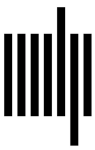
The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.



The MIT Press

3.1 Propositional and First-Order Logic

Florian Steinberger

Summary

This chapter addresses the question how (if at all) propositional logic (PL) and first-order logic (FOL) relate to epistemic rationality.[1] Rationality, it is often held, demands that our attitudes cohere in particular ways. Logic is often invoked as a source of such coherence requirements when it comes to belief: an ideally rational agent's beliefs are consistent and closed under logical consequence (i.e., the logical consequences of the agent's beliefs are also believed). However, this traditional picture has been challenged from various quarters. I begin by briefly reviewing the key concepts involved in PL and FOL. I then critically examine two distinct approaches to justifying logic-based requirements of rationality. The first lays down a set of desiderata codifying our intuitions and then seeks to formulate a principle articulating the link between logic and rational belief that satisfies them. The second starts by identifying our most fundamental epistemic aim and seeks to derive requirements of rationality based on their ability to promote this aim.

1. Logic and Rationality

Rational belief and correct belief are not the same thing. I may rationally believe a falsehood (for example, when the evidence available to me is misleading). If correctness for beliefs is truth, my belief may be correct though irrational (as when I engage in wishful thinking and, merely by a fluke, form a true belief). Part of what it means for beliefs to be rational is for them to cohere with the evidence. Another part is for them to cohere with each other.[2] Logic is often thought to be connected to the second sense of rationality. Logic is concerned with the relationships between the truth-values of different propositions. It tells us which propositions can and cannot be jointly true by dint of certain of their structural properties. For example, a chief concern of logic is logical consequence. If a proposition *C* is a logical consequence of a set of propositions, all of which I believe, it is, in a specific sense, impossible for my beliefs to be true without *C* also being true. Conversely, if I know *C* to be false, I thereby know that my antecedent beliefs cannot all be true. Relatedly, if my beliefs are inconsistent (and I recognize them to be so), then, merely by virtue of the logical structure of my beliefs' contents, I am in a position to know that I must be mistaken with respect to at least one of them—my representation of the world is not merely false, but there is no logically possible world that makes all of my beliefs true. Logical coherence, then, is typically a necessary (though generally insufficient) condition for the truth of one's belief. And while I may not be rationally criticizable for having incorrect beliefs, many maintain that if my beliefs fail to be logically coherent, I am less than ideally rational. It is in this sense that logic is thought to offer up requirements of rationality.[3]

The relation between logic and epistemic rationality relies on our assumption that the entities with which logic is concerned are also (or are systematically related to) the contents of our rationally evaluable mental states. I use "proposition" to designate the type of entity that constitutes both the premises and the conclusions of logical arguments, on the one hand, and the contents of propositional attitudes, on the other. I assume there is a type of entity capable of performing both of these roles. Furthermore, as my talk of structure in the foregoing makes plain, I assume that propositions are (logically) structured entities. That said, what I go on to say is compatible with other, nonpropositional accounts of truthbearers.

The aim of this chapter is to provide a critical survey of the discussion over whether logic really does give rise to norms of epistemic rationality. More specifically, I ask whether classical propositional logic (PL) and its extension, first-order logic (FOL; also called "[first-order] predicate logic"), can be said to do so. The force of restricting ourselves to these logics comes to this. Logic, we said, is concerned with possible truth-value distributions across propositions based on their logically relevant structural

features or *logical form*. The restriction to these particular logics is relevant in that different logics are sensitive to different structural features. The richer a logic's language, the more logical structure it is able to discern and hence the more rationally constraining it will be.

Another elementary but important observation is that PL and FOL are not themselves explicitly about rational belief formation and revision—their logical vocabulary contains no symbols that are naturally interpreted as designating doxastic states. In this they contrast with more sophisticated formalisms such as belief revision theory (Alchourrón, Gärdenfors, & Makinson, 1985), ranking theory (Spohn, 2012), and various Bayesian approaches whose express aim is to model the dynamics of belief management. The existence of these theories does not, however, render the present question nugatory. On the contrary, the fact that such theories incorporate implicit assumptions about the relationship between logic and rational belief renders the exploration of these foundational questions all the more urgent.

The chapter is structured as follows. The first two sections provide brief reviews of PL and FOL. Section 4 offers an overview of recent attempts at articulating the relationship between PL and FOL and rational belief as well as criticisms that have been leveled at such attempts. Section 5, finally, explores the relation of logical requirements of rationality to recent work in formal epistemology.

2. Propositional Logic

For our purposes, we can conceive of a logic as a formal language along with a semantic (model-theoretic) and a syntactic (proof-theoretic) relation of logical consequence. When it comes to formal languages, the symbols making up the alphabet are usually divided into three separate categories:

1. Descriptive symbols
2. Logical symbols
3. Auxiliary symbols

Descriptive symbols, intuitively, are the expressions of the language that have variable semantic values and so may be used to express truth-evaluable claims about the world. PL treats propositions as atoms, that is, PL is blind to the internal structural features of propositions. For example, it is insensitive to the distinction between propositions expressed by sentences involving only a unary predicate and singular term ("Rachel is rational") and sentences composed of binary predicates and two singular terms ("Rachel is taller than Steve"). Consequently, the only descriptive symbols contained in PL's language are formulas that serve to represent propositions (as opposed to expressions serving to express subpropositional content). PL distinguishes between *atomic* propositions, whose logical form cannot be further analyzed (using the resources of PL) and which are represented by atomic formulas, and *complex* ones, represented by complex formulas. Extending the chemistry metaphor, complex formulas are molecular compound formulas held together by the bonds that are logical connectives. The logical connectives are the following unary or binary operators (their approximate English cognates are given in brackets): "$\neg$" ("not"), "$\wedge$" ("and"), "$\vee$" ("or"), "$\rightarrow$" ("if . . . , then . . ."), and "$\leftrightarrow$" ("if and only if"). The only auxiliary symbols are parentheses, which are needed to avoid ambiguity among our expressions (e.g., to distinguish $(A \wedge (B \vee C))$ and $((A \wedge B) \vee C)$). In summary, the language contains

1. propositional variables: $\mathscr{P} = \{p_1, p_2, p_3, \ldots\}$ (in practice, we typically use "p," "q," "r," . . .);
2. logical connectives (or logical constants): $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$; and
3. auxiliary symbols: (,).

This alphabet constitutes the language of PL, $\mathscr{L}_{\text{PL}}$. Our next order of business is to specify a grammar for our language. Call a grammatically correct expression in our formal language a formula. We now define the set of formulas of PL, $\mathscr{F}_{\text{PL}}$, by means of the following inductive definition:

1. If $A \in \mathscr{P}$, then $A \in \mathscr{F}_{\text{PL}}$.
2. If $A \in \mathscr{F}_{\text{PL}}$, then so is $\neg A$.
3. If $A \in \mathscr{F}_{\text{PL}}$ and $B \in \mathscr{F}_{\text{PL}}$, then so are $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$, and $(A \leftrightarrow B)$.
4. Only such strings of symbols of our alphabet as can be constructed in accordance with rules 1 to 3 are formulas.

It is a direct consequence that every formula has a set of immediate subformulas (ISFs):

- If $A \in \mathscr{P}$, then $\text{ISF}(A) = \varnothing$.
- If $A = \neg B$, then $\text{ISF}(A) = \{B\}$.
- If $A = (B * C)$ (where $* \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$), then $\text{ISF}(A) = \{B, C\}$.

2.1 Semantics of PL

The semantics of PL mirrors its syntax in the following sense. The truth-value of a complex formula is determined by the truth-values of its immediate subformulas and

the logical connective governing it. Logical connectives are unary (negation) or binary truth functions from the truth-values of the subformula(s) to the truth-value of the complex formula to be evaluated. The atomic formulas, by contrast, are assigned their truth-values directly, by way of an interpretation. An interpretation in PL is a function $\mathfrak{I}$ whose domain is the set $\mathscr{P}$ of propositional variables and whose codomain is the set of {t, f} of truth-values:

$\mathfrak{I}: \mathscr{P} \rightarrow \{t, f\}$.

Our grammar guarantees that every complex formula can be broken down uniquely into subformulas of decreasing complexity until the process bottoms out in the propositional variables figuring within it. Consequently, the truth-value of a complex formula too will ultimately depend on the truth-values assigned to the propositional variables it contains and on the particular mode of composition of the formula. Our aim now is to use this insight to extend each interpretation function into a corresponding valuation function that assigns to each formula a truth-value. More precisely, every interpretation $\mathfrak{I}$ in PL determines a unique corresponding valuation $\mathfrak{V}_\mathfrak{I}$ (as can easily be proved).

A valuation in PL (relative to an interpretation $\mathfrak{I}$), then, is a function $\mathfrak{V}_\mathfrak{I}: \mathscr{F}_{PL} \rightarrow \{t, f\}$ that satisfies the following semantic rules:

1. $\mathfrak{V}_\mathfrak{I}(p_i) = t$ iff $\mathfrak{I}(p_i) = t$,
2. $\mathfrak{V}_\mathfrak{I}(\neg A) = t$ iff $\mathfrak{V}_\mathfrak{I}(A) = f$,
3. $\mathfrak{V}_\mathfrak{I}((A \wedge B)) = t$ iff $\mathfrak{V}_\mathfrak{I}(A) = t$ and $\mathfrak{V}_\mathfrak{I}(B) = t$,
4. $\mathfrak{V}_\mathfrak{I}((A \vee B)) = t$ iff $\mathfrak{V}_\mathfrak{I}(A) = t$ or $\mathfrak{V}_\mathfrak{I}(B) = t$,
5. $\mathfrak{V}_\mathfrak{I}((A \rightarrow B)) = t$ iff $\mathfrak{V}_\mathfrak{I}(A) = f$ or $\mathfrak{V}_\mathfrak{I}(B) = t$,
6. $\mathfrak{V}_\mathfrak{I}((A \leftrightarrow B)) = t$ iff $\mathfrak{V}_\mathfrak{I}(A) = \mathfrak{V}_\mathfrak{I}(B)$.

Clauses 1 to 6 are also called "semantic rules." Rule 1 says that propositional variables are to be evaluated via $\mathfrak{V}_\mathfrak{I}$ just as the underlying interpretation $\mathfrak{I}$ dictates. In rules 2 through 6, complex decomposable formulas are evaluated in conformity with the meanings of the logical connectives.

With this definition in place, we can now define the central concepts of consequence, validity, and consistency:

1. For all formulas $A_1, \ldots, A_n$ and B of $\mathscr{F}_{PL}$: B is a *logical consequence* in PL of $A_1, \ldots, A_n$ (in symbols: $A_1, \ldots, A_n \vDash_{PL} B$) iff for all interpretations $\mathfrak{I}$ with $\mathfrak{V}_\mathfrak{I}(A_1) = t, \ldots, \mathfrak{V}_\mathfrak{I}(A_n) = t$, it holds that $\mathfrak{V}_\mathfrak{I}(B) = t$.
2. An argument form $A_1, \ldots, A_n \therefore B$ of PL is *valid* in PL iff $A_1, \ldots, A_n \vDash_{PL} B$.
3. A set of formulas $\{A_1, \ldots, A_n\}$ is *consistent* in PL iff there exists an interpretation $\mathfrak{I}$ such that $\mathfrak{V}_\mathfrak{I}(A_1) = t, \ldots, \mathfrak{V}_\mathfrak{I}(A_n) = t$.

It is not hard to see that the concepts of logical consequence and consistency are intimately related:

- A set of formulas $\{A_1, \ldots, A_n, B\}$ is inconsistent (i.e., not consistent) in PL iff $A_1, \ldots, A_n \vDash_{PL} \neg B$.

To see this, observe that $A_1, \ldots, A_n \vDash_{PL} \neg B$ iff there exists no interpretation $\mathfrak{I}$ such that $\mathfrak{V}_\mathfrak{I}(A_i) = t$ for $i = 1, \ldots, n$ and $\mathfrak{V}_\mathfrak{I}(\neg B) = f$. But $\mathfrak{V}_\mathfrak{I}(\neg B) = f$ iff $\mathfrak{V}_\mathfrak{I}(B) = t$. Hence, $A_1, \ldots, A_n \vDash_{PL} \neg B$ iff there exists no interpretation $\mathfrak{I}$ such that $\mathfrak{V}_\mathfrak{I}(A_i) = t$ for $i = 1, \ldots, n$ and $\mathfrak{V}_\mathfrak{I}(B) = t$, but that is tantamount to saying that $\{A_1, \ldots, A_n, B\}$ is inconsistent.

2.2 Natural Deduction for PL

Let us now turn to the syntactic or proof-theoretic analogues of these notions. The proof system introduced here is a variant of natural deduction. It originates with Gerhard Gentzen (1934/1969) and was designed as an alternative to existing Frege–Hilbert axiomatic systems, with the aim of representing deductive reasoning more closely. "$\Gamma \vdash A$" is to be read as "The formula A is derivable in our system from the set of formulas Γ." For the most part, each connective is characterized in terms of a pair of rules. The introduction rules (. . . -I) state the conditions under which a proposition with the connective in question figuring as its main connective can be deduced; the elimination rules (. . . -E) state the deductive consequences of having derived such a proposition. The system, which we refer to as ND_{PL}, is constituted by the following deductive rules:

$$\wedge\text{-I}\ \frac{\Gamma_1 \vdash A \quad \Gamma_2 \vdash B}{\Gamma_1 \cup \Gamma_2 \vdash A \wedge B} \qquad \wedge\text{-E}\ \frac{\Gamma \vdash A_1 \wedge A_2}{\Gamma \vdash A_i}$$

$$\vee\text{-I}\ \frac{\Gamma \vdash A_i}{\Gamma \vdash A_1 \vee A_2} \qquad \vee\text{-E}\ \frac{\Gamma_1 \vdash A \vee B \quad \Gamma_2, A \vdash C \quad \Gamma_3, B \vdash C}{\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \vdash C}$$

$$\rightarrow\text{-I}\ \frac{\Gamma, A \vdash B}{\Gamma \vdash A \rightarrow B} \qquad \rightarrow\text{-E}\ \frac{\Gamma_1 \vdash A \quad \Gamma_2 \vdash A \rightarrow B}{\Gamma_1 \cup \Gamma_2 \vdash B}$$

$$\neg\text{-I}\ \frac{\Gamma, A \vdash \bot}{\Gamma \vdash \neg A} \qquad \neg\text{-I}\ \frac{\Gamma_1 \vdash A \quad \Gamma_2 \vdash \neg A}{\Gamma_1 \cup \Gamma_2 \vdash \bot}$$

$$\text{ECQ}\ \frac{\Gamma \vdash \bot}{\Gamma \vdash A} \qquad \text{DNE}\ \frac{\Gamma \vdash \neg\neg A}{\Gamma \vdash A}$$

where, in ($\wedge$-E) and ($\vee$-I), $i = 1$ or $i = 2$.

"$\bot$" denotes a contradictory proposition. "ECQ" abbreviates "ex contradictione [sequitur] quodlibet" (i.e., "from a contradiction anything follows"). "DNE" abbreviates "double negation elimination." We can then say that

- for all formulas $A_1, \ldots, A_n$ and B of $\mathscr{F}_{PL}$: B is *derivable from* $A_1, \ldots, A_n$ in ND_{PL} (in symbols: $A_1, \ldots, A_n \vdash_{ND_{PL}} B$) just in case there is a derivation employing only ND_{PL} rules and relying solely on premises $A_1, \ldots, A_n$.

ND_{PL} can be proved to be *sound*:

Soundness: If $A_1, \ldots, A_n \vdash_{\mathrm{ND_{PL}}} B$, then $A_1, \ldots, A_n \vDash_{\mathrm{PL}} B$,

and *complete* with respect to PL:

Completeness: If $A_1, \ldots, A_n \vDash_{\mathrm{PL}} B$, then $A_1, \ldots, A_n \vdash_{\mathrm{ND_{PL}}} B$.

3. First-Order Logic

The language of FOL is expressively richer than that of PL, which manifests itself in two principal respects: (1) FOL is sensitive to logically relevant subsentential structure, and (2) FOL is able to represent quantified sentences such as "All donkeys have soft noses" or "Some people annoy everybody." Since PL lacked the expressive resources to represent such sentences, it had to treat the propositions so expressed as unanalyzable. Consequently, PL is blind to the validity even of simple syllogisms such as

All donkeys have soft noses. Camillo is a donkey. Therefore, Camillo has a soft nose.

FOL allows us to capture not only arguments such as these but also propositions involving multiple generality, where several quantifiers interact, as in our previous example "Some people annoy everybody."

Let us turn to the vocabulary of the language of FOL, $\mathscr{L}_{\mathrm{FOL}}$:[4]

1. Descriptive symbols:
   (a) Countably many individual constants: $a_1, a_2, a_3, \ldots$
   (b) Countably many predicate symbols: $P^n_1, P^n_2, P^n_3, \ldots$ (for all arities $n \geq 1$)
2. Logical symbols:
   (a) Connectives: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$
   (b) Quantifiers: $\exists$ ("there is"/"some"), $\forall$ ("all"/"every")
   (c) Countably many individual variables: $x_1, x_2, x_3, \ldots$
3. Auxiliary symbols:
   (a) ( )
   (b) ,

Different applications of FOL will necessitate different choices of descriptive symbols. Each choice determines a specific language of first-order logic. What is common to all such languages are the logical symbols. The propositions expressed by "Camillo is a donkey" and "All donkeys have soft noses" can be represented by the formulas "$D(c)$," where "c" is a constant—comparable to a proper name—that denotes Camillo, "D" is the predicate "is a donkey," and "S" is the predicate "has a soft nose." The latter sentence can then be represented by "$\forall x\ (D(x) \rightarrow S(x))$," where "$\forall x \ldots$" is to be read as "for all $x, \ldots$"; roughly, the sentence can be paraphrased as "For all things x, if x is a donkey, then x has a soft nose."

In the following, I use lowercase letters "t," "c," "v" as meta-variables for singular terms, constants, and variables, respectively. Note that for all t, t is a singular term in $\mathscr{L}_{\mathrm{FOL}}$ iff t is an individual constant or an individual variable. We are now in a position to define the set of formulas of FOL, $\mathscr{F}_{\mathrm{FOL}}$:

If P^n is an n-ary predicate and $t_1, \ldots, t_n$ are singular terms, then $P^n(t_1, \ldots, t_n) \in \mathscr{F}_{\mathrm{FOL}}$. These are the atomic formulas of $\mathscr{L}_{\mathrm{FOL}}$.

If $A \in \mathscr{F}_{\mathrm{FOL}}$, then so is $\neg A$.

If $A \in \mathscr{F}_{\mathrm{FOL}}$ and $B \in \mathscr{F}_{\mathrm{FOL}}$, then so are $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$, and $(A \leftrightarrow B)$.

If $A \in \mathscr{F}_{\mathrm{FOL}}$ and v an individual variable, then $\exists vA \in \mathscr{F}_{\mathrm{FOL}}$.

If $A \in \mathscr{F}_{\mathrm{FOL}}$ and v an individual variable, then $\forall vA \in \mathscr{F}_{\mathrm{FOL}}$.

Only strings of symbols constructed in conformity with the clauses above qualify as well-formed formulas (wffs).

We can again define the notion of an immediate subformula in FOL:

- If A is atomic, then $\mathrm{ISF}(A) = \varnothing$.
- If $A = \neg B$, then $\mathrm{ISF}(A) = \{B\}$.
- If $A = (B \bullet C)$ (where $\bullet \in \{\wedge, \vee, \rightarrow, \leftrightarrow\}$), then $\mathrm{ISF}(A) = \{B, C\}$.
- If $A = \exists vB$, then $\mathrm{ISF}(A) = \{B\}$.
- If $A = \forall vB$, then $\mathrm{ISF}(A) = \{B\}$.

3.1 Semantics for FOL

The semantics for FOL takes a slightly different form due to the fact that we need an apparatus to assign semantic values to subsentential expressions. Moreover, we must specify a domain of objects for our quantifiers to range over.

- An *interpretation*, $\Im$, in FOL is a pair $\langle D, \varphi \rangle$ where
  1. D is a nonempty set of objects,
  2. φ is an interpretation function such that
  3. for all individual constants c, $\varphi(c) \in D$,
  4. for all n-ary predicates P^n, $\varphi(P^n) \subseteq D^n$, where D^n is the n-fold Cartesian product of D, that is, the set of n-tuples of elements of D.

Consider the atomic formula $P(x)$. Suppose we put $D = \mathbb{N}$ (where $\mathbb{N}$ is the set of natural numbers) and interpret P by letting $\varphi(P)$ be the set of prime numbers. The formula $P(x)$ cannot be said to be true or false, because we do not know what x refers to. We could prefix the formula with a quantificational expression as in $\exists x\ P(x)$, in which case the variable is said to be "bound" (as opposed to "free") and

the formula does obtain a truth-value: it is true because the set of prime numbers is nonempty. Alternatively, we could replace x by a constant c, which we might interpret by setting $\varphi(c) = 28$. The result, $P(c)$, again has a truth-value (it is false because 28 is no prime number). Individual variables thus function somewhat like demonstratives (e.g., "that"): without the benefit of appropriate contextual information, we cannot determine their referent. But here is the challenge. As in the case of PL, we want to explain the truth-value of an FOL formula as a function of the truth-values of its immediate subformulas. But to account for the truth-value of, say, $\exists x\, P(x)$, we must then first account for "the" truth-value of $P(x)$. To handle such cases, we need the notion of a variable assignment. A variable assignment σ relative to a domain D is a function that assigns to each variable an arbitrary referent in D. More precisely:

- A variable assignment σ under an interpretation $\Im = \langle D, \varphi \rangle$ is a function that assigns to every individual variable v an element d of D.

In our example, the question as to the truth-value of $\exists x\, P(x)$ (relative to the above interpretation) turns on the question whether there exists a σ such that $\sigma(x)$ is a prime number. That is, we effectively account for quantification by appeal to quantification over variable assignments in the metalanguage.

For convenience sake, let us merge our variable assignments with our interpretation function. That is, given an interpretation $\Im = \langle D, \varphi \rangle$, the interpretation function φ and any variable assignment σ (under $\Im$) can be folded into a single function, φ_σ, which now covers all singular terms—individual constants and individual variables:

- Let $\Im = \langle D, \varphi \rangle$ be an interpretation in FOL and σ a variable assignment under $\Im$; then φ_σ is defined as follows:
  1. for all individual constants c: $\varphi_\sigma(c) = \varphi(c)$, and
  2. for all individual variables v: $\varphi_\sigma(v) = \sigma(v)$.

We have now outlined the machinery needed for determining whether a formula—any formula, including those with free variables—is true or false relative to a variable assignment. Just as in the case of propositional logic, what we are after is a function that assigns to every formula a truth-value on the basis of an interpretation and a variable assignment under that interpretation. We will again use φ_σ to designate that function. So, $\varphi_\sigma(A)$ designates the truth-value of the formula A relative to the interpretation $\Im = \langle D, \varphi \rangle$ and given the variable assignment σ. It is important to be clear about the double role the function φ_σ is playing here. In the last paragraph, φ_σ played the role of a function that took singular terms—individual constants and individual variables—as inputs. Now we are charging it with the further task of taking formulas as inputs and outputting truth-values (relative to the interpretation and the variable assignment in question). In practice, the context will make it abundantly clear whether the function φ_σ is being applied to a singular term or to a formula.

With this we are in a position to define valuation functions for first-order logic:

- Let $\Im = \langle D, \varphi \rangle$ be an interpretation in first-order logic and σ a variable assignment under $\Im$. A *valuation* in first-order logic (relative to $\Im$ and σ) is a function φ_σ defined for all singular terms (as we explained in the previous section), which moreover assigns to every formula A of $\mathscr{L}_{\text{FOL}}$ of a given language of first-order logic the value t or f so as to satisfy the following semantic rules:
  1. $\varphi_\sigma(P^n(t_1, \ldots, t_n)) = \text{t}$ iff $\langle \varphi_\sigma(t_1), \ldots, \varphi_\sigma(t_n) \rangle \in \varphi(P^n)$,[5]
  2. $\varphi_\sigma(\neg A) = \text{t}$ iff $\varphi_\sigma(A) = \text{f}$,
  3. $\varphi_\sigma((A \wedge B)) = \text{t}$ iff $\varphi_\sigma(A) = \varphi_\sigma(B) = \text{t}$,
  4. $\varphi_\sigma((A \vee B)) = \text{t}$ iff $\varphi_\sigma(A) = \text{t}$ or $\varphi_\sigma(B) = \text{t}$,
  5. $\varphi_\sigma((A \rightarrow B)) = \text{t}$ iff $\varphi_\sigma(A) = \text{f}$ or $\varphi_\sigma(B) = \text{t}$,
  6. $\varphi_\sigma((A \leftrightarrow B)) = \text{t}$ iff $\varphi_\sigma(A) = \varphi_\sigma(B)$,
  7. $\varphi_\sigma(\forall vA) = \text{t}$ iff for all variable assignments σ' under $\Im$, if σ' is a v-variant of σ, then $\varphi'_\sigma(A) = \text{t}$,
  8. $\varphi_\sigma(\exists vA) = \text{t}$ iff there is a variable assignment σ' under $\Im$ such that σ' is a v-variant of σ and $\varphi'_\sigma(A) = \text{t}$.

If $\Im = \langle D, \varphi \rangle$ and $\varphi_\sigma(A) = \text{t}$, we say that A is *true* relative to the interpretation $\Im$ and the variable assignment σ under $\Im$ (similarly for falsity).

Of course, to complete our definition, we must still define the notion of a v-variant:

- Let σ, σ' be variable assignments under an interpretation $\Im$. Then σ' is a *v-variant* of σ iff for all individual variables $v' \neq v$ it is the case that $\sigma'(v') = \sigma(v')$.

Hence, if σ' is a v-variant of σ, then σ' agrees with σ on all individual variables except possibly with respect to v, where the two may (but need not) diverge. It follows that every variable assignment is a v-variant of itself (with respect to any individual variable v).

We are now in a position to define our key logical notions for FOL:

- For all formulas $A_1, \ldots, A_n, B$: B is a *logical consequence* of $A_1, \ldots, A_n$ ($A_1, \ldots, A_n \vDash_{\text{FOL}} B$) iff for all interpretations $\Im = \langle D, \varphi \rangle$ and all variable assignments σ under $\Im$, if $\varphi_\sigma(A_1) = \text{t}, \ldots, \varphi_\sigma(A_n) = \text{t}$, then $\varphi_\sigma(B) = \text{t}$.

- A set of formulas $\{A_1, \ldots, A_n\}$ is *consistent* in FOL iff there exists an interpretation $\Im = \langle D, \varphi \rangle$ and a variable assignment σ under $\Im$ such that $\varphi_\sigma(A_1) = \mathrm{t}, \ldots, \varphi_\sigma(A_n) = \mathrm{t}$.

3.2 Natural Deduction for FOL

Our system of natural deduction $\mathrm{ND}_{\mathrm{PL}}$ can be extended into a system for FOL, $\mathrm{ND}_{\mathrm{FOL}}$, by adding the following two pairs of rules for the quantifiers.

In the application of ∃-E, one must ensure that a does not occur in $\exists x\, A(x)$, in Γ_2, or in B.

$$\exists\text{-I}\ \frac{\Gamma \vdash A(t)}{\Gamma \vdash \exists x\, A(x)} \qquad \exists\text{-E}\ \frac{\Gamma_1 \vdash \exists x A(x) \quad \Gamma_2,\, A(a) \vdash B}{\Gamma_1 \cup \Gamma_2 \vdash B}$$

$$\forall\text{-I}\ \frac{\Gamma \vdash A(a)}{\Gamma \vdash \forall x A(x)} \qquad \forall\text{-E}\ \frac{\Gamma \vdash \forall x A(x)}{\Gamma \vdash A(t)}$$

In ∀-E, every free occurrence of x in A is replaced by t. In the application of ∀-I, one must ensure that a does not occur in Γ, and x must be uniformly substituted for a and must not be bound by any other quantifier in A. If a were allowed to occur among the hypotheses Γ upon which the conclusion depends, the following would be possible:

$$\forall\text{-I}\ \frac{A(t) \vdash A(t)}{A(t) \vdash \forall x\, A(x)}$$

Similar examples can readily be given for the remaining restrictions. As in the case of PL, FOL can be shown to be sound and complete with respect to $\mathrm{ND}_{\mathrm{FOL}}$.[6]

4. Logic and Rationality

Having thus reminded ourselves of the key concepts and features of PL and FOL, let us return to the supposed connection between logic and rationality we sketched in the introduction.[7] How exactly do the laws of logic express constraints on rational belief?

We do sometimes talk as if logic were itself a theory of rational belief. For instance, the rules of our natural deduction systems are typically referred to as "inference rules" and are read accordingly (e.g., "From A and B, infer $A \wedge B$"). We already noted above that such talk cannot be taken literally. According to Gilbert Harman (1984, 1986), our talk of "inference rules" and the underlying identification of logic and rational belief formation is simply a category mistake. Logic is one thing; a theory of epistemic rationality is quite another. Logic has an abstract subject matter; it concerns itself with certain properties and relations of (sets of) propositions. A theory of rationality, by contrast, aims to give a normative account of how we should form and revise our beliefs. Once we recognize the fundamental difference between these two theoretical enterprises, according to Harman, we realize that an explanatory gap separates them. Harman is doubtful that there is a substantive and informative story to be told as to how that gap might be closed.

But perhaps this is too quick. Even if deductive logic and a theory of rationality do not come to the same thing, there may still be an interesting normative connection between the two. Following John MacFarlane (2004), let us call a general principle that seeks to articulate such a relation between logic and norms governing belief, a *bridge principle*. Schematically, a bridge principle can be represented as follows:

(*) If $A_1, \ldots, A_n \vDash C$, then $N(\alpha(A_1), \ldots, \alpha(A_n), \beta(C))$.

Such principles take the form of a material conditional the antecedent of which expresses an instance of logical consequence and the consequent of which states a requirement of rationality governing the doxastic attitudes whose contents are so related: "N" stands for the governing norm, and "α" and "β" are (possibly distinct) doxastic attitudes borne toward the propositions A_i. On the basis of this blueprint, we can generate a number of distinct bridge principles by varying the following parameters:

- Doxastic attitude: Does the principle govern full beliefs or credences (i.e., degrees of belief)?
- Constraint on antecedent of the conditional: How (if at all) is the antecedent to be constrained? Does the norm cover all logical consequences? Or perhaps only those that are known, believed, obvious, and so on?
- Deontic operator: Which deontic modal operator features in the requirement expressed in the consequent? Is it a claim about what the agent *ought*, has *permission*, or has (defeasible) *reason* to "do"?
- Scope of operator: Typically, the principle's consequent itself takes the form of a conditional—call it the embedded conditional. Ought the deontic operator O take wide scope with respect to it (i.e., $O(A \rightarrow B)$)? Ought it to take narrow scope $(A \rightarrow O(B))$? Should it attach to both the consequent and the antecedent $(O(A) \rightarrow O(B))$? Or should we think of the normative claim as involving a primitive, undecomposable conditional operator $(O(B \mid A))$?
- Polarity: What is the polarity of the normative claim? Is it a positive demand that the agent have a particular attitude? Or is it a negative demand that the agent not have a particular attitude?
- Synchronic–diachronic: Are the principles synchronic, instructing us how the agent's doxastic state ought to

be at any given moment in time? Or are they requirements about how one's beliefs should evolve over time?

A few examples may be useful.

1. If $A_1, \ldots, A_n \vDash C$, then S has reason to believe C if S believes each of the A_i.
2. If $A_1, \ldots, A_n \vDash C$, then S's credences ought to be such that[8] $cr(C) \geq \sum_{1 \leq i \leq n} cr(A_i) - (n-1)$.
3. If S knows that $A_1, \ldots, A_n \vDash C$, then S ought not simultaneously disbelieve C and believe each of the A_i.
4. If S believes that $A_1, \ldots, A_n \vDash C$, then, if S is permitted to believe each of the A_i, S is permitted to believe C.

Principles 1 and 2 have unconstrained antecedents. In the first, the normative claim in the consequent features the reason-operator, which takes narrow scope with respect to the consequent of the embedded conditional and governs full belief. The second principle is familiar from probability logic (Adams, 1998) and has been advanced as a bridge principle by Field (2009, 2015). It governs credences, and ought effectively takes wide scope in the consequent. Principle 3 differs in that it is restricted to known consequences and of negative polarity. It too employs the ought-operator, which takes wide scope over the embedded conditional, although it governs full belief. Principle 4 is relativized to a nonfactive attitude. It features the permission-operator, which attaches both to the antecedent and to the consequent of the embedded conditional and also governs full beliefs.

For simplicity, I am assuming that all principles are to be understood as synchronic principles. That is, principle 1 is to be read as

(Synchronic) If S recognizes (at t) that $A_1, \ldots, A_n \vDash C$, then, if S believes each of the A_i at t, S has reason to believe C at t

as opposed to

(Diachronic) If $A_1, \ldots, A_n \vDash C$, then, if S believes each of the A_i at t and t slightly precedes t', S has reason to believe C at t' (and from t' onward).

If there is an interesting connection between logic and rationality, there should be a viable bridge principle that articulates it. But what counts as a viable bridge principle? We can distinguish two approaches. On one approach (Field, 2009, 2015; MacFarlane, 2004; Milne, 2009; Steinberger, 2017b), we lay down various criteria of adequacy for our bridge principles. The principles that perform sufficiently well against these criteria are our contenders. The second approach—which I consider in section 6—proceeds from more fundamental epistemological principles and seeks to derive a bridge principle from them (assuming such a principle is to be had).

5. What Makes for a Viable Bridge Principle?

Let us begin by considering the former approach. The adequacy criteria appealed to in the literature derive in part from Harman's objections to potential bridge principles and in part from MacFarlane (2004). We can summarize them as follows:

Belief Revision: Suppose I believe both A and $A \rightarrow B$ (as well as accepting modus ponens). The mere fact that I have these beliefs and that I recognize them to jointly entail B does not normatively compel any particular attitude toward B on my part. In particular, it is not generally the case that I ought to, or have permission to, believe B. After all, B may be at odds with my evidence, and so it would be unreasonable of me to follow modus ponens slavishly by, as it were, "adding B to my belief box." The rational course of "action," rather, when B is untenable, is for me to relinquish my belief in at least one of my antecedent beliefs A and $A \rightarrow B$ on account of their unpalatable implications. Belief revision seems to tell against narrow-scope principles, at least against strict versions of such principles involving ought and permission. Similarly, because of the reflexivity of the consequence relation, these principles would imply that one ought to (or is permitted to) believe anything one in fact believes, which clearly seems problematic. Pinder (2017) defends a reasons-based narrow-scope principle (namely, principle 1 above) against such objections.[9]

Excessive Demands: Principles whose antecedents are unrestricted pose excessive demands on agents whose resources of time, computational power, stamina, and so on are limited. Take an ought-based narrow-scope principle according to which one's beliefs ought to be closed under logical consequence. Anyone who believes the axioms of Peano–Dedekind arithmetic ought to believe every last one of its theorems, even if a theorem's shortest proof has more steps than there are particles in the universe. But if the logical *ought* implies *can* (in the sense of what agents even remotely like us can do), such principles must be rejected.[10]

Clutter Avoidance: A related worry is this. Any of the propositions I believe entails an infinite number of propositions that are of no significance to me whatsoever. Not only do I not care about, say, the disjunction "Vienna is the capital of Austria or pigs can fly" entailed by my belief that Vienna is the capital of Austria, but it would be positively irrational for me

to squander my meager cognitive resources on inferring trivial implications of my beliefs that are of no relevance to my goals.

Epistemic Paradoxes: Some maintain that there are various types of epistemic situations in which it is arguably not merely excusable for an agent to have logically incoherent beliefs but where such incoherence is permissible or even rationally mandated. The Preface Paradox arguably is a case in point (see Makinson, 1965). Here is a summary: suppose I am the author of a nontrivial nonfiction book. Having scrupulously checked the evidence for every one of my claims $A_1, \ldots, A_n$, I am highly confident (and believe) in each of them. I am also highly confident that at least one of my claims is fallacious. After all, I am fallible. Call the proposition that $\neg(A_1 \wedge \ldots \wedge A_n)$ the "preface proposition" (P). Clearly, my beliefs are inconsistent. Moreover, given reasonable assumptions, in believing P, I ought to disbelieve a straightforward logical consequence of my beliefs, namely, their conjunction $(A_1 \wedge \ldots \wedge A_n)$. The Preface Paradox presumably poses problems for all principles involving full belief (although credence-based principles may get around it).

The Strictness Test: At least when it comes to ordinary, readily recognizable logical implications leading to conclusions that the agent has reason to consider, the logical obligation should be strict: there is something amiss about an agent who endorses the premises and yet disbelieves the conclusion on account of stronger countervailing reasons (MacFarlane, 2004, p. 12).[11] The Strictness Test prima facie represents a strike against principles featuring the reason-operator, which countenances cases in which an agent believes the premises but disbelieves the conclusion on account of sufficient independent reasons for doing so.

The Priority Question: The attitudinal variants have a distinctive advantage when it comes to dealing with Excessive Demands worries. But relativizing one's logical obligations to, for example, one's believed or recognized logical consequences invites problems of its own, according to MacFarlane (2004): "We seek logical knowledge so that we will know how we ought to revise our beliefs: not just how we *will* be obligated to revise them when we acquire this logical knowledge, but how we are obligated to revise them even now, in our state of ignorance" (p. 12).

Logical Obtuseness: Suppose someone professes to believe A and B but refuses to take a stand on (neither believes nor disbelieves) the conjunction $A \wedge B$. Intuitively, such a person is liable to criticism. Whereas the weaker (positive) reason-based principles fail to live up to the Strictness Test, they do not commit the sin of logical obtuseness since one at least has reason to believe, or not disbelieve, $A \wedge B$. Not so for principles with negative polarity. So long as the agent does not actively disbelieve $A \wedge B$, our negative bridge principles find no fault with cases like these. If this intuition carries any weight, negative principles may prove to be ultimately too weak (at least on their own).

There is little consensus over which bridge principle fares best in light of these desiderata (for an overview, see Steinberger, 2017b). One important sticking point is how to deal with the Preface Paradox and similar cases. MacFarlane (2004) argues that we must simply resign ourselves to the existence of an irresolvable normative conflict between the demands placed on us by logic and other epistemic norms. Field (2009) endorses principle 2, which constrains credences. The principle allows for one's credences in a consequence jointly entailed by all the propositions (e.g., their conjunction) to be very low, despite one's having high credence in each of the individual propositions. The principle is compatible with a broadly Bayesian view of credences, according to which a rational agent's credence function is (or is extendible to) a probability function. Another important issue is the question of the permissible level of idealization of a requirement of rationality. If our aim is to formulate principles of ideal rationality, Clutter Avoidance and Excessive Demands may matter less. But how, then, do these principles relate to ordinary agents like you and me (Harman, 1986)?

I argue (Steinberger, 2017a, 2019) that the approach as a whole is marred by our failure to distinguish three importantly different ways in which logic might be normative:

Directives provide first-personal guidance in the process of doxastic deliberation.

Evaluations serve as objective third-personal standards or ideals for classifying acts or states into correct and incorrect ones.

Appraisals serve as the basis for our (equally third-personal) criticisms of our epistemic peers and so underwrite our attributions of praise and blame.

To illustrate, consider the following act-utilitarian principle:

AU You ought to act in such a way as to maximize net happiness.

The principle might be an apt evaluative norm in that it serves as a metric for what is to count as a right action. Yet the norm is often of little help to an agent trying to figure out what to do. Typically, it will not be transparent to her which of the actions available to her maximize happiness. The norm offers the agent little by way of guidance and therefore is not fit to play the role of a directive.[12] Similarly, if our agent violates the evaluative utilitarian norm, she may nevertheless not be liable to criticism. Despite having violated the norm, she may have acted reasonably in light of how the situation presented itself to her. Conversely, she might have acted recklessly and yet, out of sheer luck, complied with the norm. In both cases, our appraisals and our evaluations come apart, and this is largely because our appraisals are, while our evaluations need not be, sensitive to the agent's perspective.

I argue (Steinberger, 2019) that this tripartite distinction reveals that the actors in this debate talk past one another on account of conceiving of the normative role of logic differently. Another advantage of the proposed analysis is this. The adequacy criteria are in tension with one another, leading participants in the debate to make unprincipled choices about which desiderata to discount in order to make a case for their favored principle(s). The threefold distinction has the virtue of explaining why our adequacy criteria are inconsistent and points us toward a philosophically well-motivated resolution. The criteria are inconsistent because they are motivated by the aforementioned incompatible conceptions of the normative role of logic. Once we distinguish between directives, evaluations, and appraisals, we find that each normative role is naturally associated with its own set of adequacy criteria, which form consistent subsets of our desiderata. Hence, for each normative role, there is a separate, well-defined, and—it is hoped—more tractable question as to whether logic is normative in that sense.

Even if my analysis is correct, though, it remains to be seen which (if any) of these normative roles could give rise to principles of rationality.

6. Logic and Epistemic Utility Theory

Let us now turn to the second approach. On this view, norms of epistemic rationality are to be derived from fundamental epistemic aims or values. Several candidates could be in the running: truth, knowledge, understanding, and so on. Also, we may ask whether we should be monists or pluralists about epistemic value. Finally, given the value(s) at the center of our epistemology, what is its normative structure and how does it give rise to a theory of epistemic rationality? Rather than attempting a necessarily superficial survey of the options here, I propose to study a concrete proposal: epistemic utility theory (EUT). The prevalent version of EUT elects truth as its sole and fundamental epistemic value. What is more, the theory's value-theoretic superstructure is broadly consequentialist: any principle of epistemic rationality must earn its keep by promoting the aim of truth.

William James's famous slogan "Believe truth! Shun error!" serves as a good starting point.[13] While belief aims at the truth, our attempts at achieving it are subject to this double imperative. If my sole aim was the maximization of true belief without any concern for error, I might as well believe any proposition whatsoever. Conversely, if I was so cautious as to wish to avoid error at all costs, I would be best off suspending belief across the board. Our challenge, then, is to strike the optimal balance between maximizing true beliefs and avoiding false ones. EUT proposes to bring the tools of decision theory to bear on the problem.[14]

Let us consider the case of full belief. Suppose for simplicity that you are entertaining propositions belonging to a finite set $\mathscr{P}$. For each proposition in $\mathscr{P}$, the agent may believe (B), disbelieve (D), or suspend belief (S). Formally, we can represent the agent's various possible "choices" over $\mathscr{P}$ as belief functions $b\colon \mathscr{P} \to \{\mathrm{B}, \mathrm{S}, \mathrm{D}\}$. Our belief functions are to be assessed based on their accuracy. Our standard of accuracy is the truth-value of the proposition at a possible world, where a world is simply represented by means of a consistent valuation function $w\colon \mathscr{P} \to \{\mathrm{t}, \mathrm{f}\}$.

In analogy with decision theory, we can conceive of each choice of a doxastic attitude relative to a world as producing a certain epistemic utility, which is represented by a numerical value. Epistemic utility can be represented by a further function that associates such a value to any attitude–truth-value pair:

$$eu\colon \quad \{\mathrm{B}, \mathrm{D}, \mathrm{S}\} \times \{\mathrm{t}, \mathrm{f}\} \to (-\infty, \infty).$$

That is, for any proposition $A \in \mathscr{P}$, *eu* returns a score as a function of one's attitude toward it and the proposition's truth-value at the world. R is the score for getting it right, $-W$ is the penalty one incurs for getting it wrong, and suspending yields a neutral score. Hence,

$$eu(\mathrm{B}, \mathrm{t}) = eu(\mathrm{D}, \mathrm{f}) = R,$$
$$eu(\mathrm{S}, \mathrm{t}) = eu(\mathrm{S}, \mathrm{f}) = 0,$$
$$eu(\mathrm{B}, \mathrm{f}) = eu(\mathrm{D}, \mathrm{t}) = -W.$$

How are we to conceive of the relative values of R and W? There are three options:

- The *epistemic radical* values true belief to a higher degree than she disvalues false belief: $R > W$.
- The *epistemic centrist* values both to the same degree: $R = W$.
- The *epistemic conservative* disvalues false belief to a higher degree than she values true belief: $W > R$.

James believed one's stance toward the question to be a matter of intellectual temperament. There is, however, a prima facie case to be made for conservativeness. Suppose I flip a fair coin. Let p be the proposition that the coin lands heads. What attitude is it rational for me to adopt with respect to p? It is clear that in the absence of further information I should suspend. However, consider the following decision matrix:

p	$\neg p$	BB	BD	SS	BS
t	f	$R - W$	$2R$	0	$R + 0$
f	t	$-W + R$	$-2W$	0	$-W + 0$

(The remaining cases—DD, DB, SB, DS, SD—are strictly analogous and so can be omitted.) For the epistemic radical, the optimal choice is to believe p and disbelieve $\neg p$, or vice versa. But blindly going out on a limb in this way seems reckless. The centrist is no better off. She is indifferent between believing (or disbelieving) both p and its contradictory, believing p and disbelieving $\neg p$ (or vice versa), and suspending.[15] Conservatism, then, is our best option: the disvalue of believing falsely outstrips the value of believing truly. Since having contradictory beliefs always adds a net negative to one's score ($R - W$), one would be irrational to do so.

Having thus fixed the relative value of our rewards and penalties, we can determine the overall epistemic utility of a belief function at a world. Epistemic utility is generally assumed to be additive:[16] the overall epistemic utility of b, $EU(b)$, is

$$EU(b) = \sum_{A \in \mathcal{P}} eu(b(A), w(A)).$$

In general, of course, we are interested in the actual world and will do our best to choose the belief function with the greatest actual epistemic utility. However, we generally do not know which of our beliefs are accurate. The score of a belief function then hinges on facts about the world of which we are uncertain. What we can be certain of, though, is that a belief function that has less epistemic utility than another, *however* the world turns out to be, can be eliminated. It would be plainly irrational to adopt it. Not only is such a belief function bound to be suboptimal, but I might have appreciated its suboptimality even in the absence of any knowledge about the actual world. Decision theory captures this form of irrationality in terms of the notion of dominance:

- A belief function b is *strictly dominated* by a belief function b' iff $EU(b', w) > EU(b, w)$ for all worlds w.
- A belief function b is *weakly dominated* by a belief function b' iff $EU(b', w) \geq EU(b, w)$ for all worlds w and there exists a world w' such that $EU(b', w') > EU(b, w')$.

Two brief remarks. First, one option's being dominated by another makes choosing it irrational only if the dominating option is not itself dominated. Consider this. A genie grants me a (single) cash reward: for any number n I care to specify, I receive exactly \$$n$. But of course, any choice of n is bound to be dominated. That is not to say, though, that there is no rational response (see Pettigrew, 2016a). Second, note that a belief function b's being dominated by another (undominated) function b' only shows that it would be irrational to opt for b. It says nothing about which belief function one should adopt (in particular, it does not in general recommend adopting b').

Following Easwaran (2016), we can say that a belief function b is *strongly coherent* just in case it is not even weakly dominated and that b is *weakly coherent* just in case it is not strongly dominated. The following rationality requirement thus falls right out of EUT's central commitments:

Strong Coherence: One ought to have strongly coherent beliefs.

Central for our purposes is the question of how logical principles of rationality relate to Strong Coherence and, more generally, whether they can be justified in the context of EUT.

First, let us observe that Strong Coherence requires that one believe all logical truths:

Logical Truth: If S's belief function is strongly coherent, then S believes all logical truths (in $\mathcal{P}$).

The following table illustrates why:

	B	S	D
A	R	0	$-W$
$\neg A$	**$-W$**	**0**	***R***

Suppose A is a logical truth. It follows that the second line in bold can be ignored because there is no logically possible world at which A is false. The score of one's

belief function is thus determined on the basis of the first row of the table alone. Thus, any belief function b that suspends belief in, or disbelieves, a tautology will have less epistemic utility at every possible world than a belief function b' that agrees with b on all propositions aside from A, which it believes.

Next let us consider the following weak single-premise bridge principle. The principle is wide in scope and of negative polarity. Moreover, it is restricted to $\mathscr{P}$.

Single-Premise Closure: For all A, $C \in \mathscr{P}$, if $A \vDash C$, then S ought not both believe A and disbelieve C.

It turns out that that Strong Coherence entails Single-Premise Closure. Suppose belief function b violates Single-Premise Closure. That is, assume that A, $C \in \mathscr{P}$ and b believes A and disbelieves C. Then b is strictly dominated by the belief function b', which suspends on both propositions, as the following table illustrates (where, as above, the second line in bold, $\boldsymbol{w_2}$, can be eliminated):

$A \vDash C$	A	C	BD	SS
w_1	t	t	$R - W$	0
$\boldsymbol{w_2}$	**t**	**f**	$\boldsymbol{2R}$	**0**
w_3	f	t	$-2W$	0
w_4	f	f	$-W + R$	0

The question is whether analogous results are to be had for the following two familiar requirements:

Multiple-Premise Closure: For all $\{A_1, \ldots, A_n, C\} \subseteq \mathscr{P}$, if $A_1, \ldots, A_n \vDash C$, then S ought not both believe all of the A_i and disbelieve C.

Consistency: If $\{A_1, \ldots, A_n\}$ is a logically inconsistent set, then S ought not believe each member of the set.

At the risk of ruining the suspense: Strong Coherence entails neither Multiple-Premise Closure nor Consistency. To see this, a small detour is necessary. First, let us introduce the notion of expected epistemic utility. The expected epistemic utility of a belief function b is the probability-weighted sum of its epistemic utilities across all possible worlds:

$$EEU_P(b) = \sum_{w \in W} P(w) EU(b, w).$$

The probability function P can be thought of as the subject's evidence or at least as partially determined by her evidence. The probability function might be interpreted as a credence function, as an evidential probability function, or as representing objective chances, depending on one's conception of evidence.

It can be shown that it is a sufficient condition for a belief function to be strongly coherent that there be a regular probability function relative to which it maximizes expected utility.[17] What is more, it can be shown that a belief function b has maximal expected utility just in case there exists a probability function P such that for any proposition A,[18]

- $b(A) = B$ iff $1 \geq P(A) \geq \dfrac{W}{R+W}$,
- $b(A) = S$ iff $\dfrac{W}{R+W} \geq P(A) \geq \dfrac{R}{R+W}$,
- $b(A) = D$ iff $\dfrac{R}{R+W} \geq P(A) \geq 0$.

It is noteworthy that if P is interpreted as the agent's credence function, this result yields a version of the so-called Lockean thesis, according to which fully (rationally) believing is (in a sense to be made precise) having credence in excess of a threshold. The threshold is set by the appropriate ratio between R and W.[19]

The upshot is that belief functions can maximize expected utility (and hence be strongly coherent) while satisfying neither Multiple-Premise Closure nor Consistency. The Preface Paradox can again serve as an example. Suppose our author entertains the propositions $A_1, \ldots, A_n$, which compose the body of her book, and that she has rational uniform credences with respect to these propositions. That is, her credence function cr is (or is extendible to) a probability function, and she has the same high credence, say, $cr(A_i) = .9$, with respect to each of the propositions. Assuming that $.9 \geq \dfrac{W}{R+W}$, the belief function b that believes all of the A_i maximizes expected utility with respect to cr and so is strongly coherent. However, provided that n is sufficiently large and that the author entertains $A_1 \wedge \ldots \wedge A_n$ and its negation, we get $cr(A_1 \wedge \ldots \wedge A_n) \leq \dfrac{R}{R+W}$, and so b, if it is to maximize expected utility, disbelieves $A_1 \wedge \ldots \wedge A_n$ (and believes its negation). The belief function b is then strongly coherent and yet violates both Multiple-Premise Closure and Consistency. We can resist this conclusion only if we ensure that $W \geq (n-1)R$. Here n (we are assuming $n > 2$) is the number of propositions in question. EUT requires consistency and full closure just in case the disvalue of error exceeds the value of true belief by a factor of $n - 1$ (see Easwaran & Fitelson, 2015).

What are we to make of this? Here is an observation. Strong Coherence and Consistency are structurally very similar. Violations are, in both cases, a priori knowable indications that one's beliefs are less accurate than alternative ones. The difference comes down to a subtle

quantifier shift: if b is an incoherent belief function, then there exists a belief function b' that has at least equal epistemic utility across all worlds and outperforms it in some. That is,

$$\exists b'(\forall w\ EU(b', w) \geq EU(b, w) \wedge \exists w'\ EU(b', w') > EU(b, w')).$$

By contrast, if b is an inconsistent belief function, then, for every world w, there exists a belief function b' such that $EU(b', w) > EU(b, w)$, that is,

$$\forall w \exists b'\ EU(b', w) > EU(b, w).$$

According to EUT, the former violation spells irrationality, while the latter does not. But why should this minor difference be of such moment? If our aim really is the truth, then why should we content ourselves with maximizing epistemic utility in cases where, because our beliefs are inconsistent, some of them are bound to be false? Not only is maximizing expected epistemic utility no necessary condition for perfect accuracy, but it may sometimes positively discourage one from holding a true belief, because doing so would come with too high an opportunity cost: it would lower the expected epistemic utility of the belief function as a whole. Cases of such counterintuitive epistemic trade-offs have been the subject of much discussion.[20]

Many of these criticisms target the epistemic value theory undergirding EUT: that epistemic rationality is teleologically structured and that principles of rationality are justified only inasmuch as they further *expected* epistemic value (see Littlejohn, 2016).

Others question the assumed monism about epistemic value: is truth really the sole source of epistemic value? And even fellow monists may disagree about the nature of the fundamental aim. Why should truth be the North Star of epistemology as opposed to, for instance, knowledge?[21] Accidental true belief is not enough, they say. We want to get it right for the right reason. Finally, Leitgeb (2013, 2017) has argued for the possibility of providing a harmonious account of doxastic rationality according to which a probabilist approach to rational credence is compatible with a logical approach to rational full belief that incorporates full consistency and closure requirements.

Notes

1. Epistemic (or theoretical) rationality, here, is contrasted with practical rationality. Roughly, the former is concerned with what we ought to believe, and the latter is concerned with what we ought to do.

2. It might be thought that the two forms of coherence really are the same. Some, most notably Niko Kolodny (2007, 2008), maintain that there is no separate demand that beliefs should cohere. To the extent that beliefs should indeed cohere, this is because beliefs that enjoy evidential support cannot fail but cohere in the right way. I will set Kolodny's challenge aside for the purposes of this discussion. Conversely, it might be thought that one's evidence is constituted only by one's beliefs. If so, a belief is evidentially supported precisely when it coheres with the entire body of relevant beliefs.

3. It is a matter of some controversy whether requirements of rationality have genuine normative force. While I allow myself to use formulations such as "norms of epistemic rationality," I do not mean to prejudge the issue.

4. For ease of exposition, I do not include function symbols, nor am I including identity among the logical constants.

5. In the case of atomic formulas of the form $P^1(t)$, we simply identify—for uniformity's sake—the "1-tuple" $\langle d \rangle$ with d. Hence, the extensions of one-place predicates continue to be sets of elements of D and not of 1-tuples of such objects.

6. For a much more detailed exposition, the interested reader may consult the excellent collaborative open-source textbook by Magnus and Button (2019) or Enderton (2001).

7. Henceforth, I use "logic" to mean FOL unless I explicitly indicate otherwise. Similarly, "$\models$" designates the consequence relation of FOL.

8. Call $u(A)$ the uncertainty in A, defined as $u(A) = 1 - cr(A)$. We can then give the following tidier reformulation of the inequality: $u(C) \leq \sum_{1 \leq i \leq n} u(A_i)$.

9. In so doing, Pinder (2017) takes issue with an argument by Steinberger (2016) to the effect that appeals to the normativity of logic do not support a case for paraconsistent revisions of our logics (logics that reject the principle of explosion, according to which from an inconsistent set of propositions any proposition whatsoever logically follows).

10. It is for worries like these that Isaac Levi (2002) spoke of doxastic commitments, which do not demand the agent to believe but rather amount to something like a promise to believe. For similar issues in epistemic and doxastic logic, see chapter 5.1 by van Ditmarsch (this handbook).

11. MacFarlane takes inspiration from Broome (2000, p. 85).

12. One need not maintain that directives must always have transparent application conditions (see Srinivasan, 2015).

13. See Pettigrew (2016c) for a formal analysis of James's insights. For an alternative analysis, see Raidl and Spohn (2019).

14. The following exposition is inspired by Easwaran (2016) and Pettigrew (2017).

15. See also Easwaran and Fitelson (2015) as to why $W \leq R$ leads to counterintuitive consequences.

16. This assumption is not essential (see, e.g., Dorst, 2019).

17. A *regular* probability function, in the present context, is one that assigns every world a nonzero probability.

18. See Easwaran (2016) for details.

19. Dorst (2019) argues for a more sophisticated variable-threshold Lockean account, which he takes to provide a metaphysical reduction of full belief to credence. Easwaran (2016), by contrast, argues in favor of the primacy of full belief.

20. See Berker (2013), Carr (2017), Greaves (2013), and Littlejohn (2012, 2015) for criticisms along these lines as they pertain to EUT and epistemic consequentialism more broadly. See Pettigrew (2016b) for a response.

21. The go-to reference for much of the contemporary debate is Williamson (2000).

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*(2), 510–530.

Berker, S. (2013). Epistemic teleology and the separateness of propositions. *Philosophical Review*, *122*(3), 337–393.

Broome, J. (2000). Normative requirements. In J. Dancy (Ed.), *Normativity* (pp. 78–99). Oxford, England: Oxford University Press.

Carr, J. R. (2017). Epistemic utility theory and the aim of belief. *Philosophy and Phenomenological Research*, *95*(3), 511–534.

Dorst, K. (2019). Lockeans maximize expected accuracy. *Mind*, *128*(509), 175–211.

Easwaran, K. (2016). Dr. Truthlove, or, How I learned to stop worrying and love Bayesian probability. *Noûs*, *50*(4), 816–853.

Easwaran, K., & Fitelson, B. (2015). Accuracy, coherence, and evidence. In T. Szabo Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 5, pp. 61–96). Oxford, England: Oxford University Press.

Enderton, H. B. (2001). *A mathematical introduction to logic* (2nd ed.). San Diego, CA: Academic Press.

Field, H. (2009). What is the normative role of logic? *Proceedings of the Aristotelian Society*, *83*, 251–268.

Field, H. (2015). What is logical validity? In C. R. Caret & O. T. Hjortland (Eds.), *Foundations of logical consequence* (pp. 33–70). Oxford, England: Oxford University Press.

Gentzen, G. (1969). Investigations into logical deduction. In M. Szabo (Ed.), *The collected papers of Gerhard Gentzen* (pp. 68–128). Amsterdam, Netherlands: North Holland. (Original work published 1934)

Greaves, H. (2013). Epistemic decision theory. *Mind*, *122*(488), 915–952.

Harman, G. (1984). Logic and reasoning. *Synthese*, *60*(1), 107–127.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.

Kolodny, N. (2007). How does coherence matter? *Proceedings of the Aristotelian Society*, *107*(13), 229–263.

Kolodny, N. (2008). Why be disposed to be coherent? *Ethics*, *118*(3), 437–463.

Leitgeb, H. (2013). The stability theory of belief. *Philosophical Review*, *123*(2), 131–171.

Leitgeb, H. (2017). *The stability of belief: How rational belief coheres with probability*. Oxford, England: Oxford University Press.

Levi, I. (2002). Commitment and change of view. In J. L. Bermudez & A. Millar (Eds.), *Reason and nature: Essays in the theory of rationality* (pp. 209–32). Oxford, England: Clarendon Press.

Littlejohn, C. (2012). *Justification and the truth-connection*. Cambridge, England: Cambridge University Press.

Littlejohn, C. (2015). Who cares what you accurately believe? *Philosophical Perspectives*, *29*(1), 217–248.

Littlejohn, C. (2016). The right in the good: A defense of teleological non-consequentialism in epistemology. In K. Ahlstrom-Vij & J. Dunn (Eds.), *Epistemic consequentialism* (pp. 23–47). Oxford, England: Oxford University Press.

MacFarlane, J. (2004). In what sense (if any) is logic normative for thought? Retrieved from https://www.johnmacfarlane.net/normativity_of_logic.pdf

Magnus, P. D., & Button, T. (with additions by A. Loftis & R. Trueman, updated by A. Thomas-Bolduc & R. Zach). (2019). *For all x: Calgary: An introduction to formal logic*. Retrieved from https://forallx.openlogicproject.org/

Makinson, D. C. (1965). The paradox of the preface. *Analysis*, *25*(6), 205–207.

Milne, P. (2009). What is the normative role of logic? *Proceedings of the Aristotelian Society*, *83*(1), 269–298.

Pettigrew, R. (2016a). *Accuracy and the laws of credence*. Oxford, England: Oxford University Press.

Pettigrew, R. (2016b). Making things right: The true consequences of epistemic consequentialism. In K. Ahlstrom-Vij & J. Dunn (Eds.), *Epistemic consequentialism* (pp. 220–239). Oxford, England: Oxford University Press.

Pettigrew, R. (2016c). Jamesian epistemology formalised: An explication of 'The will to believe.' *Episteme*, *13*(3), 253–268.

Pettigrew, R. (2017). Epistemic utility and the normativity of logic. *Logos & Episteme*, *8*(4), 455–492.

Pinder, M. (2017). A normative argument against explosion. *Thought*, *6*(1), 61–70.

Raidl, E., & Spohn, W. (2020). An accuracy argument in favor of ranking theory. *Journal of Philosophical Logic*, *49*, 283–313.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford, England: Oxford University Press.

Srinivasan, A. (2015). Normativity without Cartesian privilege. *Philosophical Issues*, *25*(1), 273–299.

Steinberger, F. (2016). Explosion and the normativity of logic. *Mind*, *125*(498), 383–419.

Steinberger, F. (2017a). Consequence and normative guidance. *Philosophy and Phenomenological Research*, *98*(2), 306–328.

Steinberger, F. (2017b). The normative status of logic. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from https://plato.stanford.edu/archives/spr2017/entries/logic-normative/

Steinberger, F. (2019). Three ways in which logic might be normative. *Journal of Philosophy*, *116*(1), 5–31.

Williamson, T. (2000). *Knowledge and its limits*. Oxford, England: Oxford University Press.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455