

## INHALT

Vorwort	vii
<i>Ansgar Beckermann</i> Die Perspektive des Richters	1
<i>Wolfgang Benkewitz</i> Intentionale Einstellungen, Bedeutung und Modalitäten	15
<i>Bernd Buldt</i> On Fixed Points, Diagonalization, and Self-Reference	47
<i>Ralf Busse</i> The Myth of the Given, the Problem of Induction, and Reference to Objects	65
<i>Christoph Fehige</i> One Simple Reason Why for Many Attitudes We Will Never Have a Satisfactory General Metric	95
<i>Wolfgang Freitag and Alexandra Zinke</i> Ranks for the Riddle? – Spohn Conditionalization and Goodman’s Paradox	107
<i>Gordian Haas</i> Epistemic Accessibility as a New Criterion of Significance	127
<i>Volker Halbach</i> Prolegomena zu einer jeden künftigen Definition von Wissen, die als Lösung des Gettierproblems wird auftreten können	147
<i>Franz Huber</i> Means–End Philosophy	173
<i>Andreas Kemmerling</i> Repräsentationale Zustände	199

<i>Manfred Kupffer</i> Two Theories of Supposition	221
<i>Hannes Leitgeb</i> A Bridge Principle for Ranking Functions and Subjective Probability	243
<i>Godehard Link</i> Humean Delusions	255
<i>Arthur Merin</i> Bedeutung und die Grenzen der Absichtlichkeit – Über Hutcheson, Goethe, Darwin, H. P. Grice und »Oops!«	285
<i>Thomas Müller</i> Sleeping Beauty in Branching Time	307
<i>Julian Nida-Rümelin</i> Rationale Kooperation: die strukturelle Interpretation	327
<i>Martine Nida-Rümelin</i> Active Animals and Human Freedom	339
<i>Hans Rott</i> »If Oswald had not killed Kennedy« – Spohn on Counterfactuals	379
<i>Holger Sturm</i> Auf der Suche nach einer natürlichen Theorie der Eigenschaften	401
<i>Thomas Ede Zimmermann</i> Painting and Opacity	425
Zu den Autorinnen und Autoren	453



Franz Huber

## MEANS–END PHILOSOPHY

### ABSTRACT

The aim of this paper is a constructive one: to point out one way of engaging with some philosophical problems that does not overly rely on intuitions. Section 1 follows Huber (2014) in illustrating the pervasive reliance on intuitions in contemporary analytic philosophy. It suggests that intuitions alone will not solve philosophical problems. Section 2 introduces means–end philosophy, which focuses on arguments instead of intuitions. Section 3 describes the relevant background for section 4, where means–end philosophy is illustrated by one particular example that is discussed in more detail in Huber (2016). It is shown that obeying a certain normative principle is a means to attaining a certain cognitive end. More specifically, it is shown that obeying a normative principle relating counterfactual conditionals and conditional beliefs, viz., the *royal rule*, is a necessary and sufficient means to attaining an end that relates true beliefs in purely nonmodal propositions and true beliefs in purely modal propositions.

### 1. INTUITIONS

Philosophers of language and philosophical logicians typically rely on intuitions when theorizing about counterfactual conditionals and the modality they express: counterfactuality. However, intuitions regarding counterfactual conditionals are notoriously shaky.

As an example consider the debate between Stalnaker (1968; 1981) and Lewis (1973) about the validity of the so-called law of conditional excluded middle. The latter says that for any two sentences  $\alpha$  and  $\gamma$ : if  $\alpha$  were true, then  $\gamma$  would be true; or if  $\alpha$  were true, then the negation of  $\gamma$ , i.e.,  $\neg\gamma$ , would be true. In symbols:

$$(\alpha \Box\rightarrow \gamma) \vee (\alpha \Box\rightarrow \neg\gamma).$$

According to Stalnaker (1968) this principle is logically valid. Lewis (1973, 77 ff) disagrees and brings the following alleged counterexample:

C It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian.

Stalnaker (1981, 91 ff) defends an analysis which says that both

C<sub>1</sub> If Bizet and Verdi had been compatriots, Bizet would have been Italian,

C<sub>2</sub> If Bizet and Verdi had been compatriots, Verdi would have been French,

»are indeterminate – neither true nor false. It seems to me that the latter conclusion is clearly the more natural one. I think most speakers would be as hesitant to deny as to affirm either of the conditionals, and it seems as clear that one cannot deny them both as it is that one cannot affirm them both. Lewis seems to agree that unreflective linguistic intuition favors this conclusion« (Stalnaker 1981, 92).

The reason for Stalnaker's last claim is that Lewis (1973, 80) says: »I want to say [C], and think it probably true . . . . But offhand, I must admit, it does sound like a contradiction. Stalnaker's theory does, and mine does not, respect the opinion of any ordinary language speaker who cares to insist that it is a contradiction.«

As Stalnaker (1981, 92) points out, »it would be arbitrary to require a choice of one of [C<sub>1</sub> and C<sub>2</sub>] over the other, but . . . this is not at issue. What is at issue is what conclusion about the truth values of the counterfactuals should be drawn from the fact that such a choice would be arbitrary.« The conclusion drawn by Lewis (1973) is that both C<sub>1</sub> and C<sub>2</sub> are false, and hence that conditional excluded middle is not logically valid. The conclusion drawn by Stalnaker (1968; 1981) is that both C<sub>1</sub> and C<sub>2</sub> are indeterminate, and that conditional excluded middle remains without counterexample and is logically valid. End of discussion.

An example from the more recent literature is the discussion between Gillies (2007) and Moss (2012) about where to draw the line between the semantics and the pragmatics of counterfactuals. Here is the relevant background (I have changed the examples).

Lewis (1973, 10), referring to Sobel (1970), uses so-called Sobel sequences to argue against the analysis of the counterfactual conditional as a strict conditional. Sobel sequences are examples such as

S<sub>1</sub> If Angela were to have wine tonight, she would have red wine,

S<sub>2</sub> If Angela were to have wine tonight, and there was only white wine left, she would not have red wine.

Lewis (1973) assumes that counterfactuals such as these can be jointly true. This is not so if the counterfactual conditional is a *strict* conditional, i. e., a material conditional that is necessarily true. Lewis (1973) concludes that the counterfactual conditional is a *variably* strict conditional, i. e., a strict conditional whose strictness varies with the antecedent.

Gillies (2007) considers *reverse* Sobel sequences:

S2 If Angela were to have wine tonight, and there was only white wine left, she would not have red wine,

S1 If Angela were to have wine tonight, she would have red wine.

For Gillies (2007, 332) an example such as »this sounds for all the world like a contradiction.« Gillies (2007) then goes on to argue that the counterfactual conditional is a strict conditional, but one whose truth values interact with context in such a way that the order in which two counterfactual conditionals are asserted matters.

Moss (2012) defends Lewis (1973)'s analysis of the counterfactual conditional as a variably strict conditional. Moss (2012) admits that reverse Sobel sequences are »generally infelicitous,« though contrary to Gillies (2007) she does not consider them to be contradictory. Moss (2012) first considers sentences other than counterfactual conditionals and notes that the order in which they are uttered matters:

M1 There is red wine in Wolfgang's cellar.

M2 But Wolfgang is always out of red wine.

For her a conversation like this is felicitous, while the reversed one is not:

M2' Wolfgang is always out of red wine.

M1' But there is red wine in Wolfgang's cellar.

According to Moss (2012, 568), »[o]ur intuitions about [an example such as M'] point towards a general principle governing assertability [which] tells us that if a speaker cannot rule out a possibility made salient by some utterance, then it is irresponsible of her to assert a proposition incompatible with this possibility.«

As before, »[w]hat is at issue is what conclusions about the truth values of the counterfactuals should be drawn« from the fact that reverse Sobel sequences are considered to be infelicitous. The conclusion drawn by Gillies (2007) is that the truth values of counterfactuals depend on the order in which they are uttered. The conclusion drawn by Moss (2012) is that the

assertability conditions, but not the truth values, of counterfactuals depend on the order in which they are uttered. End of discussion.

Both examples<sup>1</sup> illustrate a common pattern. Two philosophers share an intuition – the arbitrariness of choosing between  $C_1$  and  $C_2$ , the infelicity of reverse Sobel sequences – but disagree on the details of the intuition. Where Lewis (1973) intuits falsity in addition to arbitrariness, Stalnaker (1981) intuits vagueness in addition to arbitrariness. Where Gillies (2007) intuits contradictoriness in addition to infelicity, Moss (2012) intuits unassertability in addition to infelicity.

This is bad news. Allegedly our intuitions are the data deciding between rival philosophical theories (Pust 2000). If we cannot agree on what the data are, whose intuitions to rely on, we cannot agree on which theory to accept. Besides discussions in philosophical logic and the philosophy of language this affects other discussions involving counterfactual conditionals: in epistemology, knowledge is analyzed in terms of counterfactuals (Nozick 1981; Roush 2005); in metaphysics, causation is analyzed in terms of counterfactuals (Collins, Hall, and Paul 2004; Paul and Hall 2013); in the philosophy of science, dispositions are analyzed in terms of counterfactuals (Mumford 1998); outside philosophy, historians use counterfactuals in thought experiments (Reiss 2009), and psychologists discuss regret and responsibility in counterfactual terms (Connolly, Ordóñez, and Coughlan 1997).

One reaction might be to go experimental (Knobe and Nichols 2008) and see which intuitions are more widespread. On my view this would not help much, because information about how intuitions are distributed across various populations will not settle the philosophical issue at hand. Indeed, on my view it would make things *worse*, because it would make us focus on what is *not* the arbiter in philosophical debates: intuitions.

Let  $p$  be a proposition of philosophical or other interest. Even if everybody always intuits that  $p$  with maximal strength, that does not make  $p$  true. Nor does it support the claim that  $p$  is true for anyone other than those who trust the intuiter's intuitions (which need not include the intuiter herself). For instance, I happily grant that, intuitively, skepticism is false. However, even the (ethical) intuitionist Moore (1942) seems to agree that having this intuition does not provide a refutation of skepticism (Baldwin 2004, sec. 6) for anyone other than the intuiter's devotees. Nor does having the intuition that there is a God do anything to prove the latter's existence

<sup>1</sup> Another example is provided by the discussion between Lewis (1973; 1981) and Stalnaker (1968; 1981) versus Kratzer (1981) and Pollock (1976) about the semantic principle of Comparability. The latter says, roughly, that any two worlds can be compared with respect to their similarity to the actual world (cf. Lewis 1981, sec. 5).

for anyone other than the faithful to the intuiter's intuitions (about the relevant subject matter). Instead, in order to establish a claim, one needs to provide an *argument* whose premises all parties can judge for themselves.

The argument from »intuitively  $p$ « to » $p$ «, like the arguments from »plausibly  $p$ « to » $p$ « and from »hopefully  $p$ « to » $p$ «, just states the conclusion  $p$  without supporting it. Indeed, the argument from »intuitively  $p$ « to » $p$ « not only merely states the conclusion without supporting it, but it does so under the false pretense that there was support for it.

At least, this is the case if there is no positive argument for the reliability of intuitions that does not already presuppose that intuitions are reliable. Such a question-begging argument for the reliability of intuitions could run as follows. Start by assuming that universally shared intuitions have true contents, or at least do so more often than not. Then proceed from this assumption to the conclusion that the intuitions held by professional philosophers have true contents more often than not, because philosophers are expert intuiters – that is, because the intuitions held by philosophers tend to pick out the most widely held intuitions.

This question-begging argument fails to establish that intuitions are reliable in the same way as the following argument fails to establish that beliefs, or opinions, are reliable: universally held beliefs have true contents, and therefore the »expert opinions« held by professional pollsters are true more often than not, because they tend to pick out the most widely held beliefs. Belief and intuition may be a matter of democracy, but truth is not. It is also not enough to explain away negative evidence that supports the claim that intuitions are unreliable, as Boyd and Nagel (2014) seem to think. To show that there is less disagreement among intuitions across various populations than had been thought initially, does not provide any positive evidence for the reliability of intuitions. The lack of evidence for the unreliability of the beliefs held by various dogmatists and theists does not do anything to refute skepticism, or to prove God's existence.

Even if we grant that intuition is for philosophy what perception is for the sciences, we should still be focusing on arguments. According to contemporary physics, what reality is like differs radically from what we perceive it to be like (not that I would understand any of the physics). Whether we focus on the macroscopic level and huge objects far away, as relativistic physics often does, or on the microscopic level and tiny objects, as quantum physics often does, these objects and their properties are not perceived, but *inferred* – that is, *inferred from physical theory*. Indeed, in many cases these objects and their properties are in principle *unobservable*. Perception is the result of evolution and a means to attaining many ends, including survival and procreation. Finding out what reality (and not just

our immediate environment) is like, let alone at a fundamental level, is not one of these ends.<sup>2</sup>

Similarly, intuition is a result of evolution and a means to attaining many ends, including survival and procreation. However, revealing philosophical truths is not one of them. Physics has made progress by focusing on theory, and by ignoring the illusions generated by perception, such as that the earth appears to be flat, and that the sun appears to revolve around the earth. If intuition really plays the role for philosophy that perception plays for the sciences, we have even more reason to focus on philosophical theory, and what we can infer from it, and to ignore the delusions generated by intuition.

An example from a different discipline illustrates that intuiting  $p$  does not provide a basis for an argument for  $p$ . To the best of my knowledge there is widespread agreement among mathematicians that Goldbach's conjecture that every even integer is the sum of two primes is true. Some mathematicians (e.g., Knuth 2001) even think of this conjecture as a witness to the incompleteness of Peano arithmetic: a statement in the language of Peano arithmetic that is true, but provable in Peano arithmetic only if the latter is inconsistent. If reliance on intuitions ever were enough to support a conjecture, then surely in a case like this where it is believed that an argument for the conjecture is impossible (or, strictly speaking, possible only if arithmetic is inconsistent). Yet no mathematician relies on Goldbach's conjecture in a proof. As its name suggests, Goldbach's conjecture is merely a conjecture, and such refreshing honesty would serve our discipline well.

What *seems to be* is not what *is*, and for philosophers to replace the study of what is with the study of what seems to be is to give up without admitting it. To be sure, it is certainly most interesting and fascinating to study which intuitions are shared by whom and under what conditions, and I admire and applaud the authors of such studies and their illuminating insights. I also have no doubts that intuitions are most useful as heuristic devices in coming up with various theories, philosophical or otherwise. However, the contents of an intuition are not made true by the having of

---

<sup>2</sup> Space does not permit a defense of, and the above does not presuppose, two theses I hold. First, every object in the external world is a theoretical entity, and every property instantiated in the external world is a theoretical property. Second, there is no objectively right or wrong way to carve up the world into individuals and properties – or other categories – which is what the adoption of a language or vehicle of representation does. There are only more or less useful languages for attaining one's ends. Let me illustrate this with vision as one example of perception (without claiming that vision ever acts in isolation): what I see are the two-dimensional ›contents‹ of my vision that I can introspect; everything else is inferred.

the intuition – nor are they made any likely for the infidel who still thinks for herself – even if the intuition is universally held by all at all times under all conditions and with maximal strength.

Perhaps there are situations where we simply lack the resources to dig deeper than to the level of intuitions. I do not want to deny this. However, in this case the honest thing to do is to say so rather than to proceed from this to the conclusion that the contents of the intuition are true, as intuiting  $p$  is as close as we can get to  $p$ . Sometimes we can dig deeper, though, and then stopping at the level of intuitions amounts to irrationality and irresponsibility if the cognitive goals of our profession include believing truths and disbelieving falsehoods. For others, as well as for us outside professional contexts, digging deeper may be a waste of time and energy if close enough is good enough. But as philosophers it is what we have signed up for.

Mathematicians mark claims that are proven or derived as theorems, the assumptions they are derived or proven from as axioms, and call conjectures such as Goldbach's by their name. Our discipline would profit from adopting such honest terminology rather than sweeping this difference under the carpet of eloquence. This is all the more pressing when the profession's lingua franca is not everybody's first language, and eloquence often shines especially bright when there is no argument in sight.

Focusing on intuitions tempts us to give up before we reach the cognitive goals of our profession, which I take to include believing truths and disbelieving falsehoods. It also tempts us to ascribe to ourselves more cognitive power than we do in fact have. If we stop at the level of intuitions when we can dig deeper, because close enough is good enough, then we risk falsely believing that we could also dig deeper but do not have to, when in fact we cannot. Such a sense of intellectual entitlement is reminiscent of the dogmatism enlightenment aimed to free us from and the speculative outbursts of metaphysics the logical positivists objected to. One need not subscribe to all claims of these movements in order to appreciate that their motivation and intent were noble and of broad social relevance, and that our discipline's relying overly on intuitions instead of arguments threatens to lead us back into times where faith and authority trumped argument and inference.

Indeed, I am inclined to think that contemporary analytic philosophy's obsession with intuitions is partly due to the secularization of our discipline, and the resulting demise in argumentative power. When Descartes (1641) argued for foundationalism he could still help himself to the idea of a benevolent God to solve the difficulties of his philosophy: what is perceived clearly and distinctly is true, he said, because God is not a

deceiver, and so clear and distinct perception is an infallible source of knowledge.

We cannot appeal to this benevolent God anymore to solve our philosophical problems, and for good reasons.

But really we continue doing so covertly when we appeal to intuitions where digging deeper becomes too cumbersome. What is intuited rationally is true, we say, because, well, because such is the nature of rational intuition: it is a fallible source of a priori knowledge (Bealer 1998). Descartes (1641) was at least honest and – in his own words, of course – admitted giving up. We just start to talk about rationality (or »pure reason«; Bonjour 1998) instead, and thereby cover our sense of intellectual entitlement that attributes supernatural powers to intuition.

## 2. MEANS–END PHILOSOPHY

Means–end philosophy is one possible way to dig deeper in some cases. Its starting point is the view that epistemology is a normative discipline, together with an instrumental understanding of rationality or normativity according to which one ought to take the means to one's ends. In means–end philosophy metaphysical theories, such as a theory of counterfactuals, are derived from normative principles in epistemology. The latter are hypothetical imperatives that are justified by being shown to be the means to attaining the cognitive ends they are conditional upon. Which cognitive ends we happen to have is a factual question, or a matter of stipulation if we consider what one ought to do if one had certain cognitive ends. In a sense metaphysics is thus subordinate to epistemology, but this is not to say that the one would be more important than the other. Ultimately both are rooted in what gives meaning: ends, including introspectively accessible cognitive desires.

Means–end philosophy derives metaphysical theories from hypothetical imperatives in epistemology. Hypothetical imperatives say that one ought to do something that is within one's reach given that one has a certain cognitive goal. For example, a hypothetical imperative might say that one ought to have consistent beliefs given that one has the cognitive goal of holding only true beliefs. A hypothetical imperative holds only if obeying the imperative, doing what it says one should do, is a means to attaining the cognitive end the imperative is conditional upon. The hypothetical imperative that one ought to have consistent beliefs given that one has the cognitive goal of holding only true beliefs holds only if having consistent beliefs is a means to attaining the end of holding only true beliefs. We justify a hypothetical imperative not by appealing to intuitions, but by

providing an argument which establishes that obeying the imperative is a means to attaining the cognitive end the imperative is conditional upon. We may be able to justify the hypothetical imperative that one ought to have consistent beliefs, given that one aims at the cognitive end of holding only true beliefs, by proving that the consistency of one's beliefs is necessary for their joint truth. Of course, if one does not aim at holding only true beliefs, this will not cut any ice. But that is besides the point: it is mistaking a hypothetical imperative for a categorical imperative.

Means–end philosophy thus justifies normative principles from epistemology not, or not exclusively, on the basis of subjective intuitions, but by showing them to be the (often objective) means to attaining some cognitive end. It is philosophy with a focus on arguments instead of intuitions. This philosophical methodology is then extended to metaphysics by showing that some metaphysical theories follow from some epistemological principles. Let me illustrate this with an example.

Let us assume that counterfactuals express propositions that have truth values, and do not merely express conditional beliefs as Edgington (2008) and Spohn (2013a; 2015) have it. The received view is that the meaning of counterfactuals is given by a definition that involves two things: possible worlds and another thing (both of which are mind-dependent constructs or ideas according to the modal idealism of Huber 2016, sec. 3). These two items can be combined in different ways, and so one also needs to assume a particular such way. Furthermore, not everybody who holds that counterfactuals express propositions subscribes to the received view (Fine 2012). The received view is, well, just that. According to Stalnaker (1968), with whom the received view originates, the other thing besides possible worlds is similarity as captured by a set of selection functions. According to Lewis (1973) it is similarity as captured by a system of spheres. Let us call this other thing, whatever it is, *counterfactual distance*. Counterfactual distance is at best similar to similarity, and it will figure as a primitive on the present account. This does not mean it is void of content, though.

Counterfactual distances can be put in relation to conditional beliefs, just as objective chances can be put in relation to subjective credences (Lewis 1980). The normative principle that entails the semantics for counterfactuals does precisely this. Huber (2014) dubs this principle the »royal rule.« It says that an ideal doxastic agent's grade of disbelief in a proposition  $A$  conditional on the information that the counterfactual distance to the closest  $A$ -worlds equals  $n$ , and on no further information that is inadmissible, ought to be equal to  $n$ .

Purely modal information about counterfactual distances is admissible with respect to the claim that the counterfactual distance to the closest

$A$ -worlds equals  $n$ , unless it contradicts this claim. In particular, the theory of counterfactuality – that is, the theory of deterministic alethic or metaphysical modality – whatever it is or says, is admissible. This implies that the counterfactual distance distribution of a possible world equals an ideal doxastic agent's subjective grading of disbelief conditional on the theory of counterfactuality of this possible world.

As will be explained in the next section, the grading of disbelief of an ideal doxastic agent with the relevant cognitive goals ought to be a ranking function (Spohn 1988; 2012). A conditional ranking function is a ranking function. Hence the counterfactual distance distribution of a possible world has the structure of a ranking function. Substituting the latter for Stalnaker (1968)'s selection functions or Lewis (1973)'s system of spheres, and combining them with possible worlds in one way, results in a semantics for counterfactuals with respect to which the basic conditional logic  $V$  is correct and complete (a different combination results in a semantics with respect to which the system  $VW$  is correct and complete).

This application of means–end philosophy thus sides with Lewis (1973) against Stalnaker (1968; 1981) as far as conditional excluded middle is concerned. Furthermore, it sides with Moss (2012) against Gillies (2007) with regard to the question where to draw the line between the semantics and the pragmatics of counterfactuals.<sup>3</sup> Intuitions are not needed for any of this, although there are, of course, several assumptions that may or may not be true.

This leaves the question why the ideal doxastic agent ought to obey the normative principle that relates counterfactual conditionals and conditional beliefs, the royal rule. Huber (2014) stresses frequently that the royal rule is not to be accepted on intuitive grounds, but rather because it is the means to attaining some pertinent cognitive end. Yet Huber (2014) leaves us in the dark as to what this pertinent cognitive end might be. Huber (2016) intends to fill this lacuna by showing that the royal rule is a necessary and sufficient means to attaining a cognitive goal that can be thought of as a counterfactual version in terms of specific conditional beliefs of what James (1896) calls »our first and great commandments as would-be knowers«, namely: »Believe truth! Shun error!«

This counterfactual version has its origin in Nozick (1981)'s sensitivity condition. Ignoring details such as the method that is used in belief formation, the latter says, for any proposition  $A$ : if  $A$  were false, then the ideal

---

<sup>3</sup> It assumes with Lewis (1973; 1981) and Stalnaker (1968; 1981), but against, e.g., Kratzer (1981) and Pollock (1976), that counterfactual distances are comparable.

doxastic agent would not believe  $A$ . Nozick (1981) intends this principle and another one, adherence, as bridging the gap between true belief on the one hand and knowledge on the other (for a recent discussion, see Roush 2005). Sosa (1999) and Williamson (2000) think a different condition, safety, is better suited for the job of describing a distinctive feature of knowledge.

Conceptual analysis is not the business of the present author, nor is explication (see Spohn 2013b for an excellent critique of the industry that followed Gettier 1963). Nor is this paper intended to suggest that there would be anything objectively desirable about sensitivity. Instead, the aim is to illustrate how establishing a means–end relationship between a normative principle on the one hand and a pertinent cognitive end on the other justifies the normative principle relative to the cognitive end. The royal rule will be seen to be necessary and sufficient for a strong, but conditional, formulation of sensitivity. This result tells an ideal doxastic agent what she should do if she has the cognitive end described by this strong, but conditional formulation of sensitivity. It does not say anything if she does not have this end (unless she aims to avoid the end).

In the next section I will briefly describe the relevant theory of conditional belief, the semantics for counterfactuals, and the royal rule that relates counterfactual conditionals and conditional belief. Readers familiar with these items can directly move on to section 4, which retells Huber (2016)'s argument for the thesis that the royal rule is a necessary and sufficient means to attaining the cognitive end of tracking the truth in the sense of a strong, but conditional version of sensitivity.

### 3. CONDITIONAL BELIEF AND COUNTERFACTUALS

Ranking functions have been introduced by Spohn (1988; 2012) in order to model qualitative conditional belief. Ranking theory is quantitative or numerical in the sense that ranking functions assign numbers to propositions, which are the objects of belief in this theory. These numbers are needed for the definition of conditional ranking functions representing conditional beliefs. As we will see, though, once conditional ranking functions are defined we can interpret everything in purely qualitative, but conditional, terms.

Consider a set of possible worlds  $W$  and an algebra  $\mathcal{A}$  of propositions over  $W$ . A function  $\rho: \mathcal{A} \rightarrow \mathbb{N} \cup \{\infty\}$  from  $\mathcal{A}$  into the set of natural numbers,  $\mathbb{N}$ , extended by  $\infty$ ,  $\mathbb{N} \cup \{\infty\}$ , is a *finitely/countable/completely minimi-*

*tive ranking function* on  $\mathcal{A}$  just in case for all finite/countable/arbitrary sets of propositions  $\mathcal{B} \subseteq \mathcal{A}$ :

$$\begin{aligned} \varrho(W) &= 0, \\ \varrho(\emptyset) &= \infty, \\ \varrho(\bigcup \mathcal{B}) &= \min\{\varrho(A) : A \in \mathcal{B}\}. \end{aligned}$$

For a non-empty or consistent proposition  $A \neq \emptyset$  from  $\mathcal{A}$  the conditional ranking function  $\varrho(\cdot | A) : \mathcal{A} \setminus \{\emptyset\} \rightarrow \mathbb{N} \cup \{\infty\}$  based on the nonconditional ranking function  $\varrho(\cdot) : \mathcal{A} \rightarrow \mathbb{N} \cup \{\infty\}$  is defined as

$$\varrho(\cdot | A) = \begin{cases} \varrho(\cdot \cap A) - \varrho(A), & \text{if } \varrho(A) < \infty, \\ \infty \text{ or } 0, & \text{if } \varrho(A) = \infty. \end{cases}$$

Goldszmidt and Pearl (1996, 63) suggest  $\infty$  in the latter clause. Huber (2006, 464) suggests 0 and stipulates  $\varrho(\emptyset | A) = \infty$  to ensure that every conditional ranking function is a ranking function on  $\mathcal{A}$ . A ranking function  $\varrho$  is *regular* if and only if for all non-empty or consistent propositions  $A$  from  $\mathcal{A}$ ,

$$\varrho(A) < \varrho(\emptyset).^4$$

Doxastically, ranks are interpreted as grades of disbelief. A proposition  $A$  is *disbelieved* just in case  $A$  is assigned a positive rank,  $\varrho(A) > 0$ . The proposition  $A$  is *believed* just in case its complement or negation,  $\bar{A}$ , is disbelieved,  $\varrho(\bar{A}) > 0$ .

A proposition  $A$  is *disbelieved conditional on* a proposition  $C$  just in case  $A$  is assigned a positive rank conditional on  $C$ , that is,  $\varrho(A | C) > 0$ .  $A$  is *believed conditional on*  $C$  just in case its complement or negation,  $\bar{A}$ , is disbelieved conditional on  $C$ , i.e.,  $\varrho(\bar{A} | C) > 0$ . It takes getting used to reading positive numbers in this ›negative‹ way, but mathematically this is the simplest formulation of ranking theory. Note that a proposition  $A$  is believed just in case  $A$  is believed conditional on the tautological proposition  $W$ . This is so because  $\varrho(\bar{A}) = \varrho(\bar{A} | W)$ .

It follows from the definition of conditional ranking functions that the ideal doxastic agent should not disbelieve a non-empty or consistent proposition  $A$  conditional on itself:  $\varrho(A | A) = \varrho(A \cap A) - \varrho(A) = 0$ . I will refer to this consequence below. In doxastic terms the first axiom says that the ideal doxastic agent should not disbelieve the tautological proposition  $W$ . The second axiom says that she should disbelieve the

<sup>4</sup> In contrast to probability theory it is always possible to define a regular ranking function, no matter how rich or fine-grained the underlying algebra of propositions.

empty or contradictory proposition  $\emptyset$  with maximal strength  $\infty$ . Given the definition of conditional ranking functions, the second axiom can be read in purely qualitative, but conditional, terms. Read this way it says that the ideal doxastic agent should disbelieve the empty or contradictory proposition conditional on any proposition with a finite rank. This implies that she should believe the tautological proposition with maximal strength, or conditional on any proposition with a finite rank.

Finite minimitivity is the weakest version of the third axiom. It states that  $q(A \cup B) = \min\{q(A), q(B)\}$  for any two propositions  $A$  and  $B$ . Part of what finite minimitivity says is that the ideal doxastic agent should disbelieve a disjunction  $A \cup B$  just in case she disbelieves both its disjuncts  $A$  and  $B$ . Given the definition of conditional ranking functions, finite minimitivity extends this requirement to conditional beliefs. As noted above, the definition of conditional ranking functions implies that the ideal doxastic agent should not disbelieve a proposition conditional on itself. Given this consequence, finite minimitivity says the following (in purely qualitative but conditional terms): the ideal doxastic agent should conditionally disbelieve a disjunction  $A \cup B$  just in case she conditionally disbelieves both its disjuncts  $A$  and  $B$ . Countable and complete minimitivity extend this requirement to disjunctions of countably and arbitrarily many disjuncts, respectively.

Interpreted doxastically, these axioms are synchronic norms for organizing the ideal doxastic agent's beliefs and conditional beliefs at a given moment in time. They are supplemented by three diachronic norms for updating her beliefs over time if new information of various formats is received. Plain conditionalization (Spohn 1988) mirrors the update rule of strict conditionalization from probability theory (Vineberg 2000): it is defined for the case where the new information comes in the form of a ›certainty,‹ a proposition that the ideal doxastic agent comes to believe with maximal strength. Spohn conditionalization (Spohn 1988) mirrors the update rule of Jeffrey conditionalization from probability theory (Jeffrey 1983): it is defined for the case where the new information comes in the form of new ranks for the elements of an ›evidential partition.‹ Shenoy conditionalization (Shenoy 1991) mirrors the update rule of Field conditionalization from probability theory (Field 1978): it is defined for the case where the new information reports the differences between the old and the new ranks for the elements of an evidential partition.

The resulting package of synchronic and diachronic norms can be justified by the consistency argument (Huber 2007) in much the same way that probability theory can be justified by the Dutch book argument. The consistency argument shows that obeying the synchronic and diachronic

rules of ranking theory is a necessary and sufficient means to attaining the cognitive end of always holding beliefs that are jointly consistent and deductively closed. To the extent that the ideal doxastic agent has this goal, she should obey the norms of ranking theory. It is not that we are telling her what and how to believe. *She* is the one who is assumed to have this goal. We merely point out the (objectively) obtaining means–end relationships. Of course, if the ideal doxastic agent does not aim at always holding beliefs that are jointly consistent and deductively closed, our response will cut no ice. But, as already mentioned before, that is beside the point: it is mistaking a hypothetical imperative for a categorical one.

Huber (2014; 2016) introduces *alethic* ranking functions, which are interpreted as counterfactual distances. Let  $\mathcal{L}_0$  be the smallest set that includes a given countable set  $PV$  of propositional variables and is closed under the classical connectives  $\neg$ ,  $\wedge$ ,  $\vee$ , and  $\supset$ . (They are used autonomously, as are all other symbols of the object language.) Let  $\mathcal{L}_1$  be the smallest set containing all well-formed formulas in  $\mathcal{L}_0$  as well as  $\alpha \square \rightarrow \beta$  for any two  $\alpha$  and  $\beta$  from  $\mathcal{L}_0$ . Let  $\mathcal{L}$  be the smallest set that includes  $\mathcal{L}_1$  and is closed under the classical connectives. So the language  $\mathcal{L}$  is built up from a countable set of propositional variables in the usual way, with the only exception that  $\alpha \square \rightarrow \gamma$  is a well-formed formula if *and only if*  $\alpha$  and  $\gamma$  are well-formed formulas *and do not contain an occurrence of*  $\square \rightarrow$ .

So much for the syntax. As to the considerably more complicated semantics,  $(F, \mathcal{A}_F, R, W, \llbracket \cdot \rrbracket)$  is a *ranking-theoretic model* for  $\mathcal{L}$  just in case  $F$  is a non-empty set of ›factual‹ worlds,  $\mathcal{A}_F$  is an algebra over  $F$ ,  $R$  is a set of ranking functions  $r: \mathcal{A}_F \rightarrow \mathbb{N} \cup \{\infty\}$ ,  $W \subseteq F \times R$  – the set of possible worlds – is such that for each  $f \in F$  there is *at least one*  $r \in R$  such that  $(f, r) \in W$ , and  $\llbracket \cdot \rrbracket: \mathcal{L} \rightarrow \wp(W)$  is an interpretation function such that for all  $\alpha$  and  $\beta$  from  $\mathcal{L}$ :

1. if  $p \in PV$ , then  $\llbracket p \rrbracket \subseteq W$  is such that: if  $(f, r) \in \llbracket p \rrbracket$  for some  $f \in F$  and some  $r \in R$ , then  $(f, r') \in \llbracket p \rrbracket$  for all  $r' \in R$  such that  $(f, r') \in W$ ; and  $\text{fact}(\llbracket p \rrbracket) \in \mathcal{A}_F$ ,
2.  $\llbracket \neg \alpha \rrbracket = W \setminus \llbracket \alpha \rrbracket$ ,  $\llbracket \alpha \wedge \beta \rrbracket = \llbracket \alpha \rrbracket \cap \llbracket \beta \rrbracket$ , and analogously for  $\vee$  and  $\supset$ ,
3.  $\llbracket \alpha \square \rightarrow \beta \rrbracket = \left\{ w = (f_w, r_w) \in W : \text{fact}(\llbracket \alpha \rrbracket)^{r_w} \subseteq \text{fact}(\llbracket \beta \rrbracket) \right\}$ ,

where  $\text{fact}(\llbracket \alpha \rrbracket) = \{ f \in F : \exists r \in R : (f, r) \in \llbracket \alpha \rrbracket \}$  and the superscript  $\gg r_w \ll$  indicates that we pick out those factual  $\llbracket \alpha \rrbracket$ -worlds that are  $r_w$ -minimal,<sup>5</sup>

$$A^{r_w} = \left\{ f \in A \in \mathcal{A}_F : \forall B \in \mathcal{A}_F : \text{if } f \in B, \text{ then } r_w(B) \leq r_w(A) \right\}.$$

<sup>5</sup> Alternatively one could pick out those factual  $\llbracket \alpha \rrbracket$ -worlds  $f$  that are  $r_w$ -minimal or no more distant according to  $r_w$  than  $f_w$  is (in the precise sense that is formalized in analogy to the definition of  $A^{r_w}$  above). This is an alternative definition, because  $f_w$  need not be assigned

The reason for this rather complicated way of setting things up is that Huber (2014) thinks it is important to distinguish between the factual components  $f_w$  and the modal components  $r_w$  of possible worlds  $w = (f_w, r_w)$  about which beliefs are formed by a subjective ranking function  $\varrho$  (for an alternative set-up, see the appendix of Huber 2016). This in turn is so because of the royal rule, the normative principle according to which an ideal doxastic agent's conditional subjective ranks are constrained or guided by the alethic ranks.

For the purposes of the present paper we may simplify things slightly by letting the alethic ranking functions  $r$  be defined on the powerset  $\wp(F)$  of  $F$ , and by letting the set of possible worlds be  $W = F \times R$ , where  $R$  is the set of *all* alethic ranking functions on  $\wp(F)$ . The ideal doxastic agent's subjective ranking function  $\varrho$  is defined on the powerset of  $W$  and assumed to be regular.

**Royal Rule for  $A$ .** *Let  $A \subseteq F$  be a fixed factual proposition, and let  $\gg r(A) = n \ll$  denote the proposition that the counterfactual distance to the closest  $A$ -worlds equals  $n$ , i.e.,  $\{w = (f_w, r_w) \in W : r_w(A) = n\}$ . Let  $n$  be a number from  $\mathbb{N} \cup \{\infty\}$ . Let  $\varrho : \wp(F \times R) \rightarrow \mathbb{N} \cup \{\infty\}$  be an ideal doxastic agent's grading of disbelief, which is assumed to be a regular subjective ranking function. Finally, let  $E$  be an arbitrary proposition that is ›admissible‹ in the sense that it is purely modal information (if  $(f, r) \in E$  for some  $f \in F$  and some  $r \in R$ , then  $(f', r) \in E$  for all  $f' \in F$ ) that is consistent with  $r(A) = n$ . Then*

$$\varrho\left(A \times R \mid (r(A) = n) \cap E\right) = n.$$

For a fixed factual proposition  $A$ , the royal rule for  $A$  says that an ideal doxastic agent's grade of disbelief for  $A$  conditional on the assumption that the counterfactual distance to the closest  $A$ -worlds equals  $n$ , and no further information that is not admissible, ought to be equal to  $n$ . An ideal doxastic agent ought to obey the royal rule for  $A$ , for every factual proposition  $A$ . Why she should do so is the topic of the next section (where I will focus on propositions and ignore sentences).

Ranks are numerical; but unlike probabilities, which are measured on an absolute scale, neither subjective nor alethic ranks utilize all the information carried by these numbers. Instead, subjective ranks, and hence alethic ranks, are at best measured on a ratio scale (Hild and Spohn 2008). I say ›at best‹ because even the choice of 0 as threshold for disbelief is

---

rank 0 by  $r_w$  and so can have a higher  $r_w$ -rank than the  $r_w$ -minimal factual  $\llbracket \alpha \rrbracket$ -worlds. In this case the resulting semantics is correct and complete with respect to the conditional logic VW that results from V by adding the axiom schema  $(\alpha \wedge (\alpha \Box \rightarrow \gamma)) \supset \gamma$ .

somewhat arbitrary, as Spohn (2015, 9) notes. Some positive, but finite, natural number would do just as well.<sup>6</sup> The royal rule is thus weaker than it might at first appear: alethic ranks guide subjective ranks *provided* the former are reported in terms of the latter's scale. Otherwise the royal rule is silent, as we would be comparing apples and oranges.

The royal rule implies that counterfactual distances that are numerical have the structure of a ranking function, and hence that talk of alethic *ranks* is appropriate. The present paper will now assume this consequence in the formulation of the cognitive end of tracking a fact. Next the paper will show that, given this consequence of the royal rule, obeying the royal rule is a means to attaining this end. The paper will then conclude that this justifies the royal rule. Therefore the argument of the present paper seems to be circular, as the justification of the royal rule seemingly relies on a consequence of this very rule.

This seeming circularity is only apparent, though. It is true that the formulation of the cognitive end of tracking a fact will make use of the above-mentioned consequence of the royal rule, and of the resulting theory of counterfactuals. After all, without some theory or other of counterfactuals the cognitive end of tracking a fact cannot be formulated. It is also true that I merely show that, given this theory of counterfactuals, obeying the royal rule is a means to attaining this end. However, it is not my argument but the formulation of the cognitive end of tracking a fact that assumes this theory of counterfactuals. Importantly, nobody is forcing the ideal doxastic agent, or anyone else for that matter, to adopt this cognitive end in its particular formulation. The royal rule is a hypothetical imperative which is justified relative to this cognitive end that one may or may not have.

#### 4. THE ROYAL RULE TRACKS THE FACTS

Consider the following formulation of Nozick (1981)'s sensitivity condition: *A-Sensitivity to A*. For a *fixed* factual proposition  $A \subseteq F$ : If  $A$  were true, then the ideal doxastic agent would not disbelieve  $A$  in the sense that  $q(A \times R) = 0$ .

The second occurrence of » $A$ « in » $A$ -sensitivity to  $A$ « indicates that we are interested in the question whether the ideal doxastic agent's belief in  $A$

---

<sup>6</sup> The corresponding fact about alethic ranks is used in the alternative definition in footnote 5, where the threshold for counterfactual distance is set equal to the maximum of the alethic rank  $r_w(\{f_w\})$  of the actual factual world and the alethic rank  $r_w(A)$  of the antecedent  $A$ , rather than equal to the latter.

is sensitive to the truth of  $A$ . The first occurrence of » $A$ « in » $A$ -sensitivity to  $A$ « indicates a different aspect in which this question depends on  $A$ . The notion of sensitivity is formulated in terms of an English counterfactual. On the present theory the meaning of this English counterfactual depends on a ranking function. As explained below, this ranking function is not one and the same for each antecedent  $A$ . Instead there will be different ranking functions for different antecedents. The first occurrence of » $A$ « in » $A$ -sensitivity to  $A$ « stresses this additional dependence on  $A$ .

To get things going let us assume that the ideal doxastic agent is *modally agnostic* and suspends judgment with respect to whether or not the alethic rank of any contingent factual proposition  $A$  is positive or not:  $q(r(A) = m) = 0$  and  $q(r(F \setminus A) = m) = 0$  for all natural numbers  $m$  in  $\mathbb{N}$ . What the royal rule asks of modally agnostic agents can be explained in analogy to Moore (1942)'s paradox. The latter is exemplified by sentences of the form » $A$ , and I don't believe that  $A$ « or, better suited for present purposes:

$A$ , and I disbelieve that  $A$ .

Moorean sentences are logically consistent, but somehow odd to be believed or asserted. On the present picture (dis-)beliefs come in grades, which leads us to *graded* versions of Moorean sentences:

$A$ , and I disbelieve  $A$  with strength  $n$ .

The oddity of such graded Moorean sentences increases with  $n$ . No oddity for  $n = 0$ , which means:  $A$ , and I don't disbelieve  $A$ . Some oddity for  $n > 0$ , which means:  $A$ , and I disbelieve  $A$  with strength  $n > 0$ . Maximal oddity for  $n = \infty$ , which means:  $A$ , and I am certain that  $A$  is false.

The final step is to replace *subjective* grades of disbelief with *alethic* counterfactual distances or »grades of bizarreness:«

$A$ , and the bizarreness of, or counterfactual distance to,  $A$  equals  $n$ .

Such counterfactual versions of graded Moorean sentences are logically consistent unless one made Lewis (1973, 14 f)'s weak or strong centering assumption, which the present paper does not.<sup>7</sup> They may, or may not,

<sup>7</sup> Lewis (1973, 14 f)'s weak centering assumption validates the axiom schema  $((\alpha \Box \rightarrow \gamma) \wedge \alpha) \supset \gamma$ . As noted previously, instead of making the weak centering assumption, this axiom schema can also be validated by changing the truth-condition for the counterfactual conditional and combining possible worlds and counterfactual distances as in footnote 5. Lewis (1973, 14 f)'s strong centering assumption validates the axiom schema  $(\alpha \wedge \gamma) \supset (\alpha \Box \rightarrow \gamma)$ . This axiom schema is not validated by the alternative truth-condition from footnote 5.

be odd to be asserted. However, they should be disbelieved according to the royal rule. And they should be disbelieved the firmer, the higher the grade  $n$ . What the royal rule asks of modally agnostic agents is that the conjunction  $\gg A$ , and  $r(A) = n \ll$  be disbelieved to grade  $n$ .

Ideal doxastic agents who are not modally agnostic are additionally asked to add to  $n$  their grade of disbelief that  $r(A) = n$ , that is,  $\gg A$ , and  $r(A) = n \ll$  should be disbelieved to grade  $n+k$ , where  $k$  is the ideal doxastic agent's grade of disbelief that  $r(A) = n$ .

In classical Moorean sentences the additional part corresponding to  $\varrho(r(A) = n) = k$  is a meta-disbelief about one's first-order disbelief in  $A$ . If made explicit it gives rise to infinitely long Moorean sentences:

$A$ , and I don't believe that  $A$ , and I don't believe that I don't believe that  $A$ , etc.

The auto-epistemological reflection principle (Spohn 2012, ch. 9, relying on Hild 1998) implies that the ideal doxastic agent is certain of what her own beliefs are. As a consequence it rules that the infinitely long and the classical Moorean sentences be disbelieved.

Let's introduce some terminology. A *case*  $\omega$  consists of three components: (i) a factual component  $f_\omega$  specifying the truth values of all factual propositions; (ii) an alethically modal component, i.e., the alethic ranking function  $r_\omega$ , specifying the truth values of all counterfactual conditionals; and (iii) a doxastically modal component, i.e., the subjective ranking function  $q_\omega$ , specifying all the ideal doxastic agent's beliefs and conditional beliefs. A case  $\omega$  can be represented as a triple  $(f_\omega, r_\omega, q_\omega)$ .

In order to make sense of  $A$ -sensitivity to  $A$  we need to assign counterfactual distances or ranks to these cases  $\omega$ . Otherwise the English counterfactual in the formulation of  $A$ -sensitivity to  $A$  cannot be assigned a truth value. This is a tricky task, as there is the threat that I am smuggling into these counterfactual distances or ranks whatever it is that I want to derive. So let me try to be as clear as possible.

The question we face can be stated as follows. We are given a set of possible cases  $\Omega = F \times R \times S$ , where  $F$  is the set of all possible factual components,  $R$  is the set of all alethically modal components (the alethic ranking functions defined on the power-set of  $F$ ), and  $S$  is the set of all doxastically modal components (the subjective ranking functions defined on the power-set of  $F \times R$ ). What we need to do is define one or more ranking function(s) *rank* on the power-set of  $F \times R \times S$ .

The standard move, sketched in the appendix of Huber (2016), is to equip each case  $\omega$  with its own ranking function  $rank_\omega$ . This results in a model of the form  $(\Omega, (rank_\omega)_{\omega \in \Omega})$ . However, doing so would run counter to the

very point of Huber (2014)'s set-up, which is designed such that modalities are not automatically iterated indefinitely. Yet precisely this would be the case for a model of such form. For this reason a different move is mandated.

The antecedents of the counterfactual conditionals we are considering are all purely factual propositions. Purely factual propositions are assigned alethic ranks in each possible case  $\omega$  by the alethic ranking functions  $r_\omega$ , but the cases themselves are not. However, in order to evaluate the English counterfactual in the formulation of  $A$ -sensitivity to  $A$ , ranks for the cases themselves, and not just for their factual components, are precisely what we need. Therefore we must assign ranks to the cases *vicariously*.

One option is to take  $rank_\omega(\{\omega'\}) = r_\omega(\{f_{\omega'}\})$ , but that is instantiating the problem mentioned above rather than avoiding it. A different option is to take  $rank(\{\omega\}) = r_\omega(\{f_\omega\})$ . This avoids the problem mentioned above and tells us how bizarre  $\omega$  is qua  $f_\omega$ -case, namely bizarre to degree  $r_\omega(\{f_\omega\})$ . However, it does not tell us how bizarre  $\omega$  is qua  $A$ -case, for an arbitrary factual proposition  $A \subseteq F$ . This information is not provided by the single value  $r_\omega(\{f_\omega\})$ . It is only provided by a case  $\omega$ 's alethic ranking function  $r_\omega$  *in its entirety*. That is, this information is only provided by the alethic ranks  $r_\omega(A)$  of *all* factual propositions  $A \subseteq F$ , and not just the alethic rank of the particular factual proposition  $\{f_\omega\}$ . The third option is to make this idea more precise, but to stay as neutral as possible and only use as much information from  $r_\omega$  as is needed in order to evaluate the English counterfactual in the formulation of  $A$ -sensitivity to  $A$ .

As indicated earlier, a consequence of taking the third option is that the ranks of a case depend on the antecedent of the English counterfactual that is evaluated. That is, we are dealing with models of the form  $(\Omega, (rank_A)_{A \subseteq F})$ , where  $\Omega = F \times R \times S$  and where, for each purely factual proposition  $A \subseteq F$ ,  $rank_A$  is a ranking function on the power-set of  $\Omega$  that specifies how bizarre or counterfactually distant a case  $\omega$  is qua case in which  $A$  is true.

Suppose two cases  $\omega_1$  and  $\omega_2$  agree that the factual proposition  $A$  is true. Under this assumption, the question we need to answer is this: Is  $\omega_1$  counterfactually less distant qua case in which  $A$  is true than  $\omega_2$  in the sense of the ranking function  $rank_A$ ? Huber (2016) proposes that this be the case if and only if the alethically modal components of  $\omega_1$  and  $\omega_2$  say so of  $A$ :

$$\{f_{\omega_1}, f_{\omega_2}\} \subseteq A \Rightarrow \left[ rank_A(\{\omega_1\}) < rank_A(\{\omega_2\}) \Leftrightarrow r_{\omega_1}(A) < r_{\omega_2}(A) \right].$$

On this proposal, being modally agnostic turns out to be a sufficient means to attaining the cognitive end of believing in a sensitive way even if the royal rule is not satisfied. However, this is a Pyrrhic victory, for the

former assumption and the royal rule together imply that the ideal doxastic agent is also factually agnostic. That is, the ideal doxastic agent who is modally agnostic and obeys the royal rule for every factual proposition also suspends judgment with respect to every contingent factual proposition. Such agnosticism is too high a price to pay.

Let us therefore drop the assumption of modal agnosticism and consider a formulation of Nozick (1981)'s sensitivity condition that uses conditional beliefs:

*Conditional A-Sensitivity to A.* For a *fixed* factual proposition  $A \subseteq F$ : If  $A$  were true, then the ideal doxastic agent would not disbelieve  $A$  conditional on the truth about the modal status of  $A$  (and, perhaps, other admissible information  $E$ ) in the sense that  $q(A \times R \mid r(A) \cap E) = 0$ .

The admissible information  $E$  is a purely modal proposition, where a proposition  $E \subseteq W = F \times R$  is purely modal if and only if, if  $(f, r) \in E$  for some  $f \in F$  and some  $r \in R$ , then  $(f', r) \in E$  for all  $f' \in F$ . The expression  $\gg r(A) \ll$ , on the other hand, does not stand for a proposition at all. This is so because what  $\gg$ the truth about the modal status of  $A$  $\ll$  depends on which case  $\omega$  the ideal doxastic agent is imagined to be in. This becomes more perspicuous if we reformulate the above condition as follows:

*Conditional A-Sensitivity to A.* For a *fixed* factual proposition  $A \subseteq F$ : A case  $\omega$  in which  $A$  is true, but in which the ideal doxastic agent  $q_\omega$  disbelieves  $A$  conditional on the truth about the modal status of  $A$  in  $\omega$  (and, perhaps, other information  $E$  that is admissible) in the sense that  $q_\omega(A \times R \mid r_\omega(A) \cap E) > 0$ , is bizarre, or counterfactually distant, i.e.,  $\text{rank}_A(\{\omega\}) > \text{rank}_A(A \times F \times S)$ .

Now  $r_\omega(A)$  is the proposition  $\{v = (f_v, r_v) \in F \times R = W : r_v(A) = r_\omega(A)\}$ . Since the admissible proposition  $E$  has to be consistent with  $r_\omega(A)$ , it depends on  $r_\omega(A)$ .

Asking an agent to do whatever it takes to have sensitive beliefs is much like asking an agent to do whatever it takes to believe all and only true propositions: it is asking too much of the agent. This is so despite my assumption that the agent is an ideal doxastic agent, which I take to mean that she does not suffer from any computational or other physical limitations, can always identify all logical and conceptual truths, gets to voluntarily decide what she believes, and never forgets any of her beliefs (see Huber 2013). Like ordinary agents, ideal doxastic agents do not foresee the future and are not omniscient.

Unlike the requirement of doing whatever it takes to be omniscient the royal rule is a rule that is within reach for ideal doxastic agents. The royal rule prescribes that an ideal doxastic agent hold various conditional

beliefs. The royal rule does not prescribe that the ideal doxastic agent hold a nonconditional belief in any contingent and purely factual proposition, or in any contingent and purely modal proposition. This is similar to the situation for the different requirement that one's beliefs be logically consistent: this norm does not prescribe that the ideal doxastic agent hold a nonconditional belief in any contingent proposition, but only that she not disbelieve *A* conditional on the assumption that she believes *A*.

A consequence of the fact that the royal rule is within an ideal doxastic agent's reach is that the ends that are furthered by the royal rule are limited. There is no rule within a nonomniscient ideal doxastic agent's reach that is necessary and sufficient for holding all and only true beliefs in all possible worlds at all times. Something similar is true for *A*-sensitivity to *A*, as we have seen. While it is within an ideal doxastic agent's reach to be modally agnostic, the royal rule then forces her to also be factually agnostic. On the other hand, conditional *A*-sensitivity to *A* is such a limited end that can be achieved by following a norm that is within the ideal doxastic agent's reach without her succumbing to agnosticism. The reason is, of course, its conditional nature. The agent does not have to believe a true factual proposition. The agent merely has to avoid disbelieving a true factual proposition conditional on what is true about this proposition's modal status in the cases the agent is imagined to be in.

The necessary-and-sufficient relationship goes both ways: what's within reach may not always be very desirable. That is, the conditional nature of conditional *A*-sensitivity to *A* may make this cognitive end appear to be too weak to be of interest. However, this is not the best way to think of it. Consider again consistency, now as an end rather than a means. Consistency alone won't sell a theory. However, imagine how hard it would be to sell a theory that was not even consistent!

Similarly, beliefs that are merely sensitive conditional on the truth about the modal status of their contents may be hard to sell, but good luck selling beliefs that are *not even* sensitive conditional on the truth about the modal status of their contents! If your belief about the modal status of *A* is true, then having sensitive beliefs in the sense of conditional *A*-sensitivity to *A* guarantees that you would not disbelieve *A* if it were true. You get two items of potential value for the price of one: you would automatically, and for free, not disbelieve a true factual proposition once you managed to believe the truth about its modal status.<sup>8</sup> However, if your beliefs are *not* sensitive in the sense of conditional *A*-sensitivity to *A*, then

---

<sup>8</sup> What to do in order to avoid believing a falsehood about the modal status of a factual proposition is explained in Huber (2015).

not even believing the truth about the modal status of  $A$  would prevent you automatically from disbelieving  $A$  if it were true. You would have to pay, or work, extra for this additional item of potential value.

The royal rule for  $A$  is a sufficient, but not a necessary, means to attaining the cognitive end of  $A$ -sensitivity to  $A$ . It is, however, a necessary and sufficient means to attain the cognitive end of  $A$ -tracking the fact that  $A$ :

*A-Tracking the Fact that A.* For a *fixed* factual proposition  $A \subseteq F$ : Of two possible cases  $\omega_1$  and  $\omega_2$  in which  $A$  is true, the former is counterfactually more distant qua case in which  $A$  is true than the latter if and only if the ideal doxastic agent disbelieves  $A$  in  $\omega_1$  conditional on the truth about its modal status in  $\omega_1$  (and, perhaps, other admissible information  $E_1$ ) to a higher degree than the degree to which her counterpart  $q_{\omega_2}$  disbelieves  $A$  in  $\omega_2$  conditional on the truth about its modal status in  $\omega_2$  (and, perhaps, other admissible information  $E_2$ ), i. e.,

$$q_{\omega_1}(A \times R \mid r_{\omega_1}(A) \cap E_1) > q_{\omega_2}(A \times R \mid r_{\omega_2}(A) \cap E_2) \Leftrightarrow \\ \text{rank}_A(\{\omega_1\}) > \text{rank}_A(\{\omega_2\}).$$

This completes my argument for the thesis that an ideal doxastic agent ought to obey the royal rule if she aims at attaining the cognitive end of believing in a conditionally sensitive way: for a fixed factual proposition  $A \subseteq F$ , the royal rule for  $A$ , and it only,  $A$ -tracks the fact that  $A$ . As always, the objection that the ideal doxastic agent may not have this cognitive end cuts no ice, but mistakes a hypothetical imperative for a categorical one. Needless to say, if the ideal doxastic agent does not have this cognitive end, there is nothing wrong with her. For means–end philosophy agrees with Old Fritz and his dictum to the effect: *Alle sollen nach ihrer Façon selig werden.*

#### ACKNOWLEDGMENTS

I am grateful to Thomas Kroedel and Christopher von Bülow for helpful comments on, and to Kevin Kuhl for proof-reading, a previous version of this paper, which heavily relies on Huber (2014) and, even more heavily, on Huber (2016).

My research was supported by the Canadian SSHRC through its Insight program and by the Connaught Foundation through its New Researcher program.

## REFERENCES

- Baldwin, Tom. 2004. »George Edward Moore«. In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta.
- Bealer, George. 1998. »Intuition and the Autonomy of Philosophy«. In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, edited by M. DePaul and W. Ramsey, 201–240. Lanham Md.: Rowman & Littlefield.
- BonJour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Boyd, Kenneth, and Jennifer Nagel. 2014. »The Reliability of Epistemic Intuitions«. In *Current Controversies in Experimental Philosophy*, edited by E. Machery and E. O’Neill, 109–127. London: Routledge.
- Collins, John, Ned Hall, and L. A. Paul, eds. 2004. *Causation and Counterfactuals*. Cambridge Mass.: MIT Press.
- Connolly, Terry, Lisa D. Ordóñez, and Richard Coughlan. 1997. »Regret and Responsibility in the Evaluation of Decision Outcomes«. *Organizational Behavior and Human Decision Processes* 70:73–85.
- Descartes, René. 1641/1998. *Meditations and Other Metaphysical Writings*. Translated by Desmond M. Clarke. London: Penguin.
- Edgington, Dorothy. 2008. »Counterfactuals«. *Proceedings of the Aristotelian Society* 108:1–21.
- Field, Hartry. 1978. »A Note on Jeffrey Conditionalization«. *Philosophy of Science* 45:361–367.
- Fine, Kit. 2012. »A Difficulty for the Possible Worlds Analysis of Counterfactuals«. *Synthese* 189:29–57.
- Gettier, Edmund L. 1963. »Is Justified True Belief Knowledge?«. *Analysis* 23:121–123.
- Gillies, Anthony S. 2007. »Counterfactual Scorekeeping«. *Linguistics and Philosophy* 30:329–360.
- Goldszmidt, Moises, and Judea Pearl. 1996. »Qualitative Probabilities for Default Reasoning, Belief Revision, and Causal Modeling«. *Artificial Intelligence* 84:57–112.
- Hild, Matthias. 1998. »Auto-Epistemology and Updating«. *Philosophical Studies* 92:321–361.
- Hild, Matthias, and Wolfgang Spohn. 2008. »The Measurement of Ranks and the Laws of Iterated Contraction«. *Artificial Intelligence* 172:1195–1218.
- Huber, Franz. 2006. »Ranking Functions and Rankings on Languages«. *Artificial Intelligence* 170:462–471.
- . 2007. »The Consistency Argument for Ranking Functions«. *Studia Logica* 86:299–329.
- . 2013. »Belief First«. *The Reasoner* 7:82.
- . 2014. »New Foundations for Counterfactuals«. *Synthese* 191:2167–2193.
- . 2015. »What Should I Believe About What Would Have Been the Case?«. *Journal of Philosophical Logic* 44:81–110.

- . 2016. »Why Follow the Royal Rule?« *Synthese*.
- James, William. 1896. »The Will to Believe«. In *The Will to Believe and Other Essays in Popular Philosophy*, edited by F. Burkhardt, F. Bowers, and I. Skrupskelis, 291–341. Cambridge Mass.: Harvard University Press 1979.
- Jeffrey, Richard C. 1983. *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.
- Knobe, Joshua, and Shaun Nichols, eds. 2008. *Experimental Philosophy*. Oxford: Oxford University Press.
- Knuth, Donald. 2001. »All Questions Answered«. *Notices of the AMS* 49:318–324. <http://www.ams.org/notices/200203/fea-knuth.pdf>
- Kratzer, Angelika. 1981. »Partition and Revision: The Semantics of Counterfactuals«. *Journal of Philosophical Logic* 10:201–216.
- Lewis, David K. 1973. *Counterfactuals*. Cambridge Mass.: Harvard University Press.
- . 1980. »A Subjectivist’s Guide to Objective Chance«. In *Studies in Inductive Logic and Probability*, vol. II, edited by R. C. Jeffrey, 263–293. Berkeley Calif.: University of Berkeley Press. Reprinted with Postscripts in D. Lewis, *Philosophical Papers*, vol. II, 83–132. Oxford: Oxford University Press 1986.
- . 1981. »Ordering Semantics and Premise Semantics for Counterfactuals«. *Journal of Philosophical Logic* 10:217–234.
- Moore, George E. 1942. »A Reply to My Critics«. In *The Philosophy of G. E. Moore*, edited by P. A. Schilpp, 535–677. La Salle Ill.: Open Court.
- Moss, Sarah. 2012. »On the Pragmatics of Counterfactuals«. *Noûs* 46:561–586.
- Mumford, Stephen. 1998. *Dispositions*. Oxford: Oxford University Press.
- Nozick, Robert. 1981. *Philosophical Explanations*. Oxford: Oxford University Press.
- Paul, L. A., and Ned Hall. 2003. *Causation: A User’s Guide*. Oxford: Oxford University Press.
- Pollock, John. 1976. »The »Possible Worlds« Analysis of Counterfactuals«. *Philosophical Studies* 29:469–476.
- Pust, Joel. 2000. *Intuitions as Evidence*. New York: Routledge.
- Reiss, Julian. 2009. »Counterfactuals, Thought Experiments, and Singular Causal Analysis in History«. *Philosophy of Science* 76:712–723.
- Roush, Sherrilyn. 2005. *Tracking Truth: Knowledge, Evidence, and Science*. New York: Oxford University Press.
- Shenoy, Prakash P. 1991. »On Spohn’s Rule for Revision of Beliefs«. *International Journal of Approximate Reasoning* 5:149–181.
- Sobel, Howard J. 1970. »Utilitarianisms: Simple and General«. *Inquiry* 13:394–449.
- Sosa, Ernest. 1999. »How to Defeat Opposition to Moore«. *Philosophical Perspectives* 13:141–154.
- Spohn, Wolfgang. 1988. »Ordinal Conditional Functions: A Dynamic Theory of Epistemic States«. In *Causation in Decision, Belief Change, and Statistics*, vol. II, edited by W. L. Harper and B. Skyrms, 105–134. Dordrecht: Kluwer.

- . 2012. *The Laws of Belief: Ranking Theory and its Philosophical Applications*. Oxford: Oxford University Press.
- . 2013a. »A Ranking-Theoretic Approach to Conditionals«. *Cognitive Science* 37:1074–1106.
- . 2013b. »50 Jahre Gettier: Reichen vielleicht«. In *Das Gettierproblem*, edited by G. Ernst and L. Marani, 179–198. Münster: mentis.
- . 2015. »Conditionals: A Unifying Ranking-Theoretic Perspective«. *Philosophers' Imprint* 15 (1): 1–30.
- Stalnaker, Robert C. 1968. »A Theory of Conditionals«. In *Studies in Logical Theory*, edited by N. Rescher (American Philosophical Quarterly Monograph Series 2), 98–112. Oxford: Blackwell.
- . 1981. »A Defense of Conditional Excluded Middle«. In *Ifs: Conditionals, Belief, Decision, Chance, and Time*, edited by W. Harper, R. Stalnaker, and G. Pearce, 87–104. Dordrecht: D. Reidel.
- Vineberg, Susan. 2000. »The Logical Status of Conditionalization and its Role in Confirmation«. In *Logic, Probability and Science*, edited by N. Shanks and R. B. Gardner (Poznań Studies in the Philosophy of the Science and Humanities 71), 77–94. Amsterdam: Rodopi.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.



## ZU DEN AUTORINNEN UND AUTOREN

- ANSGAR BECKERMANN ist Professor für Philosophie an der Universität Bielefeld Statt Flattersatz doch lieber Blocksatz?
- WOLFGANG BENKEWITZ ist Privatdozent für Philosophie an der Universität Konstanz
- BERND BULDT ist Professor für Philosophie und Department Chair an der Indiana University–Purdue University Fort Wayne Habe »(IPFW), USA« weggelassen; schien mir unnötig.
- RALF BUSSE ist Professor für Philosophie an der Johannes Gutenberg-Universität Mainz
- CHRISTOPH FEHIGE ist Professor für Praktische Philosophie an der Universität des Saarlandes
- WOLFGANG FREITAG ist Professor für Erkenntnistheorie und Wissenschaftstheorie am University College Freiburg und an der Albert-Ludwigs-Universität Freiburg
- GORDIAN HAAS ist Privatdozent für Philosophie an der Universität Bayreuth
- VOLKER HALBACH ist Professor für Philosophie an der Universität Oxford und Fellow am New College, Oxford
- FRANZ HUBER ist Assistant Professor an der Universität Toronto
- ANDREAS KEMMERLING ist Professor für Philosophie an der Universität Heidelberg
- MANFRED KUPFFER ist Wissenschaftlicher Angestellter an der Universität Konstanz
- HANNES LEITGEB ist Professor für Philosophie an der Ludwig-Maximilians-Universität München Munich Center for Mathematical Philosophy auch erwähnen??
- GODEHARD LINK ist Professor i.R. für Logik und Wissenschaftstheorie an der Ludwig-Maximilians-Universität München Munich Center for Mathematical Philosophy auch erwähnen??
- ARTHUR MERIN lehrt am Munich Center for Mathematical Philosophy der Ludwig-Maximilians-Universität München

THOMAS MÜLLER ist Professor für Theoretische Philosophie an der  
Universität Konstanz

JULIAN NIDA-RÜMELIN ist Professor für Philosophie an der Ludwig-  
Maximilians-Universität München

MARTINE NIDA-RÜMELIN ist Professorin an der Université de Fribourg

HANS ROTT ist Professor für Theoretische Philosophie an der Universität  
Regensburg

HOLGER STURM ist Professor für Analytische Philosophie an der  
Universität des Saarlandes

THOMAS EDE ZIMMERMANN ist Professor für Formale Semantik an der  
Johann Wolfgang Goethe-Universität Frankfurt am Main

ALEXANDRA ZINKE ist Wissenschaftliche Mitarbeiterin an der  
Universität Konstanz