception under normal conditions is very much like that we would not have perceptual knowledge. This, however, seems false. This objection is still a relatively "inexpensive" shot against the safety account. But there are more serious worries.

There are cases where it seems rather clear that the subject has knowledge even though no virtue was exercised. This would also make

[3] Credit for a performance has two dimensions: the ability and the effort. We may praise someone for their performance because of the great ability that was exercised even if little effort went into it but we may also praise someone for the remarkable effort in the use of more restricted abilities.

the idea doubtful that the exercise of a virtue is necessary for knowledge. Suppose Joe is thinking hard about some difficult mathematical problem. Suddenly he can "see" the solution. He didn't exercise special mathematical abilities; rather, the solution "just came to him", a bit like in Kekulé's case when—according to some accounts—he suddenly "saw" the structure of the benzene molecule in a dream or dream-like state. Joe thus comes to know the conclusion—without the exercise of an epistemic virtue. Sure, very often there are epistemic virtues in the background in such cases, like in Kekulé's case. But even then, the epistemic virtues were not exercised (though somehow causally effective). In Joe's case we can even assume that the relevant epistemic abilities or virtues weren't even present in the background. But there still seems to be knowledge (though probably failing the standards of professional mathematicians' knowledge) even without (an exercise of) epistemic virtue. Basic perceptual knowledge might be another case. Furthermore, if Joe forgets everything again after five minutes and never ever manages to reconstruct the solution despite repeated serious attempts, we would be even more inclined to deny that he had the epistemic virtue but we wouldn't deny that he knew the solution even if only for a short while. Knowledge can be short-lived but the corresponding virtue would be a more stable disposition. There are no virtues on one occasion only.[4]

What about the sufficiency of apt justified true belief for knowledge? Here is a counter-example. Sue is an expert on Rembrandt. Nobody comes close to her ability of telling whether something is a true Rembrandt or rather an imitation produced by a member of his school. Recently, two paintings have been discovered: one an original Rembrandt and the other one an astonishing copy by a pupil. All the other leading experts had been asked and failed to tell which is which. Sue however uses her very special abilities and, after some time, comes up with the correct answer. She has a justified, true and apt belief which qualifies as knowledge. What, however, if we add the following aspect to our example: Unbeknownst to her, some jokester at the museum has used the latest high tech tools to produce reproductions of the same Rembrandt painting and put them next to the real one and the copy by the pupil. Nobody, not even Sue, can distinguish between the original and its high tech reproduction (if unmarked as such). Suppose these reproductions are next to the two paintings Sue has been examining. Easily, she could have been, by accident, presented with a high tech reproduction. It seems that under such circumstances she doesn't know that the painting in front of her is an authentic Rembrandt. But it also seems plausible to say that her belief is justified, true, and apt: Hasn't she used her extraordinary epistemic virtues to arrive at a justified

[4] In all these cases, Joe was lucky in some respect and his knowledge was "lucky" knowledge in that respect. (The idea that some knowledge can be lucky in a certain sense might seem quite unorthodox; whether it is and in what way, should be investigated further—but not here).

true belief? But then justified, true and apt belief is not sufficient for knowledge.

One might object that even though Sue has some remarkable epistemic abilities she does not have the "right" or "relevant" ones here: abilities which would enable her to tell a Rembrandt from a high tech reproduction. But what, one would want to ask back, counts as the "right" or "relevant" abilities here? Does the subject need those abilities which lead her to the truth in a non-accidental way? If that is the answer, then it comes with a prize: It is not clear anymore whether reference to epistemic virtues does the work it was supposed to do in the first place, namely to explain the non-accidentality of a given true belief. The accidentality problem would remain even given the exercise of some virtue. More would have to be said here if one wanted to defend a virtue theory of knowledge.

Sosa (2007: 96, fn. 1) holds that in fake barn cases (structurally similar to the Rembrandt case above) only reflective knowledge is missing but not animal knowledge (cf. for this distinction: Sosa 1988: 182–184; 1991b: 240; 1994: 29–30; 1997: 422, 427; 2004: 290–292; 2007: ch. 2 Sosa 2009a: ch. 7). Reflective knowledge that p is justified true and apt (second-order) belief that one's true and justified first-order belief that p is apt.[5] One can apply this kind of response easily to the counterexample above. However, I don't find this very plausible: It seems very plausible to say that Sue also lacks "animal" knowledge that it is a real Rembrandt, too.

Perhaps one might want to reply that the conditions for the exercise of the epistemic virtue were not the right ones in Sue's case. Sosa (2007: 33) introduces condition C:

> For any correct belief that p, the correctness of that belief is attributable to a competence only if it derives from the exercise of that competence in appropriate conditions for its exercise, and that exercise in those conditions would not then too easily have issued a false belief.

Perhaps the circumstances were not appropriate in Sue's case. But why should we say that? Is it because she could have easily been wrong? But then the notion of safety rather than the notion of an epistemic virtue is doing the crucial bit of the work here. And we would be back with the problems for the safety view mentioned above. We also should not say that circumstances for the exercise are only appropriate if the subject gains knowledge under those circumstances. Given that we are trying to understand the nature of knowledge, this move would, again, trivialize the virtue account.

[5] Sosa (2007: 32) also characterizes reflective knowledge that p as second-order animal knowledge: animal knowledge that one has animal knowledge that p. So, reflective knowledge that p is based on animal knowledge of the second order. This might seem a bit odd, if only because it is not clear whether second-order attitudes could be "animal" ones at all.

5. *Virtuous Scepticism*

Do epistemic virtues help against scepticism? Here is an argument to the effect that one traditional form of scepticism only arises for the more virtuous or reflective. Take the template of Cartesian sceptical arguments (with "o" for an ordinary proposition and "s" for a sceptical proposition):

(1) S doesn't know that not-s
(2) If S doesn't know that not-s, then S doesn't know that o
(3) Hence, S doesn't know that o.

(2) is based on the assumption that

(4) S knows that (o → not-s).

If S knows that o, then S also knows—given closure (C) and (4)—that not-s. In other words, given closure and (4), if follows that

(2) If S doesn't know that not-s, then S doesn't know that o

Let us take a closer look at (4) and take a popular example. George knows that if he has hands, then he is not merely (and thus falsely) dreaming that he has hands. Now, knowledge requires understanding the known proposition. Whoever does not know that merely dreaming that p involves the false belief that p does not understand what merely dreaming is and thus does not understand that if they have hands, then they are not merely dreaming that they have hands. Therefore, they do not qualify as knowing (or even believing) that proposition. Hence, if it is true that

(5) George knows that if he has hands then he is not merely dreaming he has hands

then it is also true that

(6) George knows that if he has hands then he does not falsely believe he has hands when he really has no hands.

This is an interesting result. It shows that the Cartesian sceptical argument only works under the assumption that the subject has second-order concepts and second-order beliefs and can form beliefs about their own beliefs. In other words, at least the traditional Cartesian scepticism presupposes reflectivity. A being which is restricted to first-order beliefs—to animal beliefs—is, ironically, not threatened by this kind of Cartesian scepticism. A lack of ability can save one from (some forms of) scepticism. Reflectivity, however, can "destroy" knowledge. Less would be more and more less.

Sosa (2007: ch. 2) argues that dream scepticism only threatens the possibility of reflective knowledge but not the possibility of animal knowledge. There is a weaker and a stronger interpretation of this claim. The stronger claim says that only those who do not reflect upon their epistemic states can retain their animal knowledge. The weaker claim says that in addition those subjects who do reflect on their epistemic state can retain their animal knowledge, too, and only "lose" their reflective knowledge. There is something to be said in favour of

the stronger claim: If one's claims to reflective knowledge are threatened by dream scepticism, then this seems to give the subject a good reason to become sceptical with respect to her first-order belief, too. Animal knowledge would collapse together with reflective knowledge.

6. *Dreams, Beliefs and Scepticism*

But is the assumption that we do believe things in our dreams accurate? Sosa has denied this (cf. Sosa 2005, 2007: ch. 1). According to him, to dream is to imagine something and not to hallucinate or falsely believe it.[6] Sosa makes the distinction between what happens in one's dream and what happens while one dreams and applies this distinction to beliefs and believing. I still believe that there are no dragons even if I am dreaming about dragons. How could I believe in my dream that there are dragons when I really don't believe that there are dragons? The assumption of an inconsistency seems unconvincing and forced. Hence, we should rather give up the idea that we do have beliefs in our dreams. Sosa extends this argument from dispositional beliefs to occurrent or manifest thoughts. We don't think or believe things in our dreams, we rather imagine things.

I don't want to go further into this imagination model of dreaming but rather discuss the way Sosa uses it as an anti-sceptical weapon (cf. Sosa 2005, 2007: ch. 1, cf. also the exchange between Cohen 2009: 124–125 and Sosa 2009b: 142–143). Consider the claim to know that one is seated. Sosa argues that it is rational to go with the assumption that one is sitting and is not merely dreaming it. Only when one is awake can one ask a question (Am I sitting? Do I know that?) and answer it. When one is dreaming one cannot even ask a question (not to mention answer it). Hence, while awake the rational thing to do is to answer questions like "Am I sitting?" in the positive (when apparently sitting).

It is not quite clear in what sense this is rational. If the subject cannot distinguish between being awake and dreaming, then there is no reason accessible to her to "assume" she is awake and having thoughts and beliefs, raising questions and answering them (cf. Ichikawa 2008, 2009).

But perhaps Sosa's idea here is rather that the subject could reflect in a decision-theoretic way (where the alternative acts are either to trust or not to trust the appearances, and where the outcomes are epistemic ones):

		Circumstances	
		Awake	Dreaming
Acts	Go with appearances	good	indifferent
	Don't	bad	indifferent

[6] McGinn (2006: ch. 6) argues for an imagination account of dreaming but in contrast to Sosa he also argues that we do have beliefs (or, at least, quasi-beliefs) in dreams (cf. ch. 7, esp. 110, 112)

According to Sosa, however, neither act is available in the case of dreams: neither can one ask or answer questions nor can one suspend judgment, go with or against appearances. How then should we frame the decision matrix?

We need a more general notion of thought, covering both the things we do when awake (genuine asking and answering of questions, suspending judgment, etc.) and when dreaming ("pseudo-asking", etc.). Let us call these kinds of thoughts "super-thoughts" ("super-asking questions", "super-answering questions", "super-suspending judgment", etc.). Now, one could have problems with such a very general conception of thought—isn't it an invention of an arbitrary category of thought when we have no reason to assume there is a unitary phenomenon here? On the other hand, if one gives up this idea, then it is even harder to explain what "rationality" could mean in this kind of context; we would, at least, be at a loss when trying to construct a decision-theoretic matrix.

Presumably, the revised version of the matrix for super-thoughts would, according to Sosa, be the following one:

		Circumstances	
		Awake	Dreaming
Acts	Super-go with appearances	good	indifferent
	Don't	bad	indifferent

Hence, the first act dominates the second one. In other words, while awake it is rational to go with appearances even if one cannot distinguish between being awake and dreaming.

Still, there are several problems with this strategy. First, why should one assume that there is only one "deviant" or sceptical circumstance? What about evil demons (old and new), brains in vats, Berkeleian worlds, etc.? It seems that we don't even have an idea of what all the possible deviant circumstances might be. And, upon reflection, we can become aware of that. How then can it be rational then to go with the first act? Second, even if we accept the claim that the circumstances in the matrix above are all there are it is not clear what the outcomes are in the deviant case. Perhaps pseudo-going with appearances will be really bad (because it has an impact on our epistemic virtues when awake)? But then no act dominates the other one anymore. We need an argument which shows that the first act is the rational one in terms of possible outcomes; for that we need an argument which tells us why we should expect the above outcomes rather than others. Third, even if from the perspective of the subject it is rational to super-go with the appearances, it is doubtful whether this is sufficient for a successful reply to the sceptic: One can be justified in one's false propositional attitudes, even if they are systematically false. The subject could thus be perfectly rational in super-going with the appearances but might still be wrong. This possibility seems sufficient to entertain legitimate sceptical worries.

7. *Conclusion*

Much more could be said about Sosa's ideas about knowledge and scepticism and about the objections raised above. But it is better to stop here and see what can be said in Sosa's defence.[7]

References

Baumann, P. 2008. "Is Knowledge Safe?". *American Philosophical Quarterly* 45: 19-31.

Cohen, S. 2009. "Knowledge as Aptness". *Philosophical Studies* 144: 121–125.

Clark, M. 1963. "Knowledge and Grounds". *Analysis* 24: 46–48.

Comesaña, J. 2005. "Unsafe Knowledge". *Synthese* 146: 395–404.

Goldberg, S. 2015. "Epistemic Entitlement and Luck". *Philosophy and Phenomenological Research* 91: 273–302.

Goldman, A. 1983. "Review of Nozick, *Philosophical Explanations*". *The Philosophical Review* 92: 81–88.

Ichikawa, J. 2008. "Skepticism and the Imagination Model of Dreaming". *The Philosophical Quarterly* 58: 519–527.

Ichikawa, J. 2009. "Dreaming and Imagination". *Mind & Language* 24: 103–121.

Kvanvig, J. 2004. "Nozickian Epistemmology and the Value of Knowledge". *Philosophical Issues* 14: 201–218.

Kvart, I. 2006. "A Probabilistic Theory of Knowledge". *Philosophy and Phenomenological Research* 72: 1–43.

McGinn, C. 2004. *Mindsight. Image, Dream, Meaning*. Cambridge: Harvard University Press.

Neta, R. and Rohrbaugh, G. 2004. "Luminosity and the Safety of Knowledge". *Pacific Philosophical Quarterly* 85: 396–406.

Pritchard, D. 2005. *Epistemic Luck*. Oxford: Clarendon.

Roush, S. 2005. *Tracking Truth. Knowledge, Evidence, and Science*. Oxford: Clarendon.

Sainsbury, R. M. 1997. "Easy Possibilities". *Philosophy and Phenomenological Research* 57: 907–919.

Sosa, E. 1988. "Beyond Scepticism, to the Best of our Knowledge". *Mind* 97: 153–188.

Sosa, E. 1991a. "Reliabilism and Intellectual Virtue". In Ernest Sosa, *Knowledge in Perspective. Selected Essays in Epistemology*. Cambridge: Cambridge University Press: 131–145.

Sosa, E. 1991b. "Knowledge and Intellectual Virtue". In Ernest Sosa, *Knowledge in Perspective. Selected Essays in Epistemology*. Cambridge: Cambridge University Press: 225–244.

Sosa, E. 1991c. "Intellectual Virtue in Perspective". In Ernest Sosa, *Knowledge in Perspective. Selected Essays in Epistemology*, Cambridge: Cambridge University Press: 270–293.

[7] For discussion and remarks on earlier drafts I am grateful to Ernest Sosa and Joe Morrison.

Sosa, E. 1992. "Generic Reliabilism and Virtue Epistemology". *Philosophical Issues* 2: 79–92.

Sosa, E. 1994. "Virtue Perspectivism: A Response to Foley and Fumerton". *Philosophical Issues* 5: 29–50.

Sosa, E. 1997. "Reflective Knowledge in the Best Circles". *The Journal of Philosophy* 94: 410–430.

Sosa, E. 1999. "How to Defeat Opposition to Moore". *Philosophical Perspectives* 13: 141–154.

Sosa, E. 2003. "The Place of Truth in Epistemology". In Michael DePaul and Linda Zagzebski (eds.), *Intellectual Virtue. Perspectives from Ethics and Epistemology*. Oxford: Clarendon: 155–179

Sosa, E. 2004. "Replies". In John Greco (ed.), *Ernest Sosa and his Critics*. Malden: Blackwell: 275–325.

Sosa, E. 2005. "Dreams and Philosophy". *Proceedings and Addresses of the American Philosophical Association* 79 (2): 7–18.

Sosa, E. 2007. *A Virtue Epistemology. Apt Belief and Reflective Knowledge*. Vol I. Oxford: Oxford University Press.

Sosa, E. 2009a. *Reflective Knowledge. Apt Belief and Reflective Knowledge*. Vol. II. Oxford: Oxford University Press.

Sosa, E. 2009b. "Replies to Commentators". *Philosophical Studies* 144: 137–147.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

Intuitions: Reflective Justification, Holism and Apriority

NENAD MIŠČEVIĆ
Department of Philosophy, Faculty of Arts,
University of Maribor, Slovenia

The paper discusses Sosa's view of intuitional knowledge and raises the question of the nature of reflective justification of intuitional beliefs. It is assumed, in agreement with Sosa, that pieces of belief of good researchers are typically reflectively justified, in addition to being immediately, first-level justified. Sosa has convincingly argued that reflective justification typically mobilizes and indeed should mobilize capacities distinct from the original capacity that has produced the belief-candidate for being justified, in order to assess the reliability of the original capacity. It has to go beyond justifiers that are of the same-kind ("homogeneous") as first-level immediate ones, in order to enlarge the circle of justification (and thus avoid viciousness), and is, therefore, holistic and coherentist. But if this holds, it seems that reflective justification of armchair beliefs, presumably produced by intuition and some reasoning, should revert to empirical considerations testifying to the reliability of intuition and reasoning. Therefore, it typically combines, in an articulated way, a posteriori elements contributing to the thinker's reflective trust in her armchair capacities. In short, the paper argues that Sosa's own view of second-order justification goes better with a more aposteriorist view, if it does not even force such a view.

Keywords: Virtue epistemology, a priori, a posteriori, two-level epistemology, intuition.

1. *Introduction*

The paper discusses Sosa's view of intuitional knowledge[1] and raises the question of the nature of reflective justification of intuitional beliefs. I agree with him on many crucial matters, and have basically followed his

[1] My deepest thanks go to professor Sosa personally, for inspiration and support.

footprints in opting for a version of virtue epistemology and indeed, a two level one, familiar from his writings.

Reflection plays a paramount role in the biography of most serious intuitions, so what ought this role to be like? How does this strongly rationalistic approach fit his second-level epistemology, and perspectivalism (which I like and endorse myself), virtue-geared or otherwise? I shall argue that it does not fit well, and that the holistic character of the second-level perspective points in a very different direction. A thinker reflecting on her first-level armchair beliefs is expected to mobilize, should the need arise, literally all sorts of first-level beliefs of hers, in the hope that these belong to her first-level, animal knowledge. This might routinely include pieces of a posteriori knowledge. Can the armchair belief, thus justified, retain its aprioristic purity? Note that Sosa himself sometimes admits "[t]hat the reflective defense of our mathematical and other beliefs will not be purely a priori (…)" (2000: 13), although he then almost takes the admission back, and never develops it to any extent.

Here, then, is is the plan of the paper. Section two offers a brief overview (with apologies for brevity) of Sosa's subtle, interesting and original views, in particular about reflective justification in general, and about intuitions and intuitional beliefs. Section three is the central one: there it is argued that the proposed accounts of the two topics just mentioned do not fit well with each other: the general story of reflective justification stresses the role of broad coherence and a holistic web of belief, the particular story of justification of intuitions insists on homogenous justification of the a priori by the a priori. In the concluding section I shall briefly sketch my own proposal that the full justification of the deliverances of intuition is a highly structured one, as one would expect from the holistic character of second-level, reflective considerations, with some a priori elements, but with a sufficiently large a posteriori component to ultimately make it predominantly a posteriori. I conclude by reiterating three main questions for professor Sosa, thanking him in advance for an answer I know will be illuminating.

2. *Sosa's proposal*

a) *reflective knowledge, broad coherence and the web of belief*

Let me first remind you of Sosa's highly original proposal of a systematic virtue epistemology, and in particular of his views on two connected subtopics that we shall discuss in the rest of the paper: first, reflective knowledge, and second, intuitions and their justification.

The basic idea, with which I very much agree, is the general framework, stressing epistemic competences-virtues, and their truth-directedness. A crucial element characterizing Sosa's approach is the idea of the general epistemic structure, contrasting and combining the basic

level of first-order ("animal") competences-cum-performances with a reflective, second-order level (2007, 2009b).[2]

What characterizes reflective knowledge is "perspectival endorsement of the reliability of one's sources". (2009b: 136). It nicely brings together coherence and understanding, Sosa argues (2009b: 138). He illustrates it by bringing in Descartes's strategy of supporting first-level intuitional beliefs, prominently mathematical ones, by theological reasoning guaranteeing the coherence of the whole, and thus turning the mere *cognitio* of mathematical truths into *scientia*. The feature of reflective knowledge which will interest us most is what he calls "broad" or "comprehensive" coherence.

This stands in contrast with other theories of reflective knowledge, of a different kind, which argue that a first-level competence can be second-level justified, just by being re-applied, or again appealed to, in a reflective manner. Sosa (1994), criticizes W. Alston, the main proponent of this same-sort or homogenous second-order justification. It is combination of competences that does the work. Here is a characteristic passage:

> How does internal coherence, of little significant epistemic value in itself, become more valuable when combined with external competence? Coherence-seeking inferential reason, like retentive memory, is valuable when combined with externally competent faculties of perception, because when so combined it, like retentive memory, gives us a more comprehensive grasp of the truth than we would have in its absence. (2009b: 191)

The broad coherence goes beyond relations among the thinker's first-order beliefs, and involves coherence between them and the thinker's experiences, as well as comprehensive inter-level coherence (2009b: 192). It is necessary, Sosa claims, for the kind of reflective knowledge traditionally desired and desirable for its contribution to truth. One important component of the broad, comprehensive coherence is explanatory coherence, the contribution of some beliefs to a deeper understanding of others, and perhaps vice versa. Interpreting Descartes in an interesting and original way, Sosa stresses the "epistemic power" of such explanatory coherence. Even our mathematical knowledge can be helped by "a view of ourselves and our place in the universe" that is sufficiently comprehensive and coherent to bring us "into the realm of higher, reflective, enlightened knowledge, or "*scientia*" (2009b: 150).

Let me finally mention the metaphor which will be crucial in the discussion section: the pipeline/web contrast (2009b: 239, reiterated in 2011b: 150). The role of reflective epistemic justification is not well

[2] In his (2011b), and then further in (2015), he introduces an additional component of, or condition upon, reflective knowledge, characterizing an epistemic performance as fully apt "only if its first-order aptness derives sufficiently from the agent's assessment, albeit implicit, of his chances of success (and, correlatively, of the risk of failure)" (2011b: 11); we shall leave this refinement aside in the present paper.

characterized by the metaphor of the pipeline. The right picture is the coherentist one of the web of belief. Sosa talks about "an intricate spider's web" with its many nodes. He mentions that the position of each node (the status of each belief, we are allowed to suppose) might depend causally (to some extent, perhaps to a small extent) on the positions of the other nodes. This would yield a "distributive dependence on each and also collective dependence on all" (2011b: 150). That option, he claims, "explains a web model for belief", adding that perhaps even more data should be added, having to do with the dynamical, historical dimension. Here, I am in broad agreement with Sosa and will appeal to these ideas when it comes to the issue of a priori justification.

We now pass to our specific topic, intuition. Sosa agrees with the tradition, in particular the rationalist one, that there are intuitions, which form a distinct group of phenomena, and that there is an intuition-disposition/competence. Intuitions are a special sort of intellectual seemings, attractions to assent to a proposition triggered simply by considering a proposition consciously with understanding. Sosa is, to my mind correctly, optimistic when it comes to the reliability of intuitions: "By analogy to the seemings delivered by our visual system, the intuitions immediately delivered by our rational competences are preponderantly true, even if occasionally false. This is why those rational mechanisms are intellectual *competences*, because they systematically lead us aright." (Sosa 2007: 60).

The epistemic quality of the intuitional (as well as perceptual or introspectional) appearances depends on the quality of the underlying virtuous dispositions, formed in normal circumstances. And this brings us to issues of justification: "All seemings delivered by such competences are thereby epistemically justified", writes Sosa in the immediate sequel.

So, appearance is fallible, but still a fundamental source of justification. It is such even in the case of paradoxes, robust, powerful and entrenched, and thus a source of justification. But sometimes it can fail to be a source of knowledge, e.g. in the evil demon world, or in a world where one memorizes arithmetic tables by using a manual containing a large number of errors. For the sake of coherence we have to find some correlation between appearance as a source of justification and appearance as a source of knowledge. When is appearance a trustworthy justifier? Intellectual appearance is a trustworthy justifier when based on understanding, we are told. And here is the gist:

> S *rationally* intuits that p if and only if S's intuitive attraction to assent to <p> is explained by a competence (an epistemic ability or virtue) on the part of S to discriminate, among contents that he understands well enough, the true from the false, in some subfield of the modally strong (the necessarily true or necessarily false), with no reliance on introspection, perception, memory, testimony, or inference (no further reliance, anyhow, than any required for so much as understanding the given proposition). (Sosa 2007: 61)[3]

[3] Here is a longer quote, Sosa's answer to the question of "Just How Can Understanding Function as a Source of Epistemic Standing for Intuitions?":

So much about the first level justification of an intuition, with apologies for brevity.

At the second, meta-cognitive level, we encounter the thinker's reflective awareness of the quality of her first-level source, e.g. her reflective questioning of or trust in her intuitions. Thinkers, including ourselves, spontaneously find their intuitions true in a very compelling manner, and therefore, on a reflexive level, consider their intuition-capacity and their reason generally de facto reliable. This reliability of the first-order source, if available, yields an external, third person justifiedness. In contrast, the reflexive or meta-cognitive, second-order trust in one's own reliability, if justified, would make us, the thinkers, reflectively justified on the second level. As reflective creatures aiming at truth, we need both levels of justification for our first-order beliefs, including the intuitional ones. Such a two-level view of justification has been probably implicit in classical epistemology (Descartes), and is nowadays proposed by various authors, not only Sosa, but also K. Lehrer, W. Alston and J. van Cleve, with a lot of difference of detail. With Sosa's general view of reflective justification one would expect that the thinker may and ought to use on the second level of reflexive questioning all the available sources in order to assess the reliability (and other virtues) of a given first-order source, in this case of intuition or reason. However, when we look at Sosa's actual pronouncements, it seems as if with intuition and introspection the reflective justification is just a matter of indubitability. As regards the latter, Sosa is clearly in favor of circular self-validation, an appeal to the deliverances of our introspective faculty under reflective perspectival consideration (2000: 10). At the second level we are prompted to see necessary infallibility (reliability) of our first level introspective thoughts. And no causal knowledge is to be involved: "For the cogito the explanation of infallible reliability… skirts both causal tracking and construction or judgment dependence."

> Fundamental, intuitive rational beliefs are based at least on understanding of the propositions believed, or so it has been argued above. It is not, however, just the understanding of a proposition, whatever its content, that gives a proper basis for believing it. Otherwise, it would also constitute a basis for believing its negation, which must be equally well understood. Not even the highest pitch of clarity and distinctness will suffice. (…) What suffices is rather the being understood (shared by a proposition and its negation pretty much equally) along with the specific content of that very proposition. Can we go beyond this to some general feature that, when combined with the being understood, will properly yield acceptance? As we have seen, neither simplicity nor truth is such a feature, either singly or in combination.
>
> Nevertheless, something distinguishes simple truths of arithmetic or geometry, for example, making them suitable objects of immediate acceptance upon understanding, and giving them their attraction to normal human minds universally (upon understanding). Whatever it is, whether innate or socially instilled, its yield is uniform and general enough to suggest dispositions at work (whether wholly individually seated, or partially socially seated). Given how epistemically benign they are, finally, such dispositions seem not inappropriately considered "competences." (Sosa 2014: 49)

(2002: 376). Note that this is not an account of first-level justification but an account of "the required epistemic perspective" on the reliability of our source of a priori justification. "It is only when we see the cogito as not just infallible, but also indubitable (upon consideration) that we grasp the fuller Cartesian account" (Ibid.). The same holds for typical intuitional beliefs. Their reliability is comprehensible "through reflection about the content-determining conditions of our thought." (2003:183). Take logic: "Results in proof theory, or in metatheory more generally, might thus explain why it is that our thoughts in the relevant fields are likely to be right, or even *bound* to be right, if we follow certain methods" (2003: 183–4). And Sosa then raises the crucial, albeit rhetorical question: "What rules out the possibility of such general understanding of our own reliability on the *a priori*, precisely by means of properly directed *a priori* theorizing?" (Ibid.) We justify the a priori by the a priori.[4] This will be the main topic of our discussion.

3. *Discussion and open questions: epistemic perspective and its requirements*

a) *First level: the sources of competence*

The first question that I would like to raise is a relatively minor one. Sosa claims that the foundationalism/coherentism contrast is a false dichotomy, and accepts coherentism at the reflective level. But what about the first component, foundationalism? In the case of intuitions one often has strong seemings in favor of some given option, for instance, in the case of the Ship of Theseus that the still sailing ship is identical with the original one. Sosa interprets these seemings as conceptually grounded inclinations to believe (e.g. in 2011a: 456; he explicitly mentions logic but the context seems to point to a general view encompassing philosophical intuitions as well). Now, he finds them justified only if they derive from reliable competence; the fact that the thinker is responsibly responding to the seeming itself, independently from the externalist considerations of reliability, does not even *prima facie* justify her (if I got Sosa right). But then, it is strange that coherentism appears at the sophisticated level, one of reflection, and the traditional foundationalism of clear and attractive appearances plays no role: the dichotomy seems to re-emerge, although it was deemed to be false. If Sosa accepted that at least some *prima facie* justification is bestowed to the belief by the seeming-attraction he would end up with two tiers, the coherentist and the foundationalist, in addition to the

[4] Here is another formulation from the same context:

"Whether that project can succeed or not, anyhow, the fact remains that its success would give us an *a priori* component for our desired epistemic perspective, a component that in the respect of being substantially *a priori* would match the Cartesian epistemic reflections traditionally accorded the highest explanatory efficacy in epistemology" (Sosa 2003: 85).

two levels. The question for professor Sosa is then whether this would be acceptable for him, and if not, why.

The second question concerns competence. It is the central factor both in the production and in the justification of intuition. But what explains competence? Sosa is hopeful that "epistemic competences can be of use in epistemology even in the absence of a detailed theory of their nature and operation." (2013: 200). He claims that we can appeal to them "even with limited understanding of their modus operandi" (Ibid.) and directs the reader to his "Minimal intuition" (1998) paper. However, this early paper offers only a minimal(ist) answer: one can restrict one's confidence in intuitions worthy of being trusted, and this will yield some perspective on one's capacities (1998: 267).

And this low level of demands for reflective justification holds for all competences, perception included. Having read *Reflective knowledge* one would have expected more. And a specific worry about intuition also arises here. For perception, people had some idea of things acting upon our senses through some kind of medium. Even a moderately sophisticated person, at any time in written history, would be aware that she sees things through the intermediary of light: no light, no seeing. The same for hearing, taste and touch. She would know that a rough surface acts upon her fingers when she is touching it, and that the touching becomes more risky if the surface is hot, turning the sensation of touch into intense pain. However, as we are painfully reminded in the discussion of Benacerraf's dilemma, there is no convincing commonsense story about the build-up of our numerical competence. What about philosophical intuitions? Metaphysical ones, moral ones and so on?

We can gauge the importance of the question by taking a glance at Sosa's main resource for dealing with criticism of intuitions for their unreliability, alleged or real. When presenting his theory that the attraction or belief is justified because it is competent (2007: 59), and facing the problem of fallacies in reasoning, Sosa introduces the Chomskian performance-competence distinction:

> Fallacies can thus be viewed as performance errors chargeable against the subject, by contrast with deliverances of a competence. Unlike the Cartesian assimilation model, this account *can* admit the fallibility of intuition, can allow that paradox-enmeshed propositional contents exert proper attraction, on which one might even base justified intuitive belief. (Sosa 2007: 59)

His account is subtle: some errors are due to performance deficits, other to natural defects of the competence (early formulation already in Sosa 1998: 261). I agree with him that this is a perfect starting point for settling problems with X-phi, and I applaud most things he has to say in this context (in his (2010), and (2011a) papers).[5]

[5] He brings in comparison with perceptual illusion, using the Mueller-Lyer one as his example. And he notes that the attraction is a deliverance of the normal human visual system, and is to be put on the account of the competence itself.

Two question arise at this point for the epistemologists constructing a theory. First, how do we distinguish between performance errors and the deficits of competence unless we know more about how the latter works? Second, if Sosa is really alluding to the Chomskian performance-competence distinction, his use of "competent" is a thick one; the attraction and belief are competent only if they are derived from the corresponding (virtuous) competence. The same presumably holds for beliefs that <3+2=5>, or that <A square has four sides>, listed earlier in the chapter (Sosa 2007: 46). It is here that the Chomskian problematic shows its bite: mental linguistic competence is reliable, even virtuous in Sosa's term, because *it* dictates what the corresponding language is and is like. The order of determination goes from competence to its product. And the order of determination secures the lion's share of justification: I am competent in believing that my linguistic judgment is OK because I am judging *my* language(modulo all the reservations having to do with thorny issues of the relation between idiolect and various sociolects; here we only rehearse the main point.)

This order of determination is not available for examples like Modus Ponens, <3+2=5>, and <A square has four sides>, unless one goes strongly anti-realist and response-dependentist about logic, arithmetic and geometry. Of course, Sosa does not take this line. Here, it is not the thinker's competence that makes it true that a square has four sides, and the like, but rather the opposite holds: the mental structure-module and its functioning is virtuous *because* it conforms to an independent mathematical fact. And indeed he reiterates the contrast between justification by testimony and justification by understanding. He distinguishes between *determination of best opinion* and *tracking by best opinion. Determination by best opinion*: 5 being prime because best opinion would think it prime. *Tracking by best opinion*: 5 being thought prime by best opinion because it is prime (Sosa 2002: 370).

But once the realist stance is taken, the issue of explanation becomes pressing. How is the tracking secured? Sosa appeals to the alleged origin of intuition from understanding, primarily a conceptual one. But we need more. To return to the Ship of Theseus example, what is it about concepts that makes one think that the still sailing ship is identical with the original one? Something about the concept "ship"? Or "material object"? But what makes our concept track reality in these matters? Similarly with simple mathematical beliefs. If our concepts have the impressive power to put us in touch with mathematical reality, we should at least have some inkling of how this is possible. Sosa's role-model epistemologist, Descartes, was ontologically committed to the existence of God, and used that commitment as the epistemic guarantee of the cogito propositions. *Scientia* (reflective knowledge) is all about such a commitment. Remember that the reflective perspective has to give us a substantive understanding of our first-level beliefs, at least in cases in which we want to say that we know full well what we are talking about.

Here, then, is the second question for professor Sosa: do you have a view about how our intuitional competence connects with the world it produces judgments (more precisely, attractions to judge) about? Intuitions seem to connect us to mathematical and modal reality, to facts of metaphysical significance (the Ship of Theseus), moral significance (the Trolley problem), and perhaps more (linguistic intuitions, etc.). Let us agree that concepts are somehow involved in the feat; but how precisely?

b) *Second level: intuition, broad coherence and the web of belief*

Let us now pass to our main topic, the reflective justification of intuitions and the role and character of epistemic perspective, and to some possible substantive disagreement**.** Since Sosa often discusses introspection in the same breath with intuition, as a related a priori source, we shall follow him and occasionally mention introspection in the context. What does reflective perspective involve in the case of intuition and introspection?

We know that it cannot be just a reiteration of the first-level thought (this much is clear from Sosa's criticism of Alston); it seems then that coherent perspective is the only candidate. The reader who has formed her impression from reading Sosa's main statements on the nature of reflective knowledge in his (Sosa 2009a and 2011b) would probably expect the stress on broad coherence: after all, it is such coherence that takes us from mere *cognitio* to *scientia*. She would also remember the metaphor of the web, and its strongly coherentist morals: every node (belief) is to some extent justificationally connected to every other. She would keep in mind that the web of belief is connecting us causally to the facts in our environment (Sosa 2011b: 150).

Of course, such a reader is in for surprises. It looks as no appeal to coherence is involved, not even the very narrow one, let alone the broad one praised as crucial for the reflective level in the general cases. How significant and how *ad hoc* this exception is can be seen from comparison with Sosa's general reading of Descartes, which stresses the importance of a circle, and of a more systematic, coherence-seeking reflection at the second level.

We noted that with intuition and introspection it looks as if justification is just a matter of indubitability and of circular self-validation, an appeal to the deliverances of our introspective faculty under reflective perspectival consideration. (Sosa 2000: 10). We are prompted to see necessary infallibility (reliability) of our first-level introspective thoughts. And no causal knowledge is to be involved: "For the cogito the explanation of infallible reliability … skirts both causal tracking and construction or judgment dependence." (Sosa 2002: 376). Note that this is not an account of first-level justification but an account of "the required epistemic perspective" on the reliability of our source of a priori justification. "It is only when we see the cogito as not just infallible, but

also indubitable (upon consideration) that we grasp the fuller Cartesian account" (Ibid.). How does this differ from Alston's same-sort reflective justification, in which a capacity is second-order justified just by the re-application of itself, the very competence to be justified? And Sosa has been quite critical of Alston's strategy (1994). He has insisted on the holistic character of reflective justification, in contrast to Alston's project of "homogenous" justification of the similar by the similar.

Again, as mentioned in section 2, we are offered an analogous account of human knowledge of elementary mathematical truths and other necessary propositions. Intellectual appearance is a trustworthy justifier when based on understanding. Sosa reiterates the claim in a more recent paper: "What distinguishes intuitive justification is that the entertaining itself (with adequate understanding) of that specific content exerts its attraction while rationally unaided. Intuitions are reason-based in a way that does not go beyond conscious grasp of the specific propositional content" (Sosa 2014: 48). Sosa is happy to note that this circumvents the challenge of Benacerraf's dilemma: we do not need to connect the Platonic facts with our knowledge of them by any sort of explanatory route, other than the claim that our concepts (or even mere symbols) can put us in touch with relevant mathematical properties (Sosa 2002: 380). Not much is left of the web metaphor, and the claim that it connects us with facts in our environment in an intelligible way. Remember, it was claimed that "[j]ustified beliefs are nodes of a web properly attached to the environing world through perception and memory" (2011b: 150).[6]

The web-pipe contrast is also gone: the understanding that produces intuition justifies it on the reflective level. Connections of every belief (node) with all other nodes seem to be totally irrelevant; even the connection of any kind with any of the non-homogenous nodes disappears!

Let me then try to reconcile Sosa's holistic, web-guided understanding of reflective justification in general with his non-holistic understanding of reflective justification for intuitions (and of introspective beliefs). First, we both agree that pieces of belief of thoughtful researchers are

[6] Sosa claims that no causal relation between our cogito (or mathematical) belief and the fact believed would explain the reliability of that belief:

> Take an intricate spider's web with its many nodes, attached at various points to various surfaces. The position of each node might then depend causally (to some extent, perhaps to a small extent) on the positions of the other nodes. Here there is distributive dependence on each and also collective dependence on all. That explains a web model for belief (though beliefs also occupy an important dynamical, historical dimension, one that requires a more complex web model). Any given belief node is in place through its connections with other nodes, but each of them is itself in place through its connections with the other nodes, including that original given node. (Sosa 2011b: 150)

Not a single element from this general characterization of reflective justification applies to the particular cases of intuition and introspection. What could justify making such an exception, without even presenting it as such?

typically reflectively justified, in addition to being immediately, first-level justified. (We also agree that reflective justification at its highest involves meta-knowledge of risks, but we shall leave that aside here).

Second, we agree that holism is feasible, the way Sosa presents it in his *Reflective knowledge* (2009a) and in *Knowing full well* (2011b). We agree very much with the following claim of his:

> Reflective endorsement may now take its place in the web with no apparent special problems. Through our growing knowledge of ourselves and of the world around us and of the relation between the two, we come to see our modes of rational basing and other belief acquisition as sufficiently reliable. This enables us to endorse such modes reflectively as truth-reliable, of a sort to lend epistemic justification to our commitments and beliefs. (Sosa 2011b: 151)

Thirdly, and most importantly, reflective justification typically mobilizes and indeed should mobilize capacities distinct from the original capacity that has produced the belief-candidate for being justified, in order to assess the reliability of the original capacity. It has to go beyond justifiers that are of the same kind ("homogeneous") as first-level immediate ones, in order to enlarge the circle of justification (and thus avoid viciousness), and is, therefore, holistic and coherentist. Sosa is quite explicit about this in his writing about broad coherence, and explications of the web-metaphor underline it, with insistence of the connection of each with every node. Moreover, such a holism is commanded by the requirement of total evidence.

If this holds, reflective justification of armchair beliefs, presumably produced by intuition and some reasoning, should revert to empirical considerations testifying to the reliability of intuition and reasoning.

This bring in the a priori/a posteriori contrast. The last paragraph, if correct, suggests a further conclusion: reflective justification of armchair beliefs typically combines, in an articulated way, a posteriori elements contributing to the thinker's reflective trust in her armchair capacities with some, presumably a priori, components.

Let us be a bit more specific. In his general exposition(s) Sosa stresses the importance of explanatory coherence. Apply this to intuitions and armchair beliefs. Note that the reliability of armchair beliefs is *prima facie* puzzling. A reflective assessment of armchair beliefs is therefore incomplete in total absence of explanation of their having and reliability (as is the case with perceptual beliefs). The explanation has to be to some extent causal or causal-like. Barring the Cartesian style *a priori* theological grounding, which is very dubious, any such explanation will involve appeal to empirically believed assumption. So, the explanation of having and reliability will have essential empirical explanatory components. Therefore, reflective justification of armchair belief will have essential empirical components. It will have an important *a posteriori* component, with a clearly defined role.

Another candidate for a reflective source of information about reliability is the well-known appeal to global unavoidability and indis-

pensability: unavoidability and indispensability of logic and elementary mathematical understanding for any kind of cognitive project, call them global unavoidability and indispensability, are an important reflective justifier of logical and mathematical beliefs and inferential propensities, perhaps the most important one. This justifier can justify the target beliefs and propensities, only if our global cognitive project is a meaningful one, with some chances to succeed. The issue of success of our global cognitive project is to a large extent an empirical matter, so that we are justified in being optimistic about it on the grounds of already achieved empirical and empirically detectable success. The issue of reflective justification of logical and elementary mathematical beliefs and inferential propensities is to be decided to a large extent on the basis of global successfulness of our cognitive effort, which is largely an *a posteriori* matter. If this holds, logic and elementary mathematical understanding are reflectively justified *a posteriori* to a significant degree.

Interestingly, Sosa himself has been aware for a long time of some of the difficulties listed above. In his 2000 paper he notes, among possible objections, the "locality of cogito": cogito is a single proposition, but we need certainty over a wider span of propositions. He answers that the relevant feature of the cogito is not restricted to a single belief. Unfortunately, he does not discuss the analogous "locality" of many necessary propositions, which seem to cry for a more holistic treatment. More importantly, he notes that "reliability of our a priori beliefs could hardly be sustained purely a priori. For we need a grasp of the mechanisms participating in the beliefs' acquisition, which is a posteriori." (Sosa 2000: 14). He seems to endorse the claim and answers: "Second-level defense (revision) of our a priori dispositions is also not purely a priori" (Sosa 2000: 13). He directs the reader to a claim stated on the preceding page: "[I] agree that the reflective defense of our mathematical and other a priori beliefs will not be purely a priori" (Sosa 2000: 13). But again, after having said this, he retreats, and stresses the possibility that in the case of the cogito, nothing will be needed beyond a priori beliefs, and no causal commitments will become prominent (Sosa 2000: 14). Elsewhere, he stresses the positive consequences of dropping the claims of apriority: "And once any claims of priority are dropped, as I am proposing, then it might well be held that *cognitio* that p and *cognitio* that one enjoys *cognitio* that *p*, are both required for *scientia* that *p*" (Sosa 2009a: 150, fn. 14).[7]

[7] And there is a streak of explanationism and interest in causal dependence in Sosa's general picture of justification:

> Epistemology too, like the aesthetics of dance, reverses the import of causality found in instrumental value. The distinctively epistemic evaluation of a cognitive performance can depend substantially on its source, unlike the instrumental evaluation that depends on effects rather than sources. Consider thus the justification of a belief derived from a good inference, as when a detective figures out who did it, or when you determine how much you owe a shopkeeper.

Let me conclude with an example which shows that Sosa does in fact recognize the importance of empirical, a posteriori data for the full reflective justification of intuitional beliefs, in this case philosophical ones. In his paper on possible intuitional foundations of philosophy (2011a) he confronts the issue of possible serious divergence in subject responses to questions in thought experiments. He comes up with a dilemma: "Either experimental inquiry will uncover serious divergence in subject responses or it will not", and argues that in the latter case there is no serious problem (Sosa 2011a: 465). If the first horn turns out to be actual, we still have a way out: explain away the disagreements by differences in semantic understanding. (I apologize for brevity of presentation). Note that the moves are made in response to empirical findings, and that the last, rescuing move, would also need empirical confirmation, namely the finding that in fact the subjects have different meanings on their minds when performing the armchair experiment. It is all a piece of clearly a posteriori reflection on philosophical intuitions, geared to offering a sophisticated second-level justification, and indicating the limits of thought experimenting, thus resulting in our knowing full well the philosophical propositions in question.

To reiterate. The puzzle concerns the question of what reflective knowledge in the domain of intuitions involves? The options seem to reduce to the following three:
(a) localistic ratification and self-validation: this is how Sosa characterizes the Reflective condition for introspection, and by implication, intuition: seeing the necessary infallibility (reliability) of our first level introspective thoughts (Sosa 2000: 9), or
(b) rather holistic coherence, in line with the general picture of reflective justification. The resulting picture is then either (i) narrow, involving only a priori materials, or (ii) wide? If (ii) wide, then reflective justification becomes to a significant degree a posteriori. If (i) narrow, it is unclear why an exception is made for intuitional beliefs.

My final question for professor Sosa is then concerned with his considered judgment about the role of the a priori and a posteriori in the reflective justification of intuitions: how are reflective knowledge, broad coherence and the web of belief related in the case of intuitional knowledge? And in particular, in the case of armchair philosophical intuitions?

> Something is then believed because it is concluded from prior information already in the thinker's possession. To draw it as a conclusion and to believe accordingly for that reason is, moreover, a broadly causal matter. It is a matter of believing such and such *because* of so and so, or *on the basis* of a prior belief that so and so. Accordingly, the conclusion belief gains its epistemic status through being *based* on the premises inferentially. One believes the conclusion at least in part on that basis, *for the reason* that, as one can see, it follows from the already accepted information. The fact that one's belief in the conclusion is thus "motivated rationally" (Sosa 2007: 80).

4. *Conclusion*

Let me conclude with two points. First, I would like to sketch the road a two-level virtue epistemologist could, and perhaps should, take, if she takes seriously the broad coherence (and in particular, its explanatory component) at the second reflective level; I myself did take such a road in writing about intuitions and the a priori (Miščević 2006, 2008). Second, I would like to reiterate my three questions to professor Sosa.

The first task first. Both Sosa and I agree that obviousness and indubitability give the thinker a *prima facie* reason for accepting one's intuitions. Neither a Sosa-style moderate externalist nor any naturalist should deny this; the latter since for a naturalist normative acceptability should follow from descriptive compellingness. In the next step, as Sosa has taught us, the thinker tries to achieve a general coherent view of her cognitive abilities and their outputs. Of course, one can distinguish degrees of reflective, meta-cognitive achievement on the second, reflective level. The lowest degree is guaranteed by the immediate attraction-compellingness of contents, i.e. of intuitional propositions. If the thinker psychologically cannot doubt some such proposition, then she is *prima facie* allowed to believe it: epistemic *ought* implies epistemic *can*. Still, a more conscientious thinker would want to have a coherent meta-cognitive perspective on deliverances of her cognitive abilities, and an explanatory view on functioning of abilities. Again, we may distinguish the immediate or folk view (of e.g. perception or intuition-ability) from a theoretical perspective on these abilities.

As Sosa puts it in his general proposal, it is the interplay of (the deliverances of) all capacities, plus the best explanation of the whole, that indicates whether a particular capacity, in this case intuition is reliable. (I assume that merely negative coherence with explanation is sufficient: in other words, if the explanation does not seriously contradict the explanandum (and we have argued that it does not), we have good reasons to trust our intuitions.) If we apply it to intuitions, we note that the explanation-based doubts about intuition, for example, make it vivid for the thinker that the immediate compellingness of an intuition need not be sufficient. But then, the indispensability and success come in. Our intuitions cohere with our empirical hypotheses, and enable these hypotheses to be tested and confirmed. Indispensability and success are thus capable of almost completely justifying the reliance on intuitional knowledge. They come in very handy since they point to the massive empirical success of everyday knowledge and of science in which such beliefs are essentially used. The success does a posteriori vindicate the certainty of elementary logical and mathematical intuitions, for which there is a massive overlap with the factual domain.[8]

[8] There is a reasonable philosophical worry that some flaw in the origin of our intuitions might annihilate their justification. What if they come from a demon, Descartes asked. What if they are just figments of our imagination? How can information coming from within have any "validity" for a mind-independent world,

What about concepts and understanding? How do our concepts guide us? My answer is more extroverted than the traditional conceptualist one: the crucial quality about our competence is that it carries correct information about the world, and concepts are just a means for encoding such information. Intuitions are concerned with their external objects, the domain of items and facts, rather than with concepts. Further, they require an explanation of having and reliability, if possible a causal one. Concepts often play a role in the process, but they are not the object of intuitions, and their role is subordinate to the role played by the external referential domain. But why are they normally so helpful? My own favorite line is summarized in the following paraphrase of J. L. Austin (where the term "words" is replaced with "concepts"):

> …the stock of our deepest concepts embodies all the distinctions men have found worth drawing, and the connections they have found worth making, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up on the spot. (Austin 1979: 182)

This accumulated wisdom then allows the philosopher to anticipate the experience from the armchair. At the same time, the double fallibility of these intuitions accounts for the limits of philosophical autonomy: armchair research should be open to corrections from empirical science.

Also, our partly innate endowment might explain at least the very origin of the intuition-capacity and the initial stages of the formation of our intuition-states with their contents. For instance, it might consist of innate structures, some corresponding to concepts and some to inner, spatio-temporal "frames", responsible for an innate spatial-geometrical know-how. Note that intuitions are often rather scenario-based than inference-based. Imagining scenarios, typically particularized ones play the main cognitive and justificatory role, whereas inference typically plays a subordinate role.

All this explains some of the objective validity of our intuitions. But nativism should be restricted to the origin of the system and to the relatively initial stages of processing. An intelligent nativist-adaptationist should allow for a wide margin of influence from individual empirical learning, which may overthrow even some deeply ingrained pre-conceptions to the contrary. And most importantly, intuition is doubly fallible. It can misrepresent the contents of our cognitive apparatus, and is thus internally fallible. But, the contents themselves—including their core, the innate assumptions—are also fallible, as Sosa has noted, yielding

Kant asked (and opted for an anti-realist solution). It is here that the evolutionary explanation comes in. Its role is remedial, i.e. to alleviate or to forestall the subtle, purely philosophical skepticism focusing upon a distantly and merely possible flaw in the causal ancestry of our intuitions. It thus removes the lingering perplexity about the mystery of scientific applicability and success of our logical and mathematical intuitions.

the external fallibility of intuitions. Our innate geometry might be false, our possibly innate folk-physics certainly is. No deep or strong apriority is involved in their deliveries. In short, we can admit an important role of intuitions, and preserve some of their special status, intimated by their phenomenology, without falling into the dangerous traps of classical Cartesianism.

Finally, concerning the justificatory status of intuitions, one needs to combine *a posteriori* considerations with the *a priori* ones; the result will be a structured justification, with distinct elements coming from distinct sources. So much about the reflective-level proposal, to some extent inspired by Sosa's stress on broad coherence and the role of the web-of-belief as a whole.

Let me conclude by re-iterating my questions for professor Sosa:

If Sosa accepted that at least some *prima facie* justification is bestowed on the belief by seeming-attraction, he would end up with two tiers, the coherentist and the foundationalist one, in addition to the two levels. The question for professor Sosa is then whether this would be acceptable for him, and if not, why.

The second question: how does our intuitional competence connect with the world it produces judgments (more precisely, attractions to judge) about? Intuitions seem to connect us to mathematical and modal reality, to facts of metaphysical significance, moral significance and perhaps more (linguistic intuitions, etc.). Let us agree that concepts are somehow involved in the feat; but how precisely?

My final question for professor Sosa is then concerned with his considered judgment about the role of the a priori and a posteriori in the reflective justification of intuitions: how are reflective knowledge, broad coherence and the web of belief related in the case of intuitional knowledge? And in particular, in the case of armchair philosophical intuitions?

References

Austin, J. L. 1979. "A Plea For Excuses". In Austin, J. L. *Philosophical Papers*. Oxford: Cleardon.

Boghossian P. 2009. "Virtuous intuitions: comments on Lecture 3 of Ernest Sosa's *A Virtue Epistemology*". *Philosophical Studies* 144: 111–119.

Miščević, N. 2006. "Intuitions: The discrete voice of competence". *Croatian Journal of Philosophy* 6 (3): 523–548.

Miščević, N. 2007. "Naturalism and philosophical analysis". *Proceedings of Kirchberg Philosophy Congress* 2005. Vienna: Holger-Pichler.

Miščević, N. 2008. "Can Concepts Ground A priori Knowledge? Peacocke's Referential Turn and its Challenges". *Acta Analytica* 23 (3): 233–256.

Sosa, E. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press.

Sosa, E. 1994. "Philosophical Scepticism". *Proceedings of the Aristotelian Society, Supplementary Volumes* 68: 263–307.

Sosa, E. 1998. "Minimal intuition". In DePaul, R. M. and Ramsey, W. (Eds.). *Rethinking Intuition*. Lanham: Rowman and Littlefield.

Sosa, E. 2000. "Modal and Other A priori Epistemology". *The Southern Journal of Philosophy* 38: 1–16.

Sosa, E. 2002. "Reliability and the *A priori*". In Szabo Gendler, T. and Hawthorne, J. (Eds.). *Conceivability and Possibility*. Oxford: Oxford University Press.

Sosa, E. 2003. "Ontology, Understanding and the *A priori*". *Ratio* 16: 178–188.

Sosa, E. 2006. "Intuitions and Truth". In Greenough P. and Lynch, M. (Eds.). *Truth and Realism*. Oxford: Clarendon Press: 208–226.

Sosa, E. 2007. "Intuitions". In Sosa, E. *A Virtue Epistemology: Apt Belief and Reflective Knowledge Vol. 1*. Oxford: Oxford University Press.

Sosa, E. 2009a. *Reflective Knowledge, Apt Belief and Reflective Knowledge, Volume II*. Oxford: Clarendon Press.

Sosa, E. 2009b. "A Defense of the Use of Intuitions in Philosophy". In Bishop, M. and Murphy M. (Eds.). *Stich and His Critics*. Oxford: Blackwell: 101—112.

Sosa, E. 2010, "Intuitions and meaning divergence", *Philosophical Psychology*, 23:4, 419–426.

Sosa, E. 2011a. "Can There Be a Discipline of Philosophy? And Can It Be Founded on Intuitions?". *Mind & Language* 26 (4): 453–467.

Sosa, E. 2011b. *Knowing full well*. Princeton: Princeton University Press.

Sosa, E. 2013. "Intuitions and Foundations". In Casullo, A. and Thurow J. C. (Eds.). *The A Priori in Philosophy*. Oxford : Oxford University Press.

Sosa, E. 2014. "Intuitions, Their Nature and Probative Value". In Booth, A. R. and Rowbottom, D. P. (Eds.). *Intuitions*. Oxford: Oxford University Press: 36–50.

Sosa, E. 2015. *Judgment and Agency*. Oxford: Oxford University Press.

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

Recognitional Identification and the Knowledge Argument

ERHAN DEMIRCIOGLU
Koç University, Istanbul, Turkey

Frank Jackson's famous Knowledge Argument moves from the premise that complete physical knowledge about experiences is not complete knowledge about experiences to the falsity of physicalism. Some physicalists (e.g., John Perry) have countered by arguing that what Jackson's Mary, the perfect scientist who acquires all physical knowledge about experiencing red while being locked in a monochromatic room, lacks before experiencing red is merely a piece of recognitional knowledge of an identity, and that since lacking a piece of recognitional knowledge of an identity does not entail lacking any pieces of knowledge of worldly facts, physicalism is safe. I will argue that what Mary lacks in her room is not merely a piece of recognitional knowledge of an identity and that some physicalists have failed to see this because of a failure to appreciate that Mary's epistemic progress when she first experiences red has two different stages. While the second epistemic stage can perhaps be plausibly considered as acquiring merely a piece of recognitional knowledge of an identity, there is good reason to think that the first epistemic stage cannot be thus considered.

Keywords: The knowledge argument, Frank Jackson, John Perry, the phenomenal concept strategy.

1. *The Recognitional Strategy*

Jackson's (1982) *Knowledge Argument* is one of the most intuitively compelling arguments against physicalism (roughly, the thesis that our world is entirely physical). This famous argument moves from the premise that complete physical knowledge about experiences is not complete knowledge about experiences to the falsity of physicalism. Consider Mary, a perfect scientist who has all the physical knowledge about experiencing red and yet who has not experienced red before. The intuition is that when Mary leaves her room and sees a ripe to-

mato, she will be surprised and exclaim "So, *this* is what it is like to see red!", and will thus acquire a new piece of information about experiencing red. And, since physicalism implies that given her complete physical knowledge, Mary knows everything about experiences of red, Jackson argues, physicalism is false.

Some physicalists have countered this argument by arguing that what Mary lacks before experiencing red is merely a piece of recognitional knowledge of an identity, and that since lacking a piece of recognitional knowledge of an identity is not, or does not entail, lacking any pieces of knowledge about worldly (coarse-grained) facts,[1] physicalism is safe. Let us call this reply to the Knowledge Argument "the Recognitional Strategy" (briefly, RS). RS finds one of its clearest and most systematic expressions in John Perry's admirable *Knowledge, Possibility, and Consciousness* (2001); and, the account offered there will be the main focus of this paper, though the lessons that will be drawn throughout will be general.

An instructive way to get a better idea of RS is to appeal to an analogy. Consider the following case from Perry (2001: 119). Perry might know that Fred Dretske wrote *Knowledge and the Flow of Information* (KFI) even if he does not know that *that* (man) [perceptually demonstrating Dretske at a party] wrote KFI. In this case, Perry fails to recognize (and hence lacks the recognitional knowledge) that *that* is Dretske. If Perry had known that *that* is Dretske, he would have inferred that *that* (man) wrote KFI from his previous knowledge that Dretske wrote KFI. However, lacking this piece of knowledge (i.e., that that (man) is Dretske) is not, or does not entail, lacking any piece of knowledge about worldly facts because Perry already has certain pieces of knowledge that have the same worldly content as that piece of knowledge (e.g., that Dretske is Dretske). (Similarly, Perry's coming to know "*that* is Dretske" is merely a matter of coming to recognize an identity and cannot thus be thought of as coming to possess any piece of knowledge about a new worldly fact.) It seems that if the Knowledge Argument had a valid form, then Perry's failing to know that *that* (man) is Dretske would entail that *that* (man) is different from Dretske while, *ex hypothesi*, they are the same.

RS claims that Mary's entire new knowledge can be expressed by "*this* is what it is like to see red" and also that this piece of knowledge is the same in kind as the piece of knowledge Perry comes to possess

[1] It is commonplace to distinguish two different ways of individuating facts. *Fine-grained* facts are individuated in terms of the concepts the subject has of the things in the world; *coarse-grained* facts are individuated in a way insensitive to those concepts. So, the fact that there is a bottle of water in my backpack and the fact that there is a bottle of H_2O in my backpack are two different facts if "fact" is understood in a fine-grained way (since a subject can believe the former without believing the latter), but are the same fact if it is understood in a coarse-grained way (since water is H_2O). When I speak of "worldly facts" or "facts," what I mean is always coarse-grained facts.

when he learns that that is Dretske. "This [Mary's] new knowledge is", Perry writes, "a case of recognitional or identificational knowledge, as in the case with my knowledge at the party with Dretske" (2001: 147). According to RS, Mary's coming to know "*that* is what it is like to see red" is merely a matter of coming to recognize an identity and cannot thus be thought of as coming to possess any piece of knowledge about a worldly fact. Through experiencing red, Mary acquires, RS claims, a new *recognitional* concept (i.e., the one she expresses by "that") of the experience of seeing red she already knew under a physical/functional concept in her room; and, accordingly, the new bit of propositional knowledge she expresses by exclaiming "that is what it is like to see red!" is, it is claimed, merely a piece of recognitional knowledge of an identity.[2]

To make clearer the purported analogy between Jackson's Mary case and Perry's Dretske case, let us suppose, as Perry (2001: 99) does, that the texts Mary reads in her room have systematically named the subjective characters (what-it-is-likenesses) of color experiences and that Q_R is defined in one of those texts as the subjective character of experiencing red. So, suppose that Mary knows, *before* leaving her room, that Q_R is what it is like to see red, while she does not know what it is like to have an experience with the subjective character Q_R because she did not have an experience with Q_R. This bit of knowledge is analogous to Perry's first piece of knowledge about Dretske (i.e., "Dretske wrote KFI") in that just as the latter is "detached from my [Perry's] perception of him [Dretske]" (Perry 2001: 119), the former is detached from "an act of attending to a subjective character" (Perry 2001: 147). Upon seeing the ripe tomato, Mary comes to know that *this* is what it is like to see red, and this is analogous to Perry's coming to know, after Dretske introduces himself, that *that* is Dretske. The latter is attached to a perception of Dretske and the former to an act of attending to a certain subjective character. And, finally, combined with her previous knowledge, Mary infers that Q_R is *this*, where *this* picks out the red experience-type, just as Perry infers from his first and second bits of knowledge that *that* (man) wrote KFI.

According to Perry, the problem Mary's case poses for physicalism is best seen as an instance of the sort of problem Frege called to our attention: how can identities be informative? Now, there are (at least *prima facie*) good reasons to think that recognition of an identity does not require coming to possess any piece of knowledge of new worldly

[2] Loar also compares Mary's lack of the relevant piece of knowledge with ordinary recognitional failures: "Margot learns about the element Au and reads that people decorate themselves with alloys of Au. But she has never seen gold and cannot visually identify it: she lacks an adequate visual conception. She later is shown some gold and forms a visual conception of it, "that stuff,' and she acquires a new piece of information…to the effect that those previously read about embellishments are made of that stuff. [I]f the knowledge argument were unrestrictedly valid, it would follow that that stuff is not identical to Au" (2004: 223).

facts (and that failure to recognize an identity does not entail failing to possess any piece of knowledge of new worldly facts). When Perry fails to recognize that that [demonstrating Dretske] is Dretske, he does not thereby fail to know about a worldly fact; and, accordingly, if he recognizes that that is Dretske, he does not thereby come to know about a new worldly fact. This is again because that that is Dretske is the same worldly fact as that Dretske is Dretske, and the latter is something Perry already knows about. The basic idea here is simply that worldly facts can be represented (or conceived) in many different ways and, in order for a subject to know about a certain worldly fact, it is not necessary that she has access to all those different ways of representing (or conceiving) it. New knowledge does not necessarily involve knowledge of new facts.[3]

In what follows, I will simply grant for the sake of argument that failure to recognize an identity does not entail failing to possess any piece of knowledge of new worldly facts. I will argue against RS that what Mary lacks in her room is *not* merely a piece of recognitional knowledge of an identity and also that some physicalists have failed to see this because of a failure to appreciate that Mary's entire epistemic progress when she first experiences red has two different stages: while the second stage of her epistemic progress can be plausibly considered as acquiring a piece of recognitional knowledge of an identity, there is good reason to think that the first epistemic stage cannot be thus considered.[4]

This paper is hereafter divided into three sections. In section 2, I will briefly describe Nida-Rümelin's (2004) Marianna case and contrast it with Mary's case to show that there are indeed two different epistemic stages in question–the one that is shared by Mary and Marianna and the one that is had only by Mary. In section 3, I will argue that, contra Nida-Rümelin, Marianna's epistemic progress is *propositional* and also that there are indeed *two* knowledge arguments, rather than one,

[3] How should we understand the epistemic progress that Perry makes when he recognizes the identity in question? I take it for granted that an adequacy constraint on such an account is that it does not postulate or entail that the fact that Jameson is Jameson is different from the fact that *that* is Jameson. As is well-known, Frege's own solution to the problem of the cognitive significance of (some) recognitions of identities was to introduce "modes of presentation" but not to postulate further worldly facts: "Now if we were to regard equality [or identity] as a relation between that which the names 'a' and 'b' designate, it would seem that a = b could not differ from a = a (i.e. provided a = b is true)…A difference [between a = a and a = b] can arise only if the difference between the signs ['a' and 'b'] corresponds to a difference in the mode of presentation of that which is designated" (Frege 1993: 23–4).

[4] I do not claim any originality in arguing for the existence of two different epistemic stages (see Nida-Rümelin (2004)) or in holding that the first epistemic stage is not a matter of acquiring a piece of recognitional knowledge of an identity (see Stalnaker (2008)). The present essay aims to contribute to the literature by providing a clear account of the *nature* of the two epistemic stages and an answer to the question of *why* the first epistemic stage is not an acquisition of a piece of knowledge of an identity. See also fn. 19.

which can be distinguished with reference to the two items of knowledge that someone situated in a monochromatic environment lacks. In section 4, I will argue that Mary's extra epistemic progress *is*, while the epistemic progress shared by Mary and Marianna is *not*, a matter of recognizing an identity.

2. *Nida-Rümelin's Marianna*

Some misunderstandings regarding the nature and content of the new knowledge Mary gains after her release can be avoided by recognizing that there are different epistemic stages one might undergo in the process of obtaining information about experiences. When it comes to pointing out these differences, Nida-Rümelin's Marianna case (2004) is more helpful than Jackson's Mary. Like Mary, Marianna lives her entire life in a monochromatic environment. The central difference is, however, that when the happy day comes, rather than seeing ripe tomatoes and bananas and grass and the sky, she is randomly visually presented with four slides showing clear cases of blue, red, green and yellow but she is not told the names of the colors. Now Marianna does not know which of the four types of color experiences she has is, say, red nor does she know that having red experiences[5] is like *that*, where the indexical in question picks out the type of experience she has when presented with the red slide.[6] But still one can plausibly say that now that she has the experience of red, there is a clear sense in which she knows what it is like to see red.

One may object that there is a sense in which Marianna does not know what it is like to see red because she cannot identify red experiences *as* red experiences and knowing what it is like to see red requires such a recognitional ability: when she experiences the red slide, she is not in a position to recognize that she experiences red and justifiably verbally report "that is red." Does this mean that she does not know, in any epistemically respectable sense of the term, what it is like to see red? No, it only suggests that I should specify more carefully what I mean when I claim that after her red experience, Marianna knows what it is like to see red. What I mean is this: Marianna knows what it is like to have *this*, where *this* picks out the red experience-type. Marianna's gaining this piece of information is analogous to the case

[5] I take it as a plausible hypothesis that the sense in which experiences are red is not the sense in which physical objects are red. So, phenomenal redness is different from physical redness (see Chalmers (2004)). At this point, one may follow Peacocke (1983: 21) and introduce a primed predicate (e.g., red′) to pick out the phenomenal property of the visual experience that is normally produced by the presence of a red object. For convenience, I will not adopt this line and simply use unprimed predicates to characterize the experiences in question.

[6] For an account of how demonstrative terms can refer to types rather than tokens, see Levine (2010). For further discussion about how type-demonstratives bear on the Knowledge Argument, see Demircioglu (2012), Levin (2007), Loar (2004), and Perry (2001).

in which Perry has a little chat with Dretske at a party without recognizing him as Dretske and hence without knowing that, referring to him, that is Dretske. There is a sense in which Perry knows the man he meets even if he lacks the recognitional knowledge that, pointing at the person he meets, that is Dretske.

Marianna's case shows that there is an intermediate epistemic stage which goes unnoticed in Mary's case.[7] It seems intuitively plausible that Marianna learns something new when she is haphazardly shown different colors and hence she makes a sort of epistemic progress. However, she still misses the information that *this* is what it is like to see red, where *this* picks out the type of experience she has when she is presented with the red slide. There is still room for Marianna to make further epistemic progress because she experiences red without knowing that what she experiences is red.

3. *Two Knowledge Arguments*

An interesting question about Marianna's case is whether the epistemic progress she makes gives her the relevant item of knowledge which is crucial for the Knowledge Argument to work. After having the relevant visual experiences, Marianna evidently gains epistemic access to various new thought contents that are not available to her before. She is now in a position, for instance, to wonder whether ripe apples appear like *that* or *that*, where the indexicals respectively refer to the colors of the red and blue slides, and to entertain new hypotheses and make new guesses. What explains the fact that Marianna gains epistemic access to new thought contents is that she now knows what it is like to see red, blue, etc. Can an argument analogous to the original Knowledge Argument from Mary's case work in Marianna's case? Or does the property dualist need a further step of epistemic progress to make a viable case for the intended metaphysical conclusion—the falsity of physicalism?

Despite her new ability to entertain the corresponding thought, Marianna does not come to know that ripe apples appear like the red slide she has the experience of. Nevertheless, does she acquire any item of propositional knowledge? Through her color experiences, she knows what it is like to see red; but it is not clear whether this piece of knowledge raises any threat to physicalism because it is not clear that this knowledge has any propositional content, which can be expressed in propositional form by a suitable that-clause.[8] According to

[7] As Nida-Rümelin writes: "A disadvantage of Jackson's example is that it fails to distinguish two steps of epistemic progress that can be distinguished clearly in Marianna's case" (2004: 254). However, as will become clear, I do not agree with Nida-Rümelin's construal of the first epistemic stage of Mary's progress.

[8] The assumption here is that the kind of knowledge that is relevant to the Knowledge Argument is propositional, that is, that the Knowledge Argument purports to pose a threat against physicalism only if the kind of knowledge that is acquired through experiencing red is propositional. This assumption is supported by the idea that

Nida-Rümelin, Marianna "has not gained any new item of the relevant propositional knowledge" (2004: 254) and hence there is nothing imperiling physicalism at this stage of her epistemic progress. After all, one may ask what form the relevant proposition can possibly be given since she does not know *that that is what it is like to see red*, where *that* picks out the relevant type of color experience.

However, contra Nida-Rümelin, the following construal of Marianna's knowledge regarding what it is like to see red appears plausible:

(P) Marianna knows that it is like Q to have *this* (where *this* picks out the red experience-type).[9]

After being shown the red slide, Marianna acquires a new phenomenal concept, which is *Q*, that picks out the phenomenal quality (or subjective character) instantiated by experiencing red in a way different in kind from the concepts she had before.[10] It is controversial whether there really are phenomenal concepts that satisfy the characteristics they are thought to have (e.g., conceptual independence from physical/functional concepts or having their referents in their modes of presentation (Loar 2004)); but, for the purposes of this paper, their existence can be harmlessly taken for granted because the explanation the proponents of the recognitional reply to the Knowledge Argument give for the epistemic progress in question rests essentially on their attributing to the subject the acquisition of those concepts.[11]

physicalism is the thesis that every fact is a physical fact and propositional knowledge is knowledge of facts. (It is widely assumed that if the bit of knowledge that Marianna acquires through seeing red is, for instance, merely a piece of knowing-how, then it is not problematic for physicalism (see Lewis (2004)) and hence it can be plausibly ignored for the purposes of the Knowledge Argument. So, showing that there is a certain epistemic stage that goes unnoticed in Mary's case is not enough: one needs also to show that that epistemic stage is also propositional in character.)

[9] See Lycan (1996: 93) for a proposal along the similar lines. I adopt (P) instead of, e.g., (R) Marinna knows that this experience is Q or (S) Marianna knows that Q is what it is like to have this experience or some other proposition, because (P) is what one gets if one follows the general recipe of transforming "know wh-…" constructions into "know that" constructions (see below). However, I will sometimes make use also of (R) or (S) when it is more convenient to do so. But nothing essential in my discussion hangs on this choice.

[10] Chalmers (2010: 267–8) calls the concept *Q* a direct phenomenal concept (and the belief that it is like Q to have this experience a direct phenomenal belief). On Chalmers' account, the concept *this* that Mary employs to pick out the red experience-type is a demonstrative concept under which that experience-type is conceived *as* the object of her demonstration. Since, on this account, conceiving an experience-type under a demonstrative concept is not thereby conceiving it as having a quality picked out by a direct phenomenal concept (in our case, *Q*), (P) turns out to be a substantive piece of knowledge that is neither a priori nor analytic. This section of the present paper is much indebted to Chalmers' account.

[11] Chalmers (2002) makes a useful distinction between Type-A materialists and Type-B materialists. Type-B materialists argue, while Type-A materialists deny, that there is an epistemic gap between physical/functional truths and phenomenal truths (and, of course, they both deny that there is a corresponding ontological

How does Marianna come to entertain, through experiencing red, the proposition that it is like Q to have *this*, where *this* picks out the red experience-type? A natural suggestion is this. Marianna has a red experience when she is presented with the slide. She then attends to it and forms the phenomenal concept *Q* that picks out the red experience-type.[12] And, by predicating the concept *Q* of the very experience she has, she forms the belief that it is like Q to have *this*, demonstrating the red experience-type in question.

Thinking Marianna's new piece of knowledge in terms of (P) is supported by certain syntactic and semantic relations between the sentential constructions "S knows wh-…" and "S knows that…" For instance, "Jack knows who Mary is" is true in virtue of Jack's knowing that Mary is so-and-so, and "Susan knows where Hector is" is true in virtue of Susan's knowing that Hector is in such-and-such place. A plausible idea is that Marianna's knowledge can be viewed as a special case which can be assimilated to this general scheme: Marianna knows what it is like to see red in virtue of knowing that it is like Q to have *this*, demonstrating a red experience.[13]

I would like to make two points regarding the nature of the new piece of knowledge captured by (P). First, the new knowledge in question is not to be confused with the stipulative sort of knowledge Marianna might acquire upon getting acquainted with the phenomenal quality instantiated by experiencing red. Marianna might simply call the phenomenal quality she experiences 'Q' and acquire the piece of knowledge expressed by the sentence "I name this quality 'Q'." It is clear that

gap). Type-B materialists account for the existence of the alleged epistemic gap by reference to the special epistemic and semantic properties of phenomenal concepts. This is what is also known as "the Phenomenal Concept Strategy" (PCS) in the literature. The proponents of the recognitional reply to the Knowledge Argument are Type-B materialists in Chalmers' sense. Perry (2001) explicitly argues that the recognitional progress Mary makes through having red experiences is partly a matter of acquiring a new phenomenal concept of having red experiences. See also Tye (2000). I will specify how what I have to say about RS bears on PCS in general in the final section of the paper.

[12] Surely concept formation is a much more complex process than merely attending to the qualities in the subject's view. However such complexities need not concern us in this paper and the very rough account sketched above will suffice for our purposes.

[13] Rosenthal writes: "Knowing 'wh' abstracts from the full content of one's knowledge; one knows *what* something is only if one knows *that* it's an F" (2004: 193). It is also worth noting that the locution "it is like" in (P) does not mean "it resembles" just as the locution "it is like" does not mean "it resembles" in the context of "what it is like to see red." Nagel writes: "[The] analogical form of the English expression "what it is *like*" is misleading. It does not mean "what (in our experience) it *resembles*," but rather 'how it is for the subject himself.'" (1974: 440, fn. 6). Reading "what it is like" as "what it resembles" is what Lewis (2004) calls "the first way to miss the point" of the Knowledge Argument among the six ways it specifies. The sense in which Marianna does not know what it is like to see red before she is presented with the red slide is the same as the sense in which she does not know, before she is presented with the red slide, that it is like *Q* to have this, demonstrating the experience-type in question.

the knowledge thus expressed does not pose any threat to physicalism. However, the new piece of knowledge expressed by the that-clause in (P) does not concern a "merely linguistic" fact but a worldly fact about her experience of seeing red. That piece of knowledge is acquired through predicating the novel concept *Q* of her experience and it is not relevantly different from the pieces of knowledge Marianna can acquire through predicating that concept to her subsequent experiences of seeing red. The fact that the token of the experience-type of seeing red through the having of which Marianna acquires the concept *Q* is the same as the token experience of which she predicates that concept should not obscure the point that the new knowledge she acquires does not concern a fact about her decision regarding the use of language but a fact about her experience.

Second, in order for (P) to pose a challenge to RS, it must be shown that the piece of knowledge that is expressed by the that-clause it contains is not knowledge of an identity (Recall that RS claims that the novel piece of knowledge acquired through seeing red is knowledge of an identity and as such does not threaten physicalism). In the next section, I will argue that the piece of knowledge captured by (P) is *not* knowledge of an identity. For the moment, however, I wish to argue that Chalmers' construal of Perry's account as an attempt "to analyze phenomenal knowledge as a sort of indexical knowledge" (2004: 184) obscures much of the point of RS. Chalmers argues that Perry thinks that the only sort of concept that can be acquired through experiencing is a demonstrative concept that functions to pick out whatever sort of experience one is currently attending to. According to Chalmers, there are also non-demonstrative "qualitative concepts of experiences" (2004: 185), which can be acquired through experiencing and are involved by the crucial new knowledge that the knowledge argument turns on. By a qualitative concept of an experience, Chalmers means the same sort of concept as *Q* involved by (P).[14] Chalmers argues that Perry's (or more generally RS's) failure to recognize qualitative concepts such as *Q* results in his (or its) exclusive focus on "the relatively uninteresting indexical knowledge" (2004: 185) that *the experience usually caused by red things is this*, where *this* is a demonstrative concept of red experiences. For Chalmers, "the substantive, non-trivial" (2004: 185) knowledge that *the experience usually caused by red things is Q* (Chalmers' candidate for the crucial new knowledge central to the knowledge argument) is simply neglected by Perry and RS.

There are a couple of points I would like to make on behalf of RS. First, RS need not and does not deny that there are qualitative concepts of experiences that can be acquired through experiencing. In his response to Chalmers, Perry emphatically puts it that Chalmers' demonstrative concepts are "not my [his] candidate for Mary's [new] concept" (2004: 219) and that "in thinking of the experience in this new way,

[14] See fn. 10 above.

she is not thinking of it as 'this experience'" (2004: 221). Accordingly, RS need not and does not deny that there are substantive (cognitively significant) pieces of knowledge that can be acquired through experiencing. According to RS, just as one's coming to know, after Dretske introduces himself, that *that is Dretske*, involves a substantive piece of knowledge, what one might acquire through experiencing might similarly involve a substantive piece of knowledge. This is because new knowledge can be substantive without necessarily involving knowledge of new facts. Second, the central point of RS is that the crucial substantive piece of knowledge acquired through experiencing, the new knowledge involving qualitative concepts, is *knowledge of an identity* and as such does not pose any threat to physicalism. No purported objection to RS that does not explicitly counter this very point gets off the ground.

An interesting result appears to follow if (P) captures what Marianna comes to know after having red experiences. There are indeed two knowledge arguments, rather than one, which can be distinguished with reference to the two items of knowledge that someone situated in a monochromatic environment from birth appears to lack. First, such a person lacks the piece of knowledge expressed by the that-clause in (P): the knowledge that it is like Q to have this, where this picks out a red experience-type. The first Knowledge Argument runs roughly like this: One who knows everything physical there is to know may still lack the knowledge that it is like Q to have this, where this picks out a red experience-type, and hence physicalism is false. Second, she also lacks the knowledge that Q is what it is like to see red. The second Knowledge Argument goes roughly like this: One who knows everything physical there is to know may still lack the knowledge that Q is what it is like to see red, and hence physicalism is false.

Having the second piece of knowledge mentioned above implies having the first but not *vice versa*. Just like the first piece of knowledge, the second piece of knowledge intuitively requires having an experience of seeing red. But there is more to the second item of knowledge than what is required to have the first. Through her experience of the red slide, Marianna knows that it is like Q to have this, where *this* picks out the red experience-type, but does not know that Q is what it is like to see red; while through her experience of ripe tomatoes, Mary knows both. The difference between the two stems from the fact that unlike Mary, Marianna is not in a position to recognize her red experience-token *as* an instance of the red experience-type because such a recognition requires either having experiences of paradigmatically red objects like ripe tomatoes or being told by others that pointing at the red slide, *that* is red, or something to that effect.

A suggestive model which explains how Mary happens to recognize her experience of red as an experience of red goes like this: when Mary sees ripe tomatoes, she learns that it is like Q to have this, where *this* picks out the red experience-type. What is common to Mary and Marianna's cases is this stage of epistemic progress. However, Mary also

knows that what she sees are ripe tomatoes (after all, she can tell by the way they look to her), and combined with her knowledge that her perceptual apparatus works normally and ripe tomatoes induce red experiences in normal perceivers, she infers that *this* is an experience of red.[15] Hence, she gains an item of knowledge which Marianna lacks.

The fact that the item of knowledge that Mary and Marianna both acquire upon experiencing red is propositional is important because that means that that item of knowledge, as I have argued, can be properly deployed in a knowledge argument against physicalism. Hence if RS fails to account for that item of knowledge then it cannot be considered as a tenable physicalist reply to that argument. In the following section, I will argue that the first epistemic stage that Mary and Marianna both go through is not, or does not consist in, recognition of an identity and hence RS cannot account for the item of knowledge acquired at that stage.

4. *The Nature of the Two Epistemic Stages*

Before addressing the question of whether the first epistemic stage consists in recognition of an identity, the question I want to raise is this: what exactly is it that Mary knows in addition to what she and Marianna both know about the experience of red? How substantive is the *extra* epistemic progress Mary makes, that is, does she come to know a new fact about the world Marianna does not know? I think these are the questions RS can be properly interpreted as addressing. Let us consider the following more closely:

(1) Q is what it is like to have this experience.
(2) Q_R is what it is like to see red.

(1) is uttered by Marianna when she is shown the red slide. And, suppose that she also knows (2) when she is in her room: she knows that there is something it is like to see red, and her textbooks call it 'Q_R'.[16] What she does not know is:

[15] One may argue that Mary's knowledge is non-inferential because she does not go through any conscious inferential process in her mind. I am inclined to reply that there are unconscious inferences as well as conscious ones and Mary's inference can be the former if not the latter. Nothing much hangs on this, however. What is important is that there are different stages of epistemic progress in the cases specified and we need an account which explains how one passes from one stage to another.

[16] One may question whether we can reasonably build the assumption that Mary learns about the subjective character of seeing red, Q_R, into our formulation of the knowledge argument. Rosenthal writes: "Unless we've established independently that Q_R is itself physical, Mary's learning about it may well be learning about something physical…Only if we've shown that Q_R is physical can Mary's textbooks teach her about it. It's question begging to build that assumption into our formulation of the knowledge argument" (2004: 195–6). The anti-materialist requires, Rosenthal argues, that Mary's textbook knowledge be exclusively physical, and if Q_R is non-physical, then Mary learns about something non-physical in her room and this

(3) Q is Q_R.[17]

On the other hand, Mary knows (3) in virtue of knowing that this (the experience type demonstrated by her and Marianna) is a red experience. That is, *what she knows but Marianna does not know is the following*:

(4) This is a red experience.

RS argues for two distinct claims: first, (4) is a piece of recognitional knowledge of an identity; second, failing to recognize identities does not have ontological costs: just as Perry's failing to know, demonstrating Dretske that *that* is Dretske is merely a failure to recognize an identity and thus does not have any ontological implications, Marianna's failing to know (4) is failing to recognize an identity and thus does not have any ontological implications. RS holds that Marianna already knows a fact identical to the fact expressed by (4) in some other way (for instance, she knows that a red experience is a red experience) just as Perry knows a fact identical to the fact that that [demonstrating Dretske] is Dretske in some other way (for instance, he knows that Dretske is Dretske).[18]

obscures any new non-physical knowledge she might get on first consciously seeing red. I think this is a reasonable worry about Perry's formulation of the knowledge argument, and if, as Rosenthal suggests, Mary cannot learn about Q_R in her room, then Perry's thesis that Mary's new knowledge is knowledge of an identity does not even get off the ground. I grant in this paper that Mary can learn about Q_R in her room, and I will show that granting even this much does not save RS. See also Perry's (2004) response to Rosenthal.

[17] It is worth making a point that supplements my discussion in the previous section of Chalmers' influential interpretation of RS. (3) captures what Mary comes to know after seeing paradigmatically red objects. Indeed, the content of (3) is one of Perry's candidates for the content of Mary's crucial new knowledge (see Perry (2001, chap. 7: 145–50)). If this is so, then RS cannot viewed as an attempt to assimilate phenomenal knowledge to indexical knowledge given that there are no demonstrative concepts involved in (3).

[18] One may wonder how Perry accounts for the cognitive significance of identities such as (3) and (4). Perry argues that in order to appreciate the cognitive significance of identities, one must reject "the subject matter assumption," according to which "the rational content of a belief is the conditions its truth puts on the subject matter of the belief, the objects the notions and concepts in the belief are of" (113–4). However, rejecting the subject matter assumption is, Perry argues, not rejecting the notion of content. In fact, for Perry, we need not jettison content but *discover more of it*, i.e., we need different kinds of truth-conditions (which he calls "reflexive truth-conditions") as a part of the rational content of a belief in order to account for the cognitive significance of beliefs about identities. Reflexive truth-conditions are, as Perry defines them, not merely conditions on the subject matter but "conditions on the utterances or thoughts *themselves*" (21). By appealing to reflexive contents, Perry tells us, we can capture differences in contents of beliefs that are not captured by holding the subject matter assumption. The merits of Perry's "reflexive-referential account of content" for the cognitive significance of identities need not be assessed in this paper because the piece of knowledge that is really problematic for physicalism is, as I will argue, not of an identity.

In order to argue against RS, one need not raise objections against the account it provides for the epistemic progress Mary makes when she gets the information in (4). This is because an attempt to understand Mary's *entire* epistemic progress in terms of acquiring a piece of recognitional knowledge of an identity rests essentially on a failure to distinguish different pieces of knowledge Mary gains when she sees ripe tomatoes: the one that she shares with Marianna and the other one she has but Marianna lacks. The latter might be a bit of recognitional knowledge of an identity and hence be devoid of ontological implications; but, as I will now argue, the former is not recognition of an identity and hence that RS cannot account for the first epistemic stage.[19]

What are the conditions under which a piece of recognitional knowledge of an identity is acquired? A natural suggestion is this. A condition for a given subject to acquire a piece of recognitional knowledge of an identity is that she has (at least) two concepts of the same thing. The acquisition of recognitional knowledge of an identity occurs when the subject recognizes that there is only one thing, rather than two, her concepts are about. Perry acquires a piece of recognitional knowledge of an identity upon his recognition that *this* (person) and Dretske are not two different things but one and the same thing. This is also what happens to Mary when, at the second epistemic stage, she recognizes that Q and Q_R are not two different things but one and the same thing.

Now does anything like Mary's recognition that there is only one thing, rather than two, her concepts are about occur at the first epistemic stage? I think the answer is definitely "No." Nothing like recognizing an identity occurs at the first epistemic stage because, at that stage, Mary does *not* recognize that Q is Q_R, and recognizing that Q is Q_R (what else?) is what Mary has to do if she is to be conceived as acquiring a piece of recognitional knowledge of an identity.

One may reasonably raise the question of what sort of information Mary acquires at the first epistemic stage if it is not a piece of recognitional knowledge of an identity. I think the answer to this question is already implicit in the account given in the previous section. What Mary learns at the first epistemic stage is that the experience-type that is referred to by her demonstrative concept *this* has a certain quality, one which is referred to by her novel concept *Q*. There is no recognition of an identity here but recognition that the experience-type that is con-

[19] Stalnaker also holds that there are indeed two different epistemic stages Mary goes through upon seeing a ripe tomato and raises doubts as to whether RS is properly applicable to the first epistemic stage. He writes: "It is at stage one that the problematic cognitive achievement—the learning "what it is like" to see red – takes place. But it is at stage two…that Mary receives information that *this* color is red, the information that is analogous to the information that Perry received, that *this* person is Dretske. So even if the analogy could help to explain what is learned at stage two, it is not clear that this would be relevant to the original puzzle." (2008: 44) However, Stalnaker does not go far enough to make a definite claim that RS can*not* account for the first epistemic stage and hence leaves unanswered the question of *why* it cannot account for that stage.

ceived by Mary *as* the object of her demonstration has a certain quality, which is non-demonstratively conceived as *Q*. The piece of knowledge Mary shares with Marianna is recognitional because the concept *Q* is a recognitional concept, which she deploys to introspectively identify her red experiences, and also because correct applications of recognitional concepts yield recognitional knowledge. However, not all recognitions are recognitions of identities: some are recognitions of property-instantiations. And it is the latter type of recognition under which Mary's progress at the first epistemic stage falls.

Typically, perceptual experiences have many properties. A given perceptual experience might have, for instance, the property of being caused by red things under such and such circumstances, the property of being the favorite experience of most people located in such and such part of the world, the property of having such and such beliefs as effects, and so on. In addition to all these relational properties, the perceptual experience in question might have a specific subjective character, a qualitative feature in virtue of which there is something it is like to undergo that experience. The phenomenal concept *Q* involved in (P) is a concept of the qualitative feature of the experience of seeing red, a novel concept that Mary acquires through having that experience. Mary gains a special sort of access to the property picked out by *Q* in virtue of having the experience of seeing red, which is what explains its novelty.[20] Of course, a commitment to *special* access does not entail, without further argument, a commitment to *unique* access. More specifically, it does not entail without further argument that the property picked out by *Q* cannot be picked out by the concepts Mary acquires in her room. Indeed, the concept Q_R that Perry stipulates is had by Mary while she is still in her room is designed to pick out whatever *Q* picks out. However, as I have argued above, no recognition of the identity that Q is Q_R occurs at the first stage of Mary's entire epistemic progress, and no other identity seems relevant. The first epistemic stage is rather to be characterized by Mary's predicating the novel concept *Q*, which is of the qualitative feature of the experience of seeing red, of the experience she undergoes. If this is so, Mary's first-stage knowledge does not have the form 'x is y' but the form 'x is an F'.

A question in the vicinity that calls for an answer is this: how does my objection to RS bear on the prospects of the Phenomenal Concept Strategy (PCS) in general?[21] As I see it, RS can be plausibly considered as a version of PCS since it subscribes to the central claim of PCS that there are some special, phenomenal concepts of experiences (or their subjective characters) the acquisition of which is not guaranteed by the acquisition of physical concepts. However, RS also makes the further

[20] Perry writes: "[T]here is a way of attending to a subjective character that is possible only when one is having an experience of which it is the subjective character" (2001: 145).

[21] See fn. 11.

claim that the knowledge that characterizes Mary's entire epistemic progress is knowledge of an identity, something like 'Q is Q_R'. This is what distinguishes RS from other (and perhaps more popular) versions of PCS. This further claim about the form of the crucial new knowledge Mary acquires through experiencing is not essential to PCS. A non-RS version of PCS can consistently claim that the knowledge in question is of the form 'x is an F' while holding that concepts Q and Q_R pick out the same property. That is, such a version might hold that it is true that Q is Q_R without claiming that knowledge of this truth characterizes Mary's crucial new knowledge (or her entire epistemic progress). The objection that I develop against RS in this paper is not, and is not intended as, an objection to a non-RS version of PCS conceived along those lines.[22]

To sum up the upshot of the paper: once the two epistemic stages one might undergo through experiencing red are clearly distinguished, it is easy to see that Mary's entire epistemic progress cannot be understood merely as coming to recognize an identity. Since the first stage of the entire epistemic progress in question involves a piece of propositional knowledge that is not knowledge of an identity, RS fails as a response to the knowledge argument.[23]

References

Chalmers, D. 2002. "Consciousness and its place in nature". In D. Chalmers (Ed.). *Philosophy of Mind*. New York: Oxford University Press: 247–272.

Chalmers, D. 2004. "Phenomenal concepts and the knowledge argument". In P. Ludlow, Y. Nagasawa, and D. Stoljar (Eds.). *There is something about Mary*. Cambridge: MIT Press: 269–298.

Chalmers, D. 2010. *The character of consciousness*. Oxford: Oxford University Press.

Demircioglu, E. 2012. "Physicalism and phenomenal concepts". *Philosophical Studies* 165 (1): 257–277.

Frege, G. 1993. "On sense and reference". Originally published 1892. In A. W. Moore (Ed.). *Meaning and Reference*. Oxford: Oxford University Press: 23–43.

Jackson, F. 1982. "Epiphenomenal qualia". *Philosophical Quarterly* 32: 127–136.

Levin, J. 2007. "What is a phenomenal concept?" In T. Alter and S. Walter (Eds.). *Phenomenal concepts and phenomenal knowledge*. New York: Oxford University Press: 145–167.

Levine, J. 2010. "Demonstrative thought". *Mind and Language* 25 (2): 169–195.

[22] For an attack against non-RS versions of PCS, see Demircioglu (2012).

[23] I would like to thank İlhan İnan, Steven Voss and two anonymous reviewers of *Croatian Journal of Philosophy* for their comments on the earlier version of this paper.

Lewis, D. 2004. "What experience teaches". Originally published 1988. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.). *There is something about Mary*. Cambridge: MIT Press.

Loar, B. 2004. "Phenomenal states". Originally published 1990/7. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.). *There is something about Mary*. Cambridge: MIT Press: 219–239.

Lycan, W. 1996. *Consciousness and experience*. Cambridge: MIT Press.

Nagel, T. 1974. "What is it like to be a bat?" *Philosophical Review* 83 (4): 435–450.

Nida-Rümelin, M. 2004. "What Mary couldn't know". Originally published 1995. In P. Ludlow, Y. Nagasawa, & D. Stoljar (Eds.). *There is something about Mary*. Cambridge: MIT Press: 219–241.

Peacocke, C. 1983. *Sense and content*. Oxford: Oxford University Press.

Perry, J. 2001. *Knowledge, possibility, and consciousness*. Cambridge: MIT Press.

Rosenthal, D. 2004. "Subjective character and reflexive content". *Philosophy and Phenomenological Research*, 68 (1): 191–198.

Stalnaker, R. 2008. *Our Knowledge of the Internal World*. Oxford University Press.

Tye, M. 2000. *Color, consciousness and content*. Cambridge: MIT Press.

Croatian Journal of Philosophy
Vol. XV, No. 45, 2015

On Aristotelian Universals and Individuals: The "Vink" that is in Body and May Be In Me

IRENA CRONIN
University of California, Los Angeles, USA

G. E. L. Owen, in his influential paper "Inherence," talks of "vink," a name he has created for a particular shade of the color pink, and this "vink" serves as an individual in the Aristotelian category of quality. Owen was one of the first to aim to discredit the belief that J. L. Ackrill and his camp espoused, the belief that Aristotle thought that "general attributes are not in individuals, particular attributes are not in more than one individual." I postulate that there is nothing here that does not preclude the existence of transferable nonsubstantial particulars, and base this view on passages from Aristotle's Categories and certain examples found in Ammonius's commentary and On Colors. Given this, a nonsubstantial particular of "vink" would not have to rely on having inhered in just one particular body to have existence, however, it would have to inhere in at least one particular body.

Keywords: Universals, particulars, Owen, Ackrill, Aristotle.

G. E. L. Owen, in his influential paper "Inherence" (Owen 1965: 97–105), talks of "vink," a name he has created for a particular shade of the color pink, and this "vink" serves as an individual in the Aristotelian category of quality.[1] Owen was one of the first to aim to discredit the belief that J. L. Ackrill and his camp espoused, the belief that Aristotle thought that "general attributes are not in individuals, particular attributes are not in more than one individual" (Owen 1965: 100).

[1] "…analogous to Socrates in the category of substances" (Owen 1965: 98). Owen replaces the word Aristotle used, *leukon*, which represents all light colors, with the word *pink*, for no single English equivalent exists for *leukon*. "*Leukon* covers all light colors as *melan* covers all dark colors: that is why the commonplace that all colors range between or are composed of *leukon* and *melan* (*Cat.* 12a17–19, *Phys.* 188b3–5, *DA.* 442a12–13) is sense…" (Owen 1965: 98).

Owen claims that, given Aristotle's schema, something must contain an individual, such as "vink," if the individual is to exist at all (Owen 1965: 105). Here "vink" represents a fully determinate universal-type color,[2] a repeatable entity that could be shared among more than one particular body. It is included as an individual by Owen because he viewed it as not being "said of" anything else. Michael Frede has offered his interpretation by saying that it is a sufficient condition that an individual, such as "vink," be found in body in general (Frede 1987: 60–61). Given both Owen's and Frede's interpretations, it could be inferred that a color, such as "vink", would inhere in particular bodies as a universal-type entity, due to the view that it is able to manifest in a number of particular bodies.

Owen's and Frede's interpretations are both different than that of J. L. Ackrill's, Michael Wedin's (Wedin 1993: 163–164) and others' which view an individual as a trope-like, that is non-recurrent, nonsubstantial particular. However, Owen, Frede and Wedin all agree on the interpretation that a quality inhering in a particular body does not entail that the quality could not exist without that particular body.

So, a major issue is how, according to Aristotle, a quality is individuated in a particular body; for the color "vink," this calls into question how it is manifested in a particular body, such as me. The question of whether or not a particular "vink" is particular insofar that it belongs to a particular body goes to the heart of the matter; if the particular body stopped existing, would that particular "vink" as well? I believe that this would not have to be the case. A relevant passage here is Aristotle's *Categories* 1a24–25:

> By "in a subject" I mean what is in something, not as a part, and cannot exist separately from what it is in.[3]

Interpreting this passage in the way it appears Frede does produces the following:

> *x* is an accident $=_{df}$ there is something, *y*, such that *x* is in *y*, *x* is not in *y* as a part, and *x* cannot exist separately from *y* (Matthews 1989: 96).

I find that there is nothing in this interpretation that does not preclude the existence of transferable nonsubstantial particulars. Given that this is the case, a nonsubstantial particular of "vink" would not have to rely on having inhered in just one particular body to have existence, however, it would have to inhere in at least one particular body. A possible example of this would be when color from a particular body

[2] Aristotle is not specific about whether the general nature of color is that of a universal, but it can be inferred, as Owen has, given Aristotle's general schema of universal to particular.

[3] There are not too many places in which Aristotle makes a "distinction between *being in a subject* and *being said, or predicated, of a subject*" (*Cat.* 1a20–b9, 2a11–14, 2a27–b6, 2b15–17, 3a7–32, 9b22–24; *Postpred.* 11b38–12a17, 14a16–18; *Top.* 127b1–4). It is typically within these bounds that the whole theory of what it means to be "in a subject" is played out in.

is transferred to another, such as is what happens with color dyeing; a similar example comes from Ammonius's commentary on the *Categories*. More on these examples will follow. In my next section, I produce the traditional view.

I

Ackrill believed that Aristotle's notion of color in body was "compressed and careless" (Aristotle [Ackrill's notes] 1963: 83) when Aristotle stated that:

> All the other things are either said of the primary substances as subjects or in them as subjects. This is clear from an examination of cases. For example, animal is predicated of man and therefore also of the individual man; for were it predicated of none of the individual men it would not be predicated of man at all. Again, color is in body and therefore also in individual body; were it not in some individual body it would not be in body at all. Thus all the other things are either said of the primary substances as subjects or in them as subjects. So if the primary substances did not exist it would be impossible for any of the other things to exist.[4]

Ackrill read this passage from the *Categories* as stating that color as a universal would be found in a particular body. However, he believed that Aristotle did not really mean what he wrote there, that is that he had mistakenly written it. He further interpreted Aristotle as having meant to say that a nonrecurrent "instance" of color would be in and dependent on a particular body and that universal color would be in and dependent on universal body. Thus, universal color would not be in a particular body for it did not depend on it (Aristotle [Ackrill's notes] 1963: 83).

So, according to Ackrill's view, the universal-type "vink" would be found in a universal body and could not exist apart from it, and similarly, an instance of the color "vink" would be found in a particular body and could not exist apart from that particular body. On this definition, each instance of "vink" would be uniquely associated with a particular body and transfer of that certain instance to another particular body would not be possible, that is, that certain instance was viewed as inseparable from the particular body it inhered in. An instance of "vink" that had inhered in my sweater thus could not be transferred my dress.

Although I agree with Ackrill's reasoning that universal color would not be found in a particular body, I am not convinced that Aristotle had Ackrill's idea of instance in mind.

II

Matthews, in his characterization of Frede's interpretation of Aristotle's "in a subject" condition, includes a scenario where a person may

[4] *Cat.* 2a34–2b6.

have a grandmother in their class who is not their grandmother, but rather the grandmother of another person. In this way it is demonstrated that not every subject that an accident inheres in is a subject that it could not exist without (Matthews 1998: 96).

I agree with Matthew's characterization of Frede, and I agree with both Frede and Matthews, as well as Owen, that universal color would still exist even if a particular body, which color had inhered in, had ceased to exist. However, I disagree with the notion that color as a universal inheres in a particular body. Rather, I postulate a view here that universal color inheres in universal body, and particular color inheres in particular body as a nonsubstantial particular, in that this particular color would not have to rely on just one particular body to have existence, that is, it could change subjects.

An example that I present in this section is found in a passage in a work entitled *On Colors*.[5] This work had traditionally been attributed to Aristotle but is now often ascribed to Theophrastus (c. 371—c. 287 BCE), Aristotle's designated successor at the Lyceum (Kuehni and Schwarz 2008: 32), or Strato (c. 335—c. 269 BCE), the third director of the Lyceum after Theophrastus's death. Outside of the question of whether Aristotle was the actual writer, it is regarded by many to be Aristotelian in nature (Desclos and Fortenbaugh 2011: 307, see also Edel 1982: 157 and Thomson and Missner 2000: 7). Even so, there are still some issues that need to be addressed when applying this example in support of my view that a nonsubstantial particular may be capable of transferring from one particular body to another.

The example is as follows—with regards to the entire process of dyeing, a particular object is moistened or heated and some of the color of that object is then transferred to another particular object. For a time then, no matter how brief, the color is transported by a liquid or by heat to a new particular object. The original particular object would continue to exist, with the same type of color or a less intense hue of the type of color that had transferred, dependent on the particular object's substance and pore-structure.[6] The color that has now transferred to another particular object may blend into the original color of this object due to this particular object's

[5] *On Colors*, 794a16–b10. A part of the pertinent passage from *On Colors* is:

All dyed things take their color from the dye. Common sources of such coloration are the flowers of plants and their roots, bark, wood, leaves, or fruit, and again, earth, foam, and inks. Sometimes coloration is due to animal juices (e.g., the juice of the purple fish, with which clothes are dyed purple), in other cases to wine or smoke, or lye mixture, or to sea-water, as happens, for instance, to the hair of marine animals, which is always turned red by the sea. In short, anything that has a color of its own may transfer that color to other things, and the process is always this, that color leaving one object passes with moisture and heat into the pores of another, which on drying takes the hue of the object which the color came.

[6] This is suggested by *On Colors*, 794a16–b10.

substance and pore structure, so it is possible that after dyeing, this object may end up with a color that is not the exact color that was transferred.[7]

Notwithstanding, however, it is clear that the color that came from one particular object had transferred into another particular object. The question is, what is the ontological picture behind this occurrence? Has some of the original particular object, the host of the color, been transferred with its color to another particular object? Another example from a later Greek commentary on the *Categories* may be useful in thinking this through.

The example of dyeing found in the passage from *On Colors* happens to be similar to an example that Ammonius (c. 435/445—c. 517/526 CE) had presented in his commentary on the *Categories* (Matthews 1989: 91–104). Ammonius's example had to do with the fragrance of an apple and how it appeared to exist separately from the apple. However, according to Aristotle, that which is in a subject could not exist separately from what it is in. One of Ammonius's solutions to this puzzle, which is a repetition of a solution that had previously been put forth by Porphyry (232—309 CE), is that some fragrance of the apple had transferred from the apple into the air. This solution, called the "tense solution" by John Ellis, "allows particular accidents to migrate to other subjects" and maintains "that a particular accident must always be in *some* subject" (Ellis 1990: 291–292). The tense solution, as explained by Porphyry and echoed by Ammonius, is:

> For neither did he [Aristotle] say, "cannot exist separately from that in which it was," but "cannot exist separately from that in which it *is*." For the fragrance can be separated from that in which it is, though it is impossible [for it] to exist separately on its own, but it either perishes or is transferred to another subject. For that it is inconceivable for an accident to exist separately on its own, this he indicated, but not that it cannot be separated (Ellis 1990: 291–292).

Ammonius also adds to this that it is perhaps our sense of smell that picks out the fragrance of the apple.

Matthews places more stock in the tense solution than another that Ammonius seems to favor (although this may be disputed) (Matthews 1989: 100; Ellis 1990: 291–302),[8] which Ellis calls the "effluence solution" (Ellis 1990: 293). The effluence solution has the fragrance first

[7] *On Colors*, 794a16–b10.

[8] Ammonius never states that he favors the effluence solution, he just wrote more lines about this solution than the tense solution. It may be that he felt he did not have to write as much on the tense solution since it was an already established and well-known solution. That is, Ammonius may have been listing and explaining solutions rather than attempting to provide a point of view on which solution to favor. Later medieval solutions move away from the tense solution to more psychological solutions, but this may be due to the particular evolution of solutions to the fragrance-in-apple problem and changes in cultural and scientific thought, rather than efforts to provide a direct reflection of what Aristotle might have had in mind as a solution during the period in which he lived.

in the apple, then in bits of the apple that are carried along with fragrance into the air.

In this case, the apple, which is the host of the fragrance, somehow has a part of itself broken up into bits, with those bits carried into the air; this is postulated in order to preserve the view that an accident could not exist separately from its particular body. However, as Matthews keenly points out (Matthews 1989: 100), the effluence solution is not the relevant solution to ensure this is the case, for the fragrance could indeed exist separately from the particular body of the apple. Particular bits of apple separated from a particular apple are like particular toes separated from a particular man. Particular toes are not equivalent to a particular man, and similarly, particular bits of apple are not equivalent to a particular apple.

The tense solution allows for a transfer of a particular accident from one particular body to another particular body; this solution does not appear to allow for the simultaneous sharing of a particular accident among two or more particular bodies. The effluence solution, on my and Matthews's interpretation, allows for a transfer of a particular accident from one particular body (a particular apple) to a particular entity (that is, a particular bit from a particular apple; I am clear not to call this a particular body, for it would be hard to imagine a universal bit of apple as part of Aristotle's ontology). Even though the transfer of the particular fragrance in the effluence solution is not due to any action on the particular fragrance's part, rather it is the particular subject in which it inheres itself that appears to have actively changed, the particular fragrance has nonetheless been transferred. In addition to the tense solution, the effluence solution also does not appear to allow for the simultaneous sharing of a particular accident in multiple particular bodies.

Both the tense solution and the effluence solution may be seen to apply to the "transfer of color by dyeing" example. In the version utilizing the tense solution, the nonsubstantial particular shade of color that is transferred from one particular body to another particular body is first transferred via some particular moisture or particular heat in which the particular color inheres. Then, this particular color is transferred once more to another particular body, the particular body that will be dyed by that particular color.

With the effluence solution, the nonsubstantial particular shade of color which is inhering in the original particular body then inheres in bits of the original particular body that have been separated from that body, which in turn are situated in some particular moisture or particular heat. After this, the particular color inhering in particular bits of the original particular body is transferred to another particular body. Any issues of dye fading or running out of this particular body after the dyeing process are not relevant, for in the effluence solution case, it is the transfer of the nonsubstantial particular shade of color between the

original particular body and the bits of the original particular body that are important. However, as in the "fragrance-in-apple" example, an issue that may never be resolved is exactly what kind of entity would Aristotle have thought a bit of apple was. Although, it is clear to me and others such as Matthews that a particular bit of apple is not equivalent to the particular original whole apple it had separated from.

III

Man, the universal, is said of a particular man. On my view, a fully determinate, universal-type color, such as "vink", is said of a nonsubstantial particular of "vink". This nonsubstantial particular of "vink" inheres in a particular body and may be transferred. I believe that color in general, moreover, is said of any of the determinate, universal-type colors that range in shades between *leukon*, all light colors, and *melan*, all dark colors.

To summarize, I interpret Aristotle to say that without some particular body or other, there would be no universal body, no color in general, and no color in particular. Furthermore, this color in particular is a nonsubstantial particular that is capable of being transferred, and only a nonsubstantial particular can inhere in some particular body.

In contrast, Wedin interprets Aristotle's schema as allowing nonsubstantial universals to inhere in particular bodies, while also allowing for "the nonrecurrent status of nonsubstantial particulars" (Wedin 1993: 164). Examples provided by Wedin of these phenomena would have been helpful, for it is difficult to grasp the dynamics of how they could appear in nature, and thus appears far-fetched; as a result, I do not believe Wedin's structure is what Aristotle had in mind. In another opposing view, Cresswell theorizes that Aristotle believed that all whiteness was only individual whiteness, with no whiteness in general existing (Cresswell 1975: 244–245); this view does not seem to fit with Aristotle's notion of a general ontology.

Conclusion

I have postulated that Aristotle, according to the *Categories* and bolstered by certain examples found in Ammonius's commentary and *On Colors*, found that universal color inhered in universal body, and that colors as nonsubstantial particulars inhered in particular bodies, though not necessarily inhering in any one particular body, and furthermore, that colors as nonsubstantial particulars may have the capability of transferring from one particular body to another particular body. So, on my interpretation, the universal-type "vink" would be found in universal body, and a nonsubstantial particular "vink" would be found in some particular body or other and may have the capability to transfer from one particular body to another.

References

Ammonius. 1991. *On Aristotle's* Categories. Gareth B. Matthews and S. Marc Cohen, (trans. and comm.). Ithaca: Cornell University Press.

Aristotle. 1963. *Categories* and *De Interpretatione*. J. L. Ackrill (trans. with notes). Oxford: Oxford University Press.

_____. 1984. *On Colours*. In *The Complete Works of Aristotle*, Vol. I, Jonathan Barnes (ed.). Princeton: Princeton University Press: 1222–23.

_____. 1984. *Topics*. In *The Complete Works of Aristotle*, Vol. I, Jonathan Barnes (ed.). Princeton: Princeton University Press.

Cresswell, M. J. 1975. "What is Aristotle's Theory of Universals?" *Australasian Journal of Philosophy* 53: 238–47.

Desclos, M.-L. and W. W. Fortenbaugh (eds.). 2011. *Strato of Lampsacus: Text, Translation, and Discussion, Rutgers University Studies in Classical Humanities,* Vol. XVI. New Brunswick: Aldine Transaction.

Edel, A. 1982. *Aristotle and his Philosophy*. Chapel Hill: University of North Carolina Press.

Ellis, J. 1990. "The Trouble with Fragrance." *Phronesis* 35: 290–302.

Frede, M. 1987. *Essays in Ancient Philosophy*. Oxford: Oxford University Press: 49–71.

Kuehni, R. G. and A. Schwarz. 2008. *Color Ordered: A Survey of Color Order Systems from Antiquity to the Present*. Oxford: Oxford University Press.

Matthews, G. B. 1989. "The Enigma of *Categories* 1a20ff and Why It Matters." *Apeiron: A Journal for Ancient Philosophy and Science* 22 (4): 91–104.

Owen, G. E. L. 1965. "Inherence." *Phronesis* 10: 97–105.

Thomson, G. and M. Missner. 2000. *On Aristotle*. Belmont, CA: Wadsworth.

Wedin, M. V. 1993. "Nonsubstantial Individuals." *Phronesis* 38 (2): 137–65.

Table of Contents of Vol. XV

Croatian Journal of Philosophy is published three times a year. It publishes original scientific papers in the field of philosophy.

Croatian Journal of Philosophy is rated INT2
(in the field of Philosophy) and INT2
(in the field of Philosophy of Science) in the
European Reference Index in the Humanities (ERIH).
It is indexed in *The Philosopher's Index*
and in *Arts & Humanities Citation Index* (*Web of Science*).

Payment may be made by bank transfer
SWIFT PBZGHR2X
IBAN HR4723400091100096268

Instructions for Contributors

All submissions should be sent to the e-mail: cjp@ifzg.hr. Submissions must be in English and formatted to be double-spaced with suitably wide margins, an A4 page size, and automatic page numbering.

Articles are normally no longer than 8,000 words of main text including bibliography. The Journal will consider longer papers, but, once these limits are exceeded, authors should bear in mind the editorial policy that the acceptance bar raises with increasing length.

Manuscripts should be compiled in the following order: cover page; title; abstract (not exceeding 200 words); keywords (3 to 6); main text; appendices (as appropriate); references.

All the authors of a paper should include their full names, affiliations, postal addresses, telephone and fax numbers and email addresses on the cover page of the manuscript. If a paper is co-written, one author should be identified as the Corresponding Author. The cover page must be submitted as a separate document. All submitted manuscripts must be prepared for blind review, with revealing acknowledgements and self-identifying references removed.

Sources are cited in the text by the author's last name, the publication date of the work cited, and a page number if needed, e.g. (Barber 2007: 324). Full details appear in the reference list in which the year of publication appears immediately after the author's name:

Barber, Alex. 2007. "Linguistic Structure and the Brain." *Croatian Journal of Philosophy* 21 (7): 317–341.

Williamson, Timothy. 2013. *Identity and Discrimination*. Oxford: Wiley-Blackwell.

ISSN 1333-1108
9 771333 110001