Boulesic Logic, Deontic Logic and the Structure of a Perfectly Rational Will

Daniel Rönnedal

Abstract: In this paper, I will discuss boulesic and deontic logic and the relationship between these branches of logic. By 'boulesic logic,' or 'the logic of the will,' I mean a new kind of logic that deals with 'boulesic' concepts, expressions, sentences, arguments and systems. I will concentrate on two types of boulesic expression: 'individual x wants it to be the case that' and 'individual x accepts that it is the case that.' These expressions will be symbolised by two sentential operators that take individuals and sentences as arguments and give sentences as values. Deontic logic is a relatively well-established branch of logic. It deals with normative concepts, sentences, arguments and systems. In this paper, I will show how deontic logic can be grounded in boulesic logic. I will develop a set of semantic tableau systems that include boulesic and alethic operators, possibilist quantifiers and the identity predicate; I will then show how these systems can be augmented by a set of deontic operators. I use a kind of possible world semantics to explain the intended meaning of our formal systems. Intuitively, we can think of our semantics as a description of the structure of a perfectly rational will. I mention some interesting theorems that can be proved in our systems, including some versions

*   Stockholm University

   📝 Department of Philosophy, Stockholm University, Universitetsvägen 10 D, 106 91 Stockholm, Sweden

   ✉ daniel.ronnedal@philosophy.su.se

of the so-called *hypothetical imperative.* Finally, I show that all sys-tems that are described in this paper are sound and complete with respect to their semantics.

Keywords: Boulesic logic; deontic logic; modal logic; practical ration-ality; the hypothetical imperative; the logic of the will; semantic tab-leaux.

1. Introduction

In this paper, I will discuss boulesic and deontic logic and the relation-ship between these branches of logic. By 'boulesic logic' (from the Greek 'boulesis'), 'the logic of the will,' 'conative logic,' or 'volative logic' I mean a kind of logic that deals with 'boulesic' concepts, words, expressions, sen-tences, principles, arguments and systems. Boulesic logic is a new kind of logic. There are hints about such a logic in the literature, but few attempts to develop the basic idea in detail. The main results in this paper are there-fore entirely new.[1]

---

[1]   In the *Nicomachean Ethics* (Book VI and VII), Aristotle mentions a special kind of practical syllogism where the conclusion is a command or act. This suggests that the Greek philosopher might have envisioned some sort of 'practical logic.' Immanuel Kant discusses some principles that it is plausible to be included in a boulesic logic, for instance, the so-called *hypothetical imperative* (see below for more on this). Some similar principles were discussed already by various medieval thinkers (see Knuuttila 2004, 3.3). In 1926, Ernst Mally published the book *Grundgesetze des Sollens*, which is generally thought to contain the first published formal deontic system ever. Mally's book has the subtitle *Elemente der Logik des Willens* (Elements of the Logic of (the) Will), which indicates that he saw important connections between deontic logic and the logic of the will. In fact, he might have thought they are the same thing. There are some similarities between boulesic logic and 'intentional logic,' even though there are also many important differences. Later in the introduction, I will say more about this. For some information on intentional logic, see, for example, (Broersen 2011; Broersen, Dastani and van der Torre 2001; Cohen and Levesque 1990; Lorini and Herzig 2008; Marra and Klein 2015; and Semmling and Wansing 2008). Harry Gens-ler develops a logic of the will as a part of a kind of 'belief logic' that is based on imperative logic and extends ideas introduced by Hector-Neri Castañeda in several works (see Gensler 2002, Chapter 10 for more on this). See also (Bratman 1999).

Some examples of boulesic words include 'wanting,' 'willing,' 'accepting,' and 'consenting.' Many words and expressions in the vicinity might also be classified as boulesic, such as 'intending,' 'desiring,' 'rejecting,' 'loving,' and 'hating,' and—more generally—'having a pro-attitude.' A boulesic concept is a concept expressed by a boulesic word. A boulesic sentence is a sentence that (essentially) includes a boulesic word. A boulesic argument is an argument that (essentially) involves a boulesic sentence, and a boulesic system is a system that (essentially) includes various boulesic axioms and/or rules of inference.[2]

Here are some examples of boulesic sentences:

- John wants to win.
- Jennifer wants it to be the case that there will be peace.
- Sonny accepts the fact that he is never going to be a professional football player.

Here are some examples of boulesic principles (not necessarily valid):

- No one wants it to be the case that A and also wants it to be the case that not-A.
- If a perfectly rational individual x wants it to be the case that A and B is a necessary means to A, then x wants it to be the case that B.
- It is permitted that you perform this action only if everyone who is perfectly rational consents to the idea that you perform this action.

Here are some examples of boulesic arguments (not necessarily valid):

Boulesic logic, in the sense that I am using the term, is similar to deontic logic, intentional logic, imperative logic (if there is such a thing) and certain forms of epistemic and doxastic logics, and there seem to be important connections between, for instance, boulesic and deontic logic (for more on this, see sections 3.4, 3.5 and 4.3). However, there are also important differences between these branches, and there are, as far as I know, no systems of the kind introduced in this paper in the literature.

[2] These are not meant to be exact definitions. There are sentences that include boulesic words that are not boulesic—for example, 'The word "want" is an English word' and 'Jim believes that Greta wants to become a doctor'—and there are arguments that include boulesic sentences that are not boulesic, etc. Hence, I have added the qualification 'essentially.'

Argument 1

1. Henrietta wants the sun to shine.
2. Henrietta wants to go to the beach.
    Hence:
3. Henrietta wants the sun to shine and go to the beach. [From 1 and 2]

Argument 2

1. Mona wants to be happy.
    Thus:
2. If Mona is perfectly rational, she consents to the idea that she is happy. [From 1]

Argument 3

1. Every person in the class wants to pass the exam.
2. Sandra is a person in the class.
3. It is necessary that Sandra passes the exam only if she studies hard. Therefore:
4. If Sandra is perfectly rational, she wants to study hard. [From 1−3]

I will concentrate on two types of boulesic expression in this paper: 'individual c wants it to be the case that A' and 'individual c accepts that it is the case that A' or 'individual c consents to the state of affairs (the fact/the idea) that A.' These expressions will be symbolised by two sentential operators, $\mathcal{W}$ and $\mathcal{A}$ respectively, which take individual terms and sentences as arguments and give sentences as values. That is, in the formal sentence $\mathcal{W}_c B$, the constant c refers to an individual and B is any well-formed sentence (and similarly for $\mathcal{A}_c B$). In other words, '$\mathcal{W}_c B$' is read as 'c wants it to be the case that B' and '$\mathcal{A}_c B$' is read as 'c accepts that it is the case that B.'

Deontic logic is a relatively well-established branch of logic. It deals with normative concepts, sentences, arguments and systems. For introductions to this branch of logic, see for example (Åqvist 1987, 2002; Gabbay et al. 2013; and Hilpinen 1971, 1981). In this paper, I will show how deontic logic can be grounded in boulesic logic in a certain sense. I will develop a set of semantic tableau systems that include boulesic and alethic operators, possibilist quantifiers and the identity predicate, and I will show how these

systems can be augmented by a set of deontic operators. I use a kind of possible world semantics to explain the intended meaning of our formal systems. Intuitively, we can think of our semantics as a description of the structure of a perfectly rational will. I mention some interesting theorems that can be proved in our systems, including several versions of the so-called *hypothetical imperative* (if you want A and B is a necessary means to or condition for A, then you also want B insofar as you are rational; more about this below). Finally, I show that all systems that are described in this paper are sound and complete with respect to their semantics.

$\mathcal{W}$ takes any sentence as its argument. In $\mathcal{W}_c B$, B can be any well-formed formula whatsoever. So, it is possible to want anything, so to speak. B can be about the present time (I want you to be here now), about the future (She wants to become a doctor [some time in the future]) or about the past (I want [hope, desire, wish] that I made the right choice yesterday (Feldman 2004, 62)); it can be about a contingent state of affairs (She wants to buy a house) or a necessary state of affairs (He wants this mathematical theorem to be true); it can be about facts concerning nature (He wants the sun to shine tomorrow) or about various mental states (I want to feel happy); it can be about c (He wants to be happy) or about some other individual or individuals (She wants her children to be happy), and so on. According to our systems, it is meaningful to speak about wanting anything, and it is (logically) possible that any sentence of the form $\mathcal{W}_c B$ is true. It is even possible for someone to want something that is impossible. It seems reasonable to me that our systems allow this. Normally, if we want B, it is probably true that B is a contingent state of affairs that is in (or about) the future. However, this does not appear to be logically necessary. The same is true about $\mathcal{A}$ (acceptance). If we assume that c is perfectly rational (reasonable or wise), things are different. It seems plausible, for example, to claim that every perfectly rational (reasonable or wise) individual only wants something if it is possible; let us call this principle the Want-Can principle ($\mathcal{WC}$).[3] Nonetheless, there is a difference between what a perfectly

---

[3]   In every system that includes the tableau rule $T - \mathcal{WC}$ (Table 18), we can prove this principle. $T - \mathcal{WC}$ is valid in the class of all models that satisfy $C - \mathcal{WC}$ (see Section 3.3.5 for more on this). Note that the Want-Can principle does not entail

rational individual wants and what some arbitrary agent desires. It appears to be possible for someone who is not perfectly rational to want something that is impossible. In fact, there are probably actual examples of people with an inconsistent will. Yet, this should come as no surprise: not everyone is perfectly rational. [4]

When we say that someone wants something (or accepts something), we usually mean that she wants (accepts) this 'thing' in an *all-things-considered* sense in this paper; we do not necessarily mean that she wants it *in itself*. It is possible to want something as a means to something else and it is possible to want something in itself and it is possible to want something all-things-considered. Someone can, for example, want to study for the exam in an all-things-considered sense even though she does not want this in itself. She wants to study for the exam because she wants to pass the exam and she believes that she will pass the exam only if she studies for it. Studying is a means to an end. Moreover, it is possible for someone to want (or accept) *A* in an all-things-considered sense even if she does not like every aspect of *A* or every consequence of *A* and even if she has some desire (a prima facie desire) for not-*A*.

---

that everyone ought to want something only if it is possible. The latter thesis is independent of the Want-Can principle. The Want-Can principle is even compatible with the proposition that some individuals that are not perfectly rational ought to want some things that are impossible (in every system in this paper). Furthermore, it may still be reasonable to think about doing something impossible, to daydream about doing something impossible, and perhaps also to wish that something impossible be the case. But (merely) thinking, wishing, daydreaming, and so on is not the same thing as wanting.

[4]   Some philosophers seem to think that desires (and wants) are always future-oriented—that is, they think that if at time t, someone S desires that p be the case, then p is future relative to t. Wayne Sumner might be an example (Sumner 1996, 128–30; Sumner 2000). Other philosophers appear to reject this thesis (see, for example, Feldman 2004, 61–63). According to our systems, wants are not necessarily concerned with the future; it is possible that they are directed at the present or the past too, for instance. Still, I am inclined to believe that all 'genuine' wants are future-oriented for *perfectly rational individuals*—at least if we assume that facts about the past and the present are historically settled. Thus, I think our systems can do justice to Sumner's intuitions, at least to some extent.

For perfectly rational individuals wanting is a 'stronger' attitude than consenting (at least, in every class of models that satisfies $C-bD$ (Table 3)). If a perfectly rational individual wants A, she also consents to A (given that we accept $C-bD$). However, it is possible for a perfectly rational individual to consent to something that she does not want. A perfectly rational individual may, for example, consent to doing some boring chore in a particular situation even though doing this chore is not something she wants. Sometimes we can use the words 'agree,' 'allow,' 'approve,' 'condone' or 'tolerate' instead of 'accept' or 'consent.' Again, in this paper, we use 'accept' etc. in an *all-things-considered* sense. It is possible for someone to consent to A even though she objects to some aspects of or consequences of A. In every system in this paper, it is possible (even for a perfectly rational individual) to accept that A and (at the same time) to accept that not-A.

If c is not perfectly rational, almost nothing interesting at all follows logically from the fact that c wants something or accepts something.[5] For instance, if c is not perfectly rational and wants it to be the case that A, it does not follow that it is not the case that c also wants it to be the case that not-A. If c is not perfectly rational and c wants it to be the case that A, it does not follow that c also wants it to be the case that B even if B is a necessary condition for A and c knows this. It is probably not rational to want it to be the case that A and also want it to be the case that not-A, etc., but it does not seem to be *logically* impossible. And, in fact, according to our systems it is not. This is as it should be. We cannot prove the proposition that no one wants it to be the case that A and wants it to be the case that not-A in any system introduced in this paper. Nevertheless, in some systems (for instance every system that includes the rule $T-bD$; Table 14), we can prove the proposition that no one that is *perfectly rational* wants it to be the case that A and wants it to be the case that not-A (for more on this principle, see Section 3.3.3). These facts do not exclude the

---

[5] I say 'almost nothing interesting,' because we can still draw all usual conclusions from this fact. For example, if c wants it to be the case that A, then it is not the case that it is not the case that c wants it to be the case that A, etc. But we do not need a special boulesic logic to draw such conclusions.

possibility that there are *psychological laws* that make it *historically* impossible to combine different attitudes. It might, for instance, be *historically* necessary that no one (not even anyone that is imperfectly rational) wants it to be the case that *A* and also at the same time wants it to be the case that not-*A*. I am inclined to believe that this is not the case, at least not always. Yet, it is not *logically* impossible according to our systems: our systems do not rule out this possibility, and I believe that this is plausible.[6]

Boulesic logic, in the sense that I am using the term, is about the *rational* will, not only about what people *actually* want and accept and what follows from this. It attempts to give a description of the will of *perfectly rational*, *reasonable* or *wise* individuals. However, we can also use the systems in this paper to symbolise propositions about what individuals that are not perfectly rational want, accept, etc. Yet, the new, interesting boulesic laws that can be proved in our systems are not empirical, psychological laws; they do not 'describe' the contingent boulesic lives of actual people, even though the contents of these laws can include claims about what individuals that are not perfectly rational want, accept, etc.; they are 'laws of rationality.' Derivatively, boulesic logic also tells us how we must structure our wants if we are to be (perfectly) rational. Of course, exactly what it means to be 'perfectly rational' is something of an open question

---

6   It does not appear to be *logically* impossible for human beings to be perfectly rational, but it is very likely that no *actual* human being is perfectly rational; it might even be *historically* impossible for human beings to instantiate this property. Still, it seems to be the case that people are not totally irrational either. Normally, it appears to be true that if someone wants it to be the case that *A*, she does not also at the same time want it to be the case that not-*A*, etc. If this were not the case, we would probably not be able to ascribe wants to anyone. Furthermore, it is hard to think that an individual that often wanted it to be the case that *A* and also wanted it to be the case that not-*A* would survive for any long period: she would likely be stuck between alternatives like Buridan's ass and starve to death or be eaten by a predator. It is primarily a question for psychologists and other scientists, not for philosophers or logicians, to find out if there are any historically necessary laws of this kind.

and different answers to this question may lead to different boulesic systems. In boulesic logic, we can investigate the consequences of various ways of making the concept of perfect rationality more precise.

There are many ideas about what it means to be rational and many concepts of rationality. According to my view, the 'essence' of rationality is consistency; to be rational is to be consistent (in a wide sense). This involves, at least, consistency with oneself, but perhaps also consistency with the world and with other individuals. According to this view of rationality, it is very plausible to assume that one cannot be perfectly rational if one believes that A and that not-A, or if one wants it to be the case that A and that not-A. This is the core of the concept of rationality that I am trying to explicate in this paper. This concept of rationality should be distinguished from the concept that is often used in, for example, game theory and similar disciplines. In game theory, one usually assumes that every individual is an 'egoist' in the sense that she is only interested in satisfying her own preferences.[7] In game theory rationality is something like enlightened self-interest. Rationality-as-consistency should also be distinguished from a kind of 'rationality' that might be called 'pragmatic.' Suppose an eccentric (and very rich) neuroscientist were to offer you 10.000.000 pounds if you were able to believe in a contradiction and want this contradiction to be true. In this situation it might be plausible to say that it is 'rational' to believe in the contradiction and want it to be true, in some sense of 'rational.' We can call this kind of rationality 'pragmatic.' Suppose you were able to believe in the contradiction and want it to be true. Then we could say that you were pragmatically rational, but you would not be perfectly rational in our sense of this term. Rationality-as-consistency seems to me to be the most basic form, even though I do not deny that it might be fruitful to talk about rationality in other senses too. Much more could be

---

[7]   Such preferences can include otherregarding or altruistic preferences. Still, if an individual does not have any otherregarding or altruistic preferences, it is not rational for her to care about other people according to standard versions of game theory.

said about different theories of rationality, but this suffices for our purposes in the present paper.[8]

One could develop a boulesic logic as a kind of normal modal system and introduce a boulesic operator for every individual, where every operator functions as a normal modal operator. Let us call a boulesic system of this kind an 'ordinary boulesic system.' Nonetheless, there are certain problems with this approach. Firstly, in a system of this kind, it seems unlikely that W can be used to symbolise what actual persons want, for in such systems everyone wants everything that is necessary, no one has conflicting wants (without wanting absolutely everything), and every individual is such that if she wants A and B is a (logically) necessary means to A (i.e. if A entails B), then she also wants B. But all of this seems false. Secondly, if we restrict a boulesic logic to perfectly rational individuals (to avoid the first problem), we cannot speak about what persons that are not perfectly rational want, accept, etc., at least not in a natural way (we would have to use atomic formulas). Thirdly, we want to be able to symbolise such sentences as 'Everyone wants to be happy,' 'No perfectly rational individual wants it to be the case that both A and not-A,' and 'Everyone in the room wants to take the course.' Fourthly, in an ordinary boulesic system we implicitly have to assume that every perfectly rational individual is necessarily perfectly rational. It is not immediately obvious that this is the case. In our systems, we can investigate what follows if there are individuals that are only contingently perfectly rational. It is also meaningful, in principle, to ask whether we *should* be perfectly rational. If we can only speak about perfectly rational individuals, this does not seem possible. Fifthly, there are many arguments that are intuitively plausible (valid) that cannot be proved in ordinary boulesic systems that can be established in our systems. *Argument* 3 above is an example. The conclusion in this argument is derivable from the premises in every logic in this paper that includes the rule $T – MW$ (see Table 18). In Section 5.1, I will show this. Yet, *argument* 3 cannot be proved in any ordinary boulesic system, at least not without adding extra, implicit

---

[8]   For more on the concept of rationality, see, for example, (Broome 2013; Horty 2015; and Mele 2004).

premises. For these (and some other) reasons, I think the logics developed in this essay are preferable. In spite of this fact, they can be seen as an elaboration of the modal approach. All our systems include an ordinary modal part, with two kinds of modal operators for absolute and historical necessity and possibility.[9]

As I mentioned in footnote 1, there are some similarities between boulesic logic and 'intentional logic,' even though there are also many important differences. My formal approach is quite different from the formal approach found in the literature on intentional logic. I want to point out some differences. (i) The systems that are developed in the literature on intentional logic are often axiomatic. I use semantic tableaux. (ii) As far as I know, no intentional system includes a distinction between perfectly rational individuals and individuals that are not perfectly rational. Hence, the same logical principles hold for everyone in such systems. In my systems, it is not necessarily the case that individuals who are not perfectly rational satisfy the same principles that perfectly rational individuals satisfy. (iii) Intentions are sometimes required to be consistent while desires are not. In such intentional systems, it is logically impossible that some individual intends to do something and also intends not to do it. In my systems, it is always logically possible that someone (who is not perfectly rational) wants A at the same time that she wants not-A. (iv) Intentional systems are often at least as strong as so called classical modal systems. This means that an individual c intends (that) A iff (if and only if) she intends everything that is logically equivalent with A. In my systems, it is possible that an individual (who is not perfectly rational) wants A even though she does not want everything that is logically equivalent with A. (v) The intentional systems are not usually combined with predicate logic. Therefore, it is not possible to quantify over agents in such systems. They cannot be used to symbolize

---

[9]   For some introductions to ordinary (alethic) modal logic, see, for example, (Blackburn, de Rijke, and Venema 2001; Blackburn, van Benthem, and Wolter 2007; Chellas 1980; Fitting and Mendelsohn 1998; Gabbay 1976; Garson 2006; Kracht 1999; and Lewis and Langford 1932). For more on modal predicate logic, see, for example, (Barcan (Marcus) 1946; Carnap 1946; Garson 1984, 2006; Hintikka 1961; Hughes and Cresswell 1968; Parks 1976; and Priest 2008).

expressions of the following kind: 'Everyone who is such and such intends to do this or that,' and 'someone who is such and such intends to do this or that.' In all systems in this paper, we can symbolise expressions of the following kind: 'Everyone who is such and such wants it to be the case that,' and 'someone who is such and such wants it to be the case that.' (vi) Intentional systems are often weaker than so-called normal modal systems (at least for desires). This means that one cannot prove that if an individual desires (intends) (that) A then she also desires (intends) (that) B even though B is a necessary means to A. Similarly, in our systems we cannot prove that if an individual wants A then she also wants every necessary condition of A. However, in some systems we can show that if a *perfectly rational* individual wants A and A necessarily implies B, then she also wants B (see the discussion about hypothetical imperatives below). (vii) At least in some intentional systems the following propositions are valid: if an individual x intends that A and x intends that A implies B then x intends that B, and if it is valid that A implies B then if x intends that A then x intends that B. In our systems, it is not generally true that if an individual x wants it to be the case that A and x wants it to be the case that A implies B then x wants it to be the case that B; nor is it necessarily the case that x wants it to be the case that B if x wants it to be the case that A given that it is valid that A implies B. However, the latter principles do hold in our systems if they are restricted to *perfectly rational* individuals. (viii) As I am using the terms, intentions and wants are not the same thing. You can want someone else to do something, but you cannot intend someone else to do something. Intentions are directed towards (our own) actions, while it is possible to want all sorts of things. Wanting to do something and intending to do it might be the same thing, but it is not obvious that this is the case. If wanting to do something and intending to do it are not the same thing, wanting to do something probably often causes an intention to do it. So, we should make a distinction between intentions and wants. These are some of the most important differences.

Furthermore, I believe, that we should make a distinction between 'wants' and 'wishes' and not only between 'wants' and 'intentions.' Wanting something is not (necessarily) the same thing as wishing it were true, even

though we may sometimes use 'wish' instead of 'want.' Wishing something impossible were true might perhaps be possible even for a perfectly rational individual, even though it seems to be reasonable to claim that no perfectly rational individual wants impossible things (in an all-things-considered sense). It seems possible that the sentence 'I wish you were here' could be true (even in a situation where it is historically impossible for you to be here [now]), while 'I want you were here' is not even grammatical.

There are many good reasons to be interested in the results in this paper, both logical and philosophical. I cannot discuss all of these reasons: instead I will focus on one to illustrate the philosophical usefulness of our technical results.

The systems in this paper can be used to analyse and shed some light upon various interpretations of some philosophically interesting principles— for instance, the so-called *hypothetical imperative.* The notion of a hypothetical imperative was introduced by Immanuel Kant. In *Grundlegung zur Metaphysik der Sitten*, Kant characterises a hypothetical imperative in the following way:

> 'Who wills the end, wills (so far as reason has decisive influence on his actions) also the means which are indispensably necessary and in his power' and '"If I fully will the effect, I also will the action required for it" is analytic.' (Kant 1991, 45 [originally published in 1785]; English translation in Paton 1948, 80–81.)

Since Kant, there has been debate about how one should formulate the hypothetical imperative and how it should be interpreted, and about whether it is true or not. I will now show how one can use boulesic-deontic logic to distinguish between several different interpretations of this famous principle. I will consider eight of the most interesting readings and then show how they can be formalised in our systems. Finally, I will indicate which versions can be proved in various systems.[10]

1. It is universally necessary that, for every x, if x wants it to be the case that A and it is necessary that if A then B, then x wants it to be the

---

[10] Kant is usually taken to mean that moral principles are necessary and universal. Hence, I will interpret the hypothetical imperative in the same way.

case that B. $U\Pi x((\mathcal{W}_x A \land \Box(A \to B)) \to \mathcal{W}_x B)$. (Translation key. U: It is universally (absolutely) necessary that. Πx: For every [possible] x. $\mathcal{W}_x$: x wants it to be the case that. $\Box$: It is historically necessary that. $\to$: Material implication.)

2. It is universally necessary that, for every x, if x wants it to be the case that A and it is necessary that if A then B, then it ought to be the case that B. $U\Pi x((\mathcal{W}_x A \land \Box(A \to B)) \to OB)$. (Translation key. The same as before. O: It ought to be the case that.)

3. It is universally necessary that, for every x, if x wants it to be the case that A and it is necessary that if A then B, then x ought to want it to be the case that B. $U\Pi x((\mathcal{W}_x A \land \Box(A \to B)) \to O\mathcal{W}_x B)$.

4. It is universally necessary that, for every x, it ought to be the case that if x wants it to be the case that A and it is necessary that if A then B, then x wants it to be the case that B. $U\Pi x O((\mathcal{W}_x A \land \Box(A \to B)) \to \mathcal{W}_x B)$.

5. It is universally necessary that, for every x, it ought to be the case that if x wants it to be the case that A and it is necessary that if A then B, then B. $U\Pi x O((\mathcal{W}_x A \land \Box(A \to B)) \to B)$.

6. It is universally necessary that, for every x, if x is perfectly rational, then if x wants it to be the case that A and it is necessary that if A then B, then x wants it to be the case that B. $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to \mathcal{W}_x B))$.

7. It is universally necessary that, for every x, if x is perfectly rational, then if x wants it to be the case that A and it is necessary that if A then B, then it ought to be the case that B. $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to OB))$.

8. It is universally necessary that, for every x, if x is perfectly rational, then if x wants it to be the case that A and it is necessary that if A then B, then x ought to want it to be the case that B. $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to O\mathcal{W}_x B))$.

(1), (2), (3), (6), (7) and (8) are so-called 'narrow-scope' readings of the hypothetical imperative; (4) and (5) are so-called 'wide-scope' readings. Note how the consequent in the various interpretations varies. In (1), the

consequent is about an attitude; in (2), the consequent is about a norm; and in (3), it is about a norm about an attitude, etc. Readings (1)–(5) cannot be proved in any system in this paper: they are not valid in any class of model that we consider. In Section 5.1, I will show that the following 'instance' of (1) is not valid (in the class of all models): $\Pi x((\mathcal{W}_x Qx \land \Box(Qx \to Dx)) \to \mathcal{W}_x Dx)$, where Q and D are two monadic predicates. In our example, Qx stands for 'x quenches her thirst' and Dx stands for 'x drinks some water.' Since $\Pi x((\mathcal{W}_x Qx \land \Box(Qx \to Dx)) \to \mathcal{W}_x Dx)$ is not valid in the class of all models, it follows that not every instance of $U\Pi x((\mathcal{W}_x A \land \Box(A \to B)) \to \mathcal{W}_x B)$ is valid in the class of all models. Nevertheless, (6)–(8) can be deduced in some systems: (6) is provable in any boulesic (or boulesic-deontic) system that includes the rule $T - M\mathcal{W}$; (7) and (8) cannot be established in any pure boulesic system: we need a boulesic-deontic system; (7) can be derived in every boulesic-deontic system that includes the rules $T - \mathcal{W}O$, $T - H\mathcal{W}$ and $T - M\mathcal{W}$ (I will establish this in Section 5.1); (8) can be proved in every boulesic-deontic system that includes the rules $T - \mathcal{W}O$, $T - H\mathcal{W}$, $T - M\mathcal{W}$, $T - a4$, $T - b4$ and $T - FTR$. (For more on these rules, see Section 4.2.)

It seems to me that interpretation (6) comes very close to Kant's own reading of the hypothetical imperative. If this is correct, we can prove that Kant was right: 'Who wills the end, wills (so far as reason has decisive influence on his actions) also the means which are indispensably necessary…' is, in this sense, a necessary universal truth in some models. In Section 3.6, I will discuss a semantic argument that shows that (6) is valid in the class of all models that satisfy $C - M\mathcal{W}$ (see Table 5).[11]

---

[11]   For more information about the hypothetical imperative, see, for example, (Bedke 2009; Broome 1999; Brunero 2010; Downie 1984; Feldman 1986, Chapter 5; Foot 1972; Gensler 1985; Greenspan 1975; Harsanyi 1958; Hill 1973, 1989; Korsgaard 2008; Marshall 1982; Shaver 2006; Schroeder 2004, 2005, 2009, 2015; Wallace 2001; and Way 2010). There are at least two interesting questions about the interpretation of the hypothetical imperative. (1) What does (or should) the 'consequent' of the hypothetical imperative say (is it a claim about an attitude, a norm about what ought to be or about what we ought to do, or a norm about an attitude)? Kant seems to think that the consequent is about an attitude, about willing. But it has also been suggested that the consequent is a norm about what we ought to do or

The discussion of the hypothetical imperative above clearly shows, I believe, that the systems introduced in this paper are philosophically interesting. In conclusion, the topic of this article is both philosophically and logically well motivated.[12]

---

a norm about an attitude. Hill (1973), for instance, suggests that the consequent might be about what 'we' *ought to want*, and Marshall (1982) and Schroeder (2004) suggest that it might be about what 'we' *ought to do*. (2) Should the imperative be given a wide-scope or a narrow-scope interpretation? There is a debate about what Kant meant. Many philosophers seem to prefer a wide-scope reading of hypothetical imperatives, see e.g. (Hill 1973; Gensler 1985; Wallace 2001; Broome 1999, 2001; Greenspan 1975). But some have also argued for a narrow-scope reading, see e.g. (Schroeder 2004, 2005, 2009). According to Schroeder, Kant should be interpreted as a narrow-scoper. I am inclined to believe that this is in fact a better interpretation of Kant's position. Whether or not this view is correct, it is a nice feature of the systems in this paper that we can clearly distinguish between these different readings.

[12]   There are many other good reasons to be interested in the systems in this paper and also some potential problems with the whole project. I cannot discuss every interesting philosophical issue that is related to the topics of this paper. However, I would like to briefly mention a potential problem that was raised by an anonymous reviewer. According to this reviewer it may in principle be interesting to devote some attention to specific logical/inferential features of sentences which speak about willing/wanting, but there is no real need for a comprehensive logical theory of expressions of this kind. It is, of course, possible that this view is correct, but is seems highly problematic to me. If it is interesting to devote some attention to specific logical/inferential features of sentences which speak about willing/wanting, it seems to me that it must also be interesting to try to develop a comprehensive logical theory of expressions of this kind. In general, it is more interesting to have a comprehensive theory of some 'phenomenon' than just a piecemeal grasp of some unconnected truths (and almost everyone, I think, agrees with this view). Consider a similar argument. 'Though in principle it may be interesting to devote some attention to specific logical/inferential features of sentences which speak about propositional (truth-functional) truths there is no real need for a comprehensive logical theory of expressions of this kind. For example, we do not need sound and complete systems of propositional logic. There is no point in trying to construct axiomatic systems or tableau systems of propositional logic. It is enough if we study the law of non-contradiction, the law of identity, the law of excluded middle, etc.' This argument is obviously highly problematic and I believe few people would be convinced by it. When philosophers and logicians started to study propositional logic systematically, constructed sound and complete systems, and proved that they were sound and

The paper is divided into seven main sections. Section 2 deals with the syntax and Section 3 with the semantics of our systems. In Section 4, I describe the proof theory of our logics, while Section 5 includes some examples of theorems. Section 6 contains soundness and completeness proofs for every system. Finally, Section 7 includes a short conclusion and summary.

2. Syntax

2.1. Alphabet

Terms
(i) A set of variables x_1, x_2, x_3, . . .
(ii) A set of constants (rigid designators) k_{d_1}, k_{d_2}, k_{d_3}, . . .

Predicates
(iii) For every natural number $n > 0$, n-place predicate symbols P_n^1, P_n^2, P_n^3, . . .
(iv) The monadic existence predicate E, and the monadic rationality predicate R.
(v) The dyadic identity predicate (necessary identity) $=$.

Connectives
(vi) The primitive truth-functional connectives $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (material implication) and $\leftrightarrow$ (material equivalence).

---

complete, this was an extremely important development in the history of logic. Why should there be any difference if one talks about 'logical/inferential features of sentences which speak about willing/wanting'? The reviewer might be right that we do not 'need' boulesic logic in some senses of this word. For example, we do not need it to survive or for society to go on functioning. But, then again, we do not need *any* comprehensive logical theory at all for these purposes. The fact that we do not 'need' boulesic logic in some senses of the term 'need,' doesn't show that the topic of my paper isn't interesting. And it certainly does not follow that the project is not worth the effort. In conclusion, this potential problem does not strike me as particularly serious.

Operators

(vii) The alethic operators U (universal necessity), M (universal possibility), $\square$ (historical necessity) and $\diamond$ (historical possibility).[13]

(viii) The deontic operators O (Ought) and P (Permission).

(ix) The boulesic operators $\mathcal{W}$ (Want) and $\mathcal{A}$ (Accept).

Quantifiers

(x) The (possibilist) quantifiers Π (For all) and Σ (For some).

Parentheses

(xi) The brackets) and (.

I will use x, y and $z \ldots$ for arbitrary variables, a, b, $c \ldots$ for arbitrary constants, and s and t for arbitrary terms (with or without primes or subscripts). For more on the set of constants, see Section 3.1. I will use F_n, G_n, $H_n, \ldots$ for arbitrary n-place predicates and I will omit the subscript if it can be read off from the context.

Π and Σ are substitutional, 'possibilist' quantifiers because the domain is the same in every possible world and every object in the domain has a name (Section 3). Thus, in effect, they vary over every object in the domain.

2.2. Languages

I will consider two languages in this paper. The first, $L1$, does not include the deontic operators; the second, $L2$, does. $L1$ is constructed from clauses (i)–(viii) and (x), and $L2$ from clauses (i)–(x) below.

(i) Any constant or variable is a term.

(ii) If $t_1, \ldots, t_n$ are any terms and P is any n-place predicate, $Pt_1 \ldots t_n$ is an atomic formula.

---

[13] U and M are standard universal modalities (see almost any introduction to modal logic). For more on the concepts of historical necessity and possibility, see, for example, (Åqvist and Hoepelman 1981; and Chellas 1969). In this paper, I will not try to combine boulesic logic and temporal logic since I want to keep things relatively simple. However, it is in principle possible to combine the systems in this paper with various tense systems. I hope to do this in future work (see the conclusion in Section 7).

(iii) If t is a term, Et ('t exists') is an atomic formula and Rt ('t is perfectly rational') is an atomic formula.

(iv) If s and t are terms, then $s = t$ ('s is identical with t') is an atomic formula.

(v) If A and B are formulas, so are $\neg A$, $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$ and $(A \leftrightarrow B)$.

(vi) If A is a formula, so are UA ('it is universally [or absolutely] necessary that A'), MA ('it is universally [or absolutely] possible that A'), $\Box A$ ('it is [historically] necessary that A') and $\Diamond A$ ('it is [historically] possible that A').

(vii) If B is any formula and t is any term, then $\mathcal{W}_t B$ ('t wants it to be the case that B') and $\mathcal{A}_t B$ ('t accepts that it is the case that B') are formulas.

(viii) If A is any formula and x is any variable, then $\Pi x A$ ('for every [possible] x: A') and $\Sigma x A$ ('for some [possible] x: A') are formulas.

(ix) If A is a formula, then OA ('it ought to be the case that A') and PA ('it is permitted that A') are formulas.

(x) Nothing else is a formula.

A, B, C . . . stand for arbitrary formulas, and Γ, Φ . . . for sets of formulas. The concepts of bound and free variables, and open and closed formulas, are defined in the usual way. $(A)[t/x]$ is the formula obtained by substituting t for every free occurrence of x in A. The definition is standard. Brackets around formulas are usually dropped if the result is not ambiguous.

2.3. Definitions

It is possible to introduce some new symbols into our languages by definitions. If we do that, the new symbols should be treated as pure metalogical abbreviations and we should not read anything more into the definitions. Here are some examples:

Deontic operators. $FA =_{df} \neg PA$. $KA =_{df} (PA \wedge P\neg A)$. $NA =_{df} \neg KA$.

Actualist quantifiers. $\forall x A =_{df} \Pi x(Ex \rightarrow A)$ and $\exists x A =_{df} \Sigma x(Ex \wedge A)$.

O and *P* are not included in *L1*; in *L2* they are treated as primitive symbols. In some (but not all) systems that I will describe, *O*, *P* and *F* are 'definable' in terms of the boulesic operators in the sense that we can prove that the following equivalences are logically true: $OB \leftrightarrow \Pi x(Rx \to \mathcal{W}_x B)$ ('It ought to be the case that *B* iff everyone who is perfectly rational wants it to be the case that *B*'); $PB \leftrightarrow \Pi x(Rx \to \mathcal{A}_x B)$ ('It is permitted that *B* iff everyone who is perfectly rational accepts that [consents to the state of affairs that] it is the case that *B*'); and $FB \leftrightarrow \Pi x(Rx \to \mathcal{W}_x \neg B)$ ('It is not permitted that *B* iff everyone who is perfectly rational wants it to be the case that not-*B*'). However, since it is not immediately obvious that these equivalences *should* hold, and since we want to know which assumptions we must make to be able to prove them, we do not introduce the deontic operators through definitions in this paper. Furthermore, when I say that *O*, *P* and *F* are 'definable' in terms of the boulesic operators, this should not be taken to imply that, for example, '*OB*' has the same meaning as '$\Pi x(Rx \to \mathcal{W}_x B)$' or that '*OB*' can be replaced by '$\Pi x(Rx \to \mathcal{W}_x B)$' in every context. (For more on this, see Section 4.2.12 and Table 30.)[14]

---

[14] One possible objection against these equivalences is that they seem to presuppose an extreme view of rationality according to which all rational agents should have essentially the same wishes. It is true that in the systems where we can prove the equivalences, all perfectly rational individuals want the same things (see Sections 3.3.7, 4.2.8 and 4.2.12). If *a* is perfectly rational and *b* is perfectly rational, then it is not the case that *a* wants *C* and *b* wants not-*C* in those systems (at least, if we also assume that they include, for example, $T - bD$ (Table 14)). But is it not the case that this leaves no scope for legitimate conflicts of interests such as two businesspeople each wishing their own company to take market shares from that of the other, or two fathers both wishing that their own child wins a competition? I cannot discuss this argument in detail in this paper, but I want to make the following remarks. (1) Even in systems where we can prove the equivalences, it is possible that an individual *a* wants *C* and another individual *b* wants not-*C* (given that not both *a* and *b* are perfectly rational). (2) Even in systems where we can prove the equivalences, it is possible that it is permitted (and even obligatory) for some individual *a* to want *C* and for some other individual *b* to want not-*C*. (3) Even in systems where we can prove the equivalences and where both *a* and *b* are perfectly rational, it is possible that *a* wants to do 'everything' *a* can to win and that *b* wants to do 'everything' *b* can to win and that both *a* and *b* want 'the best man' to win, even

3. Semantics

3.1. Models

Definition 1 *A model $\mathcal{M}$ is a relational structure $\langle D, W, \mathfrak{R}, \mathfrak{A}, v \rangle$, where D is a non-empty set of individuals (the domain), W is a non-empty set of possible worlds, $\mathfrak{R}$ is a binary alethic accessibility relation ($\mathfrak{R}$ is a subset of $W \times W$), $\mathfrak{A}$ is a ternary boulesic accessibility relation ($\mathfrak{A}$ is a subset of $D \times W \times W$), and v is an interpretation function.*

A supplemented model $\mathcal{M}_S$ is a relational structure $\langle D, W, \mathfrak{R}, \mathfrak{S}, \mathfrak{A}, v \rangle$, where D, W, $\mathfrak{R}$, $\mathfrak{A}$ and v are as in an ordinary model, and $\mathfrak{S}$ is a dyadic deontic accessibility relation ($\mathfrak{S}$ is a subset of $W \times W$).

$\mathfrak{R}$ 'corresponds' to the alethic operators $\square$ and $\diamondsuit$, $\mathfrak{S}$ to the deontic operators O and P, and $\mathfrak{A}$ to the boulesic operators $\mathcal{W}$ and $\mathcal{A}$. Informally, $\mathfrak{R}\omega\omega'$ says that the possible world ω' is alethically (historically) accessible from the possible world ω, $\mathfrak{S}\omega\omega'$ says that the possible world ω' is deontically accessible from the possible world ω, and $\mathfrak{A}\delta\omega\omega'$ says that the possible world ω' is acceptable to the individual δ in (or relative to) the possible world ω, or that δ accepts ω' in (or relative to) ω.

In Section 3.4, we will see how $\mathfrak{S}$ and $\mathfrak{A}$ can be defined, and we will explore the consequences of these definitions.

The valuation function v assigns every constant c an element $v(c)$ of D, and every possible world ω in W and an n-place predicate P a subset $v_{\omega}(P)$ (the extension of P in ω) of D^n. In other words, $v_{\omega}(P)$ is the set of n-tuples that satisfy P in the world ω. Hence, every constant is a kind of rigid designator: it refers to the same individual in every possible world. Nonetheless, the extension of a predicate may change from world to world and it may be empty in a world. Let $\mathcal{M}$ be an ordinary or supplemented model. Then the

---

though it is not possible that both a and b win. (4) All systems I discuss are compatible with the proposition that it is possible for perfectly rational individuals to *wish* for incompatible things even though it is not possible for them to *want* incompatible things. So, it is not the case that those systems leave *no* scope for legitimate conflicts of interests. Whether or not they leave *enough* scope is, of course, debatable. Personally, I am inclined to believe that they *do* leave enough scope.

language of $\mathcal{M}$, $\mathcal{L}(\mathcal{M})$, is obtained by adding a constant k_d, such that $v(k_d)$ = d, to the language for every member $d \in D$. Hence, every object in the domain of a model has at least one name in our language, but several different constants may refer to one and the same object.

The predicate R has a special interpretation in our systems. 'Rc' says that c is *perfectly rational*, *perfectly reasonable* or *perfectly wise*. If $v(c)$ is in the extension of R at the possible world ω, this means that $v(c)$ is perfectly rational, reasonable or wise in ω. Exactly what this means will depend on the conditions we impose on the boulesic accessibility relation $\mathfrak{A}$ (Section 3.3). R functions as an ordinary predicate. Hence, an individual δ may be in R's extension in one possible world even though δ is not in R's extension in every possible world. Accordingly, the fact that an individual δ is perfectly rational, reasonable or wise in a possible world does not *entail* that δ is perfectly rational, reasonable or wise in *every* possible world. In Section 3.3.8, we will see what happens if we add the extra assumption that every perfectly rational individual is necessarily perfectly rational (the semantic condition C–UR guarantees that this is the case: see Table 8). In the light of the definitions of the truth conditions for sentences of the forms $\mathcal{W}_a C$ and $\mathcal{A}_a C$ (see Section 3.2, conditions (ii), (xi) and (xii)), it should be obvious that R plays an important role in our systems. It will become even clearer when we introduce the various tableau rules in Section 4.2. Whether or not we can draw any interesting consequences from the fact that an individual c wants (or accepts) something (in a possible world) will depend on whether or not c is perfectly rational, that is, whether or not c is in R's extension (in this possible world).

The valuation function assigns extensions to so-called matrices. Given any closed boulesic formula of the form $\mathcal{W}_t C$ or $\mathcal{A}_t C$, we shall construct its matrix as follows. Let m be the least number greater than every n such that x_n occurs bound in C. From left to right, replace every occurrence of an individual constant with x_m, x_{m+1}, etc. The result is the formula's matrix. Here are some examples: the matrix of $\mathcal{W}_d Pc$ is $\mathcal{W}_{x_1} Px_2$; the matrix of $\mathcal{A}_c Pdc$ is $\mathcal{A}_{x_1} Px_2x_3$; the matrix of $\mathcal{W}_c(Pa \leftrightarrow (Pa \wedge Pa))$ is $\mathcal{W}_{x_1}(Px_2 \leftrightarrow (Px_3 \wedge Px_4))$; the matrix of $\mathcal{W}_c \Sigma_{x_1}(Fc \rightarrow G_{x_1})$ is $\mathcal{W}_{x_2} \Sigma x_1(Fx_3 \rightarrow Gx_1)$; the matrix

of $\mathcal{W}_c\mathcal{W}_d\Pi x_3 Px_3$ is $\mathcal{W}_{x_4}\mathcal{W}_{x_5}\Pi x_3 Px_3$, the matrix of $\mathcal{W}_c\Pi x_1\mathcal{W}_{x_1}\Sigma x_2 Px_1 x_2$ is $\mathcal{W}_{x_3}\Pi x_1\mathcal{W}_{x_1}\Sigma x_2 Px_1 x_2$, etc.

Let A be any formula. Then, $(A)[a_1, \ldots, a_n/x_1, \ldots, x_n]$ is the result of replacing every free occurrence of x_1 by a_1, and $\ldots$, and every free occurrence of x_n by a_n in A. $(A)[a_1, \ldots, a_n/x_1, \ldots, x_n]$ will be abbreviated as $(A)[a_1, \ldots, a_n/\vec{x}]$ (parentheses around A will sometimes be dropped). Here are some examples. Let A be $\mathcal{W}_{x_1}Px_2$. Then, $(A)[d, c/x_1, x_2] = \mathcal{W}_d Pc$. Let A be $\mathcal{A}_{x_1}Px_2 x_3$. Then, $(A)[c, d, c/x_1, x_2, x_3] = \mathcal{A}_c Pdc$. Let A be $\mathcal{W}_{x_4}\mathcal{W}_{x_5}\Pi x_3 Px_3$. Then, $(A)[c, d/x_4, x_5] = \mathcal{W}_c\mathcal{W}_d\Pi x_3 Px_3$, etc.

If M is any matrix of the form $\mathcal{W}_t C$ or $\mathcal{A}_t C$ with free variables $x_1, \ldots, x_n$, then $v_\omega(M) \subseteq D^n$. Intuitively, this means that M is interpreted as a predicate and not as a (closed) sentence. Note that M always includes at least one free variable. Let M be a matrix where x_m is the first free variable in M and a_m is the constant in $M[a_1, \ldots, a_n/\vec{x}]$ that replaces x_m. Then the truth conditions for closed boulesic formulas of the form $M[a_1, \ldots, a_n/\vec{x}]$, when $v_\omega(Ra_m) = 0$, are defined in terms of the extension of M in ω. If $v_\omega(Rc) = 1$, then $\mathcal{W}_c$ in $\mathcal{W}_c B$ ($\mathcal{A}_c$ in $\mathcal{A}_c B$) will behave as a modal operator in ω. (See conditions (ii), (xi) and (xii) in Section 3.2 below for more on this.)[15]

$v_\omega(=) = \{\langle d, d\rangle : d \in D\}$, i.e. the extension of the identity predicate is the same in every possible world (in a model). It follows that all identities (and non-identities) are both absolutely and historically necessary. The existence predicate E functions as an ordinary predicate. The extension of this predicate may vary from one world to another. 'Ec' is true in a possible world iff $v(c)$ exists in this world.

3.2. Truth conditions

We now extend the interpretation function. Every closed formula, A, is assigned exactly one truth-value (1 = True or 0 = False), $v_\omega(A)$, in each world ω.

Here are the truth conditions for some sentences in our language. (The truth conditions for the omitted truth-functional connectives are the usual

---

[15]   See Priest (2005, Ch. 1–2) and Section 5.1 in this paper for more on matrices.

ones. '$\forall \omega' \in W$' means 'for every possible world ω' in W'; and '$\exists \omega' \in W$' means 'for some possible world ω' in W.')

(i) $v_\omega(Pa_1 \ldots a_n) = 1$ iff $\langle v(a_1), \ldots, v(a_n) \rangle \in v_\omega(P)$.

Let M be a matrix where x_m is the first free variable in M and a_m is the constant in $M[a_1, \ldots, a_n/\vec{x}]$ that replaces x_m. Then the truth conditions for closed boulesic formulas of the form $M[a_1, \ldots, a_n/\vec{x}]$, when $v_\omega(Ra_m) = 0$, are given in (ii) below.

(ii) $v_\omega(M[a_1, \ldots, a_n/\vec{x}]) = 1$ iff $\langle v(a_1), \ldots, v(a_n) \rangle \in v_\omega(M)$.
(iii) $v_\omega(UA) = 1$ iff $\forall \omega' \in W$: $v_{\omega'}(A) = 1$.
(iv) $v_\omega(MA) = 1$ iff $\exists \omega' \in W$: $v_{\omega'}(A) = 1$.
(v) $v_\omega(\Box A) = 1$ iff $\forall \omega' \in W$ s.t. $\Re \omega \omega'$: $v_{\omega'}(A) = 1$.
(vi) $v_\omega(\Diamond A) = 1$ iff $\exists \omega' \in W$ s.t. $\Re \omega \omega'$: $v_{\omega'}(A) = 1$.
(vii) $v_\omega(OA) = 1$ iff $\forall \omega' \in W$ s.t. $\mathfrak{S} \omega \omega'$: $v_{\omega'}(A) = 1$.
(viii) $v_\omega(PA) = 1$ iff $\exists \omega' \in W$ s.t. $\mathfrak{S} \omega \omega'$: $v_{\omega'}(A) = 1$.
(ix) $v_\omega(\Pi x A) = 1$ iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega(A[k_d/x]) = 1$.
(x) $v_\omega(\Sigma x A) = 1$ iff for some $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega(A[k_d/x]) = 1$.

Note that O and P are not included in the language $L1$, while they are primitive in the language $L2$.

Here are the truth conditions for $\mathcal{W}_a C$ and $\mathcal{A}_a C$.

(xi) $v_\omega(\mathcal{W}_a C) = 1$ iff for all ω' such that $\mathfrak{A} v(a) \omega \omega'$: $v_{\omega'}(C) = 1$, given that $v(a)$ is an element in $v_\omega(R)$, if $v(a)$ is not an element in $v_\omega(R)$, then $\mathcal{W}_a C$ is assigned a truth value in ω in a way that does not depend on the value of C (see condition (ii) above).

(xii) $v_\omega(\mathcal{A}_a C) = 1$ iff for at least one ω' such that $\mathfrak{A} v(a) \omega \omega'$: $v_{\omega'}(C) = 1$, given that $v(a)$ is an element in $v_\omega(R)$, if $v(a)$ is not an element in $v_\omega(R)$, then $\mathcal{A}_a C$ is assigned a truth value in ω in a way that does not depend on the value of C (see condition (ii) above).

Intuitively, conditions (xi) and (xii) can be interpreted in the following way: if $v(a)$ is not perfectly rational in a possible world, $\mathcal{W}_a C$ and $\mathcal{A}_a C$ behave as if they are ordinary predicates in this world; and if $v(a)$ is perfectly rational in a possible world, $\mathcal{W}_a$ and $\mathcal{A}_a$ behave as ordinary modal

operators in this world. So, the truth value of $\mathcal{W}_a C$ ($\mathcal{A}_a C$) in the possible world ω when a is not perfectly rational in ω is not determined by anything that goes on in some other world. It is, for example, logically possible for someone who is not perfectly rational to want C without wanting B even though B is a necessary means to C (see Section 5.1 for more on this).

3.3. Conditions on models

In this section, I will consider some conditions that might be imposed on our models. These conditions concern the formal properties of the accessibility relations, the relationships between the various accessibility relations and the relationships between the accessibility relations and the valuation function. In the formulas in this section, we can think of c and d as varying over individuals in D, and x, y, z and w as varying over possible worlds in W. Table 1 and Table 2 include information about the alethic and the deontic accessibility relations. The well-known conditions introduced in these tables are mentioned in almost any introduction to modal and deontic logic (see the introduction for some references). The clauses in Table 4, which concern the relationships between the alethic and the deontic accessibility relations, have been discussed by Rönnedal (2012), for instance. All other conditions are new.

The clauses in this section can be combined in many different ways, generating many different boulesic and boulesic-deontic systems (sections 3.5 and 4.3). Exactly which conditions *should* we accept? The answer to this question will depend on what it means to be perfectly rational. I think there might be good reasons to accept *all* (or almost all) conditions in this section. In Section 3.4, I will consider one such reason. However, it might also be interesting to see what follows if we accept some other, smaller class. The conditions in this section should be more or less self-explanatory. Nevertheless, I have added a few comments about some of the new clauses. There are many interesting relationships between the various conditions that I do not have space to discuss in this paper. In Section 3.4, I will consider what follows if we define the ternary boulesic accessibility relation $\mathfrak{A}$ in terms of the alethic accessibility relation $\mathfrak{R}$ and a binary acceptance predicate, and the deontic accessibility relation $\mathfrak{S}$ in terms of the ternary boulesic accessibility relation. This will show how deontic logic can in a certain sense be grounded in boulesic logic.

3.3.1. Conditions on the relation $\mathfrak{R}$

Condition	Formalisation of condition
$C-aT$	$\forall x \mathfrak{R}xx$
$C-aD$	$\forall x \exists y \mathfrak{R}xy$
$C-aB$	$\forall x \forall y (\mathfrak{R}xy \rightarrow \mathfrak{R}yx)$
$C-a4$	$\forall x \forall y \forall z ((\mathfrak{R}xy \wedge \mathfrak{R}yz) \rightarrow \mathfrak{R}xz)$
$C-a5$	$\forall x \forall y \forall z ((\mathfrak{R}xy \wedge \mathfrak{R}xz) \rightarrow \mathfrak{R}yz)$

Table 1

3.3.2. Conditions on the relation $\mathfrak{S}$

Condition	Formalisation of condition
$C-dD$	$\forall x \exists y \mathfrak{S}xy$
$C-d4$	$\forall x \forall y \forall z ((\mathfrak{S}xy \wedge \mathfrak{S}yz) \rightarrow \mathfrak{S}xz)$
$C-d5$	$\forall x \forall y \forall z ((\mathfrak{S}xy \wedge \mathfrak{S}xz) \rightarrow \mathfrak{S}yz)$
$C-dT'$	$\forall x \forall y (\mathfrak{S}xy \rightarrow \mathfrak{S}yy)$
$C-dB'$	$\forall x \forall y \forall z ((\mathfrak{S}xy \wedge \mathfrak{S}yz) \rightarrow \mathfrak{S}zy)$

Table 2

3.3.3. Conditions on the relation $\mathfrak{A}$

Condition	Formalisation of condition
$C-bD$	$\forall d \forall x \exists y \mathfrak{A}dxy$
$C-b4$	$\forall d \forall x \forall y \forall z ((\mathfrak{A}dxy \wedge \mathfrak{A}dyz) \rightarrow \mathfrak{A}dxz)$
$C-b5$	$\forall d \forall x \forall y \forall z ((\mathfrak{A}dxy \wedge \mathfrak{A}dxz) \rightarrow \mathfrak{A}dyz)$
$C-bT'$	$\forall d \forall x \forall y (\mathfrak{A}dxy \rightarrow \mathfrak{A}dyy)$
$C-bB'$	$\forall d \forall x \forall y \forall z ((\mathfrak{A}dxy \wedge \mathfrak{A}dyz) \rightarrow \mathfrak{A}dzy)$

Table 3[16]

---

[16] 'C' in '$C-bD$' stands for 'condition' and 'b' for 'boulesic.' $C-bD$ is called '$C-bD$' because it is similar to the well-known condition D (as in 'deontic') in ordinary

The conditions in Table 3 correspond to the tableau rules in Table 14. Note that $\mathfrak{A}$ is a ternary relation. $C-bD$ (Table 3) says: for every (individual) d and for every (possible world) x there is a (possible world) y such that d accepts y in x. According to this condition, every individual always accepts at least one possible world, no matter what situation she is in. If all possibilities are in some sense 'bad,' she will accept the possibility (or possibilities) that is (are) 'least bad,' so to speak. In all classes of models that satisfy this condition, the following sentence (schema) is valid: $\Pi x(Rx \rightarrow \neg(\mathcal{W}_x B \wedge \mathcal{W}_x \neg B))$ ('For every x: if x is perfectly rational, then it is not the case that x wants it to be the case that B and x wants it to be the case that not-B'). This is an intuitively plausible principle. If c wants it to be the case that B and also wants it to be the case that not-B, not all of c's wants can be satisfied. There is no possible world in which both B and not-B are true, and c cannot see to it that B and see to it that not-B.

3.3.4. Conditions concerning the relation between $\mathfrak{R}$ and $\mathfrak{S}$

Condition	Formalisation of condition
$C-MO$	$\forall x \forall y(\mathfrak{S}xy \rightarrow \mathfrak{R}xy)$
$C-OC$	$\forall x \exists y(\mathfrak{S}xy \wedge \mathfrak{R}xy)$
$C-OC'$	$\forall x \forall y(\mathfrak{S}xy \rightarrow \exists z(\mathfrak{S}yz \wedge \mathfrak{R}yz))$
$C-MO'$	$\forall x \forall y \forall z((\mathfrak{S}xy \wedge \mathfrak{S}yz) \rightarrow \mathfrak{R}yz)$
$C-ad4$	$\forall x \forall y \forall z((\mathfrak{R}xy \wedge \mathfrak{S}yz) \rightarrow \mathfrak{S}xz)$
$C-ad5$	$\forall x \forall y \forall z((\mathfrak{R}xy \wedge \mathfrak{S}xz) \rightarrow \mathfrak{S}yz)$
$C-PMP$	$\forall x \forall y \forall z((\mathfrak{S}xy \wedge \mathfrak{R}xz) \rightarrow \exists w(\mathfrak{R}yw \wedge \mathfrak{S}zw))$
$C-OMP$	$\forall x \forall y \forall z((\mathfrak{R}xy \wedge \mathfrak{S}yz) \rightarrow \exists w(\mathfrak{S}xw \wedge \mathfrak{R}wz))$

Table 4

alethic (modal) logic. Similar remarks apply to the other conditions in this section. It is usually binary relations that are called serial, transitive, Euclidean, etc. Nonetheless, we will extend these concepts to ternary relations. If $\mathfrak{A}$ satisfies $C-b4$, we will call $\mathfrak{A}$ transitive, and so on. If it is clear from the context that we are talking about a semantic condition, I will often omit the initial C.

3.3.5. Conditions concerning the relation between $\mathfrak{R}$ and $\mathfrak{A}$

Condition	Formalisation of condition
$C - M\mathcal{W}$	$\forall d\forall x\forall y(\mathfrak{A}dxy \rightarrow \mathfrak{R}xy)$
$C - \mathcal{W}C$	$\forall d\forall x\exists y(\mathfrak{A}dxy \wedge \mathfrak{R}xy)$
$C - \mathcal{W}C'$	$\forall d\forall x\forall y(\mathfrak{A}dxy \rightarrow \exists z(\mathfrak{A}dyz \wedge \mathfrak{R}yz))$
$C - M\mathcal{W}'$	$\forall d\forall x\forall y\forall z((\mathfrak{A}dxy \wedge \mathfrak{A}dyz) \rightarrow \mathfrak{R}yz)$
$C - ab4$	$\forall d\forall x\forall y\forall z((\mathfrak{R}xy \wedge \mathfrak{A}dyz) \rightarrow \mathfrak{A}dxz)$
$C - ab5$	$\forall d\forall x\forall y\forall z((\mathfrak{R}xy \wedge \mathfrak{A}dxz) \rightarrow \mathfrak{A}dyz)$
$C - \mathcal{A}MP$	$\forall d\forall x\forall y\forall z((\mathfrak{A}dxy \wedge \mathfrak{R}xz) \rightarrow \exists w(\mathfrak{R}yw \wedge \mathfrak{A}dzw))$
$C - \mathcal{W}MP$	$\forall d\forall x\forall y\forall z((\mathfrak{R}xy \wedge \mathfrak{A}dyz) \rightarrow \exists w(\mathfrak{A}dxw \wedge \mathfrak{R}wz))$

Table 5[17]

The conditions in Table 5 are similar to the conditions in Table 4. However, the clauses in Table 5 concern the relationship between the boulesic accessibility relation and the alethic accessibility relation. The conditions in Table 5 correspond to the tableau rules in Table 18. $C - M\mathcal{W}$ says: 'For every (individual) d, for every (possible world) x and for every (possible world) y, d accepts y in x only if y is alethically accessible from x.' In other words, if $C - M\mathcal{W}$ holds, then it is not the case that d accepts y in x if y is not alethically accessible from x. In every class of models that satisfies this condition, the following version of the hypothetical imperative is valid: $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$ (see the introduction; in Section 3.6, I will prove this claim). So, this condition is philosophically quite interesting.

$C - \mathcal{W}C$ is another philosophically interesting condition. According to $C - \mathcal{W}C$, for every (individual) d, for every (possible world) x there is

---

[17] '$M\mathcal{W}$' in '$C - M\mathcal{W}$' stands for 'Must Want,' and '$\mathcal{W}C$' in '$C - \mathcal{W}C$' for 'Want Can.' $C - ab4$ (as in 'alethic boulesic 4') is called '$C - ab4$' because it is similar to the well-known alethic (modal) condition $C - 4$ and the alethic deontic condition $C - ad4$, and similarly for $C - ad5$. '$\mathcal{A}MP$' in '$C - \mathcal{A}MP$' is an abbreviation of 'Acceptance Must Permutation,' and '$\mathcal{W}MP$' in '$C - \mathcal{W}MP$' is an abbreviation of 'Want Must Permutation.'

a (possible world) y such that d accepts y in x and y is alethically accessible from x. In other words, in every possible world, d accepts at least one possible world that is alethically accessible. This condition is similar to condition $C-bD$ (Table 3). $C-\mathcal{WC}$ entails $C-bD$, but $C-bD$ (in itself) does not entail $C-\mathcal{WC}$. In every class of models that satisfies this condition, the following schema is valid: $\Pi x(Rx \rightarrow (\mathcal{W}_x A \rightarrow \Diamond A))$ ('For every x: if x is perfectly rational, then x wants it to be the case that A only if A is (historically) possible'). Hence, according to this condition, a perfectly rational individual does not want anything impossible. This is an intuitively plausible principle. If c wants something that is impossible, c's want will inevitably be frustrated; c can never see to it that anything impossible is the case. In the introduction, we called this principle the 'Want-Can principle' ($\mathcal{WC}$).

Space does not allow me to discuss every philosophically interesting argument for and against the Want-Can principle, but I would like to address one possible counterexample (this problem was raised by an anonymous reviewer). If $\mathcal{WC}$ is valid, then we must accept the following instance of this principle: If a perfectly rational individual wants a unicorn to exist then it is (historically) possible that a unicorn exists. But this instance is counterintuitive. Hence, $\mathcal{WC}$ cannot be valid. I agree that this instance seems somewhat strange, even perhaps counterintuitive, at a first glance. But I do not think that this fact refutes the principle. Let me explain why. The expression 'If, then' in this instance should be interpreted as material implication. Sometimes we read more into this expression in English, for example, some kind of causal relation. But we should avoid this in our case. Suppose that a perfectly rational individual c wants it to be the case that A. $\mathcal{WC}$ does not entail that we have to assume that c's attitude *causes* it to be the case that it is historically possible that A. If we are idealists about possibilities, we might want to make such an assumption. But we do not have to be idealists to accept the systems in this paper. We *can* read $\mathcal{WC}$ in the following way: 'For every x: if x is perfectly rational, then *if x wants it to be the case that A then A* is (historically) possible.' But, as I suggested above, it might be more plausible to read it in the other direction, that is, in the following way: 'For every x: if x is perfectly rational, then x wants it to be the case that A *only if A* is (historically) possible.' We can think of

our example in this way. Then we should read our instance of $\mathcal{WC}$ in the following way: 'If someone is perfectly rational, then she wants a unicorn to exist only if it is (historically) possible that a unicorn exists.' Suppose that it is not (historically) possible that a unicorn exists. (In our possible world it seems to be the case that it is logically but *not* historically possible that there exists a unicorn.) Then a perfectly rational individual will adjust his or her attitudes to this fact. Hence, he or she will not want a unicorn to exist. Furthermore, recall that we use 'want' in an all-things-considered sense in this paper. So, the fact that a perfectly rational individual does not want anything that is (historically) impossible according to $\mathcal{WC}$ does not necessarily entail that he or she cannot daydream about unicorns, think about what it would be like if a unicorn existed, believe that it would be cool if a unicorn existed, etc. (see footnote 3). However, he or she will not want a unicorn to exist in an all-things-considered sense. So, I do not think that this example refutes $\mathcal{WC}$.

3.3.6. Conditions concerning the relation between $\mathfrak{S}$ and $\mathfrak{A}$

Condition	Formalisation of condition
$C-O\mathcal{W}$	$\forall x \forall y (\exists d\mathfrak{A}dxy \rightarrow \mathfrak{S}xy)$
$C-\mathcal{W}O$	$\forall x \forall y (\mathfrak{S}xy \rightarrow \exists d\mathfrak{A}dxy)$
$C-\mathfrak{A}\Sigma$	$\forall x \forall y (\mathfrak{S}xy \leftrightarrow \exists d\mathfrak{A}dxy)$
$C-O\mathcal{W}'$	$\forall x \forall y (\forall d\mathfrak{A}dxy \rightarrow \mathfrak{S}xy)$
$C-\mathcal{W}O'$	$\forall x \forall y (\mathfrak{S}xy \rightarrow \forall d\mathfrak{A}dxy)$
$C-\mathfrak{A}\Pi$	$\forall x \forall y (\mathfrak{S}xy \leftrightarrow \forall d\mathfrak{A}dxy)$

Table 6[18]

The conditions in Table 6 are concerned with some possible relationships between the deontic accessibility relation $\mathfrak{S}$ and the boulesic accessibility relation $\mathfrak{A}$. According to $C-\mathfrak{A}\Sigma$, y is deontically accessible from x iff y is acceptable to at least one individual in x; and according to $C-\mathfrak{A}\Pi$, y is

---

[18]   '$O\mathcal{W}$' is an abbreviation of 'Ought Want,' and '$\mathcal{W}O$' of 'Want Ought.'

deontically accessible from x iff y is acceptable to all individuals in x. $C-$ $\mathfrak{A}\Pi$ is an immediate consequence of ($Def\ \mathfrak{S}$), which is a definition that we will introduce in Section 3.4. $C-\mathfrak{A}\Sigma$ follows from $C-O\mathcal{W}$ and $C-\mathcal{W}O$, and $C-\mathfrak{A}\Pi$ follows from $C-O\mathcal{W}'$ and $C-\mathcal{W}O'$. If the condition $C-HW$ (Table 7) holds, then $C-\mathfrak{A}\Sigma$ and $C-\mathfrak{A}\Pi$ are equivalent, for then a possible world y is boulesically accessible from a possible world x to some individual iff y is boulesically accessible from x to every individual. $C-O\mathcal{W}$ corresponds to the tableau rule $T-O\mathcal{W}$ and $C-\mathcal{W}O$ to the tableau rule $T-\mathcal{W}O$ (Table 19). In Sections 3.4 and 4.2.12, we will consider some consequences of $C-$ $\mathfrak{A}\Pi$ ($Def\ \mathfrak{S}$). In every model that satisfies $C-O\mathcal{W}$ the following sentence is valid: $OA \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_xA)$, which says that if it ought to be the case that A then everyone who is perfectly rational wants it to be the case that A. If a model satisfies $C-\mathcal{W}O$ (and $C-HW$ and $C-\Sigma R$ in Table 7), then $\Pi x(Rx \rightarrow \mathcal{W}_xA) \rightarrow OA$ is valid in this model. $\Pi x(Rx \rightarrow \mathcal{W}_xA) \rightarrow OA$ says that it ought to be the case that A if everyone who is perfectly rational wants it to be the case that A. The intuitive idea behind the conditions in Table 6 is that there might be some interesting connections between what perfectly rational individuals want and accept and various normative 'facts' about what ought to be the case, about what is permitted and about what is not permitted. In our systems, we can explore those possible connections in a systematic and precise way. (For more on this, see Sections 3.4 and 4.2.12.)

3.3.7. One more condition on the relation $\mathfrak{A}$ and the being of a perfectly rational individual

Condition	Formalisation of condition
$C-HW$	$\forall c\forall d\forall x\forall y(\mathfrak{A}cxy \rightarrow \mathfrak{A}dxy)$
$C-\Sigma R$	In every possible world, ω, there is at least one individual, d, such that d is in R's extension in ω.

Table 7[19]

---

[19] 'HW' in '$C-HW$' is an abbreviation of '[the] Harmony of the Wills' and 'R' in '$T-\Sigma R$' is an abbreviation of 'perfectly rational' or 'perfectly reasonable.'

$C - HW$ says that if the possible world y is acceptable to the individual c in the possible world x, then y is acceptable to any individual d in x. $C - HW$ corresponds to the tableau rule $T - HW$ (Section 4.2.8, Table 15).

$C - \Sigma R$ says the following: in every possible world, there is at least one individual that is perfectly rational. This does not entail that there is one individual that is perfectly rational in every possible world. However, if there is one individual that is perfectly rational in every possible world, then obviously $C - \Sigma R$ holds. $C - \Sigma R$ corresponds to the tableau rule $T - \Sigma R$ (Section 4.2.8, Table 15). When I say that 'there is at least one individual,' I do not mean that this necessarily entails that this individual exists in this world. The expression is interpreted as a kind of 'possibilist quantifier' that is supposed to range over every possible object. Obviously, though, if there exists an individual that is perfectly rational (in some possible world), then there is such an individual (in this world).

3.3.8. Conditions on the valuation function v in a model

Condition	Formalisation of condition
$C - FTR$	If $\Re\omega_1\omega_2$ and Rc is true in ω_1, then Rc is true in ω_2 (for any c).
$C - UR$	If Rc is true in ω_1, then Rc is true in ω_2 (for any c).

Table 8

The semantic conditions $C - FTR$ and $C - UR$ correspond to the tableau rules $T - FTR$ and $T - UR$, respectively. (See Section 4.2.10, Table 17, for more on this.)

3.4. Relations between semantic conditions

There are many interesting relationships between the conditions introduced in Section 3.3. It is not possible to go through them all, but I will mention some of the most interesting. Due to considerations of space, proofs are omitted. Let us begin by saying a few words about the alethic accessibility relation.

Remark 2   *The following facts are well-known. (a) If $\mathfrak{R}$ is reflexive (C−aT), then $\mathfrak{R}$ is serial (C−aD). (b) $\mathfrak{R}$ is an equivalence relation iff (i) $\mathfrak{R}$ is reflexive (C−aT), symmetric (C−aB) and transitive (C−a4), iff (ii) $\mathfrak{R}$ is reflexive (C−aT) and Euclidean (C−a5), iff (iii) $\mathfrak{R}$ is serial (C−aD), symmetric (C−aB) and transitive (C−a4), iff (iv) $\mathfrak{R}$ is serial (C−aD), symmetric (C−aB) and Euclidean (C−a5).*

Since $\square$ and $\diamondsuit$ are interpreted as historical necessity and possibility, it is reasonable to treat $\mathfrak{R}$ as an equivalence relation. If we assume this, $\square$ and $\diamondsuit$ will behave as so-called $S5$-operators.

The following theorem says something about the relationships between the conditions on the deontic accessibility relation and between the conditions on the boulesic accessibility relation.

Remark 3   *The following facts are well-known. (i) If $\mathfrak{S}$ is Euclidean (C−d5), then $\mathfrak{S}$ is almost (secondarily) reflexive (C−dT′) and almost (secondarily) symmetric (C−dB′). Likewise, it is easy to prove the following facts. (ii) If $\mathfrak{A}$ is Euclidean (C−b5), then $\mathfrak{A}$ is almost (secondarily) reflexive (C−bT′) and almost (secondarily) symmetric (C−bB′).*

Now we will introduce some definitions. First, we define the ternary boulesic accessibility relation $\mathfrak{A}$ and the binary deontic accessibility relation $\mathfrak{S}$, and then we investigate the consequences of these definitions. We also introduce a new semantic condition called *the Accessibility Condition*, or $C−bdD$ (Definition 6), which has some interesting implications.[20]

Definition 4   *Def* $\mathfrak{A}$. $\forall c \forall x \forall y (\mathfrak{A}cxy =_{df} (\mathfrak{R}xy \wedge Tcy))$.[21]

In this definition (Definition 4), T is a binary predicate that says that (the individual) c accepts (the possible world) y or that y is acceptable to c. *Def* $\mathfrak{A}$ can be read as: 'For every (individual) c and for all (possible worlds) x and y: y is acceptable to c in x iff y is alethically accessible from

---

[20]   Note that the definitions in this paper are 'theoretical' rather than 'lexical' or 'descriptive.' Furthermore, the 'definiens' does not have to have the same meaning as the 'definiendum' in every respect.

[21]   Note that this definition does not entail any of the following propositions: $\forall c \exists y Tcy$, $\exists c \exists y Tcy$, $\forall c \exists y (Rcy \wedge Tcy)$, $\exists c \exists y (Rcy \wedge Tcy)$.

x and c accepts y.' This is a definition of the ternary boulesic accessibility relation $\mathfrak{A}$ in terms of the alethic accessibility relation $\mathfrak{R}$ and the binary accessibility relation T.[22]

Definition 5 *Def* $\mathfrak{S}$. $\forall x \forall y (\mathfrak{S}xy =_{df} \forall c \mathfrak{A}cxy)$.

Def $\mathfrak{S}$ (Definition 5) was mentioned already in Section 3.3.6. It should be obvious that $C-\mathfrak{A}\Pi$ is an immediate consequence of *Def* $\mathfrak{S}$. *Def* $\mathfrak{S}$ is a definition of the deontic accessibility relation $\mathfrak{S}$ in terms of the boulesic accessibility relation $\mathfrak{A}$. Informally, Definition 5 says the following: 'For all (possible worlds) x and y: y is deontically accessible from x iff every (individual) c accepts y in x' (or iff y is acceptable to every individual c in x). The intuition behind this definition is that the aim of morality is to create a possible world that everyone (who is perfectly rational) accepts (or can accept). This is an idea that might be attractive to at least some ideal observer theorists, Kantians, contractualists, moral idealists, constructivists, response dependent theorists, and divine will theorists. The definition has several interesting formal consequences (see Theorem 7 and Theorem 8 below), but it does not tell us anything about which worlds various

---

[22] *Def* $\mathfrak{A}$ is compatible with many different theories about what it means for an individual to accept a possible world (for a world to be acceptable to an individual). T is not necessarily a primitive, undefined relation. Here are some possible definitions: y is acceptable to c iff the utility of y for c is positive, or above a certain threshold or as high as possible, or iff c does not prefer any other possible world to y, or . . . . The important thing for our purposes in this paper is that all definitions of this kind share the same form. The definition is also consistent with the proposition that different individuals accept different worlds and that different individuals might have different reasons for accepting a possible world. Perhaps c accepts y because the utility of y for c is above a certain threshold, and perhaps d accepts z because z does not contain any serious violations of human rights. It is an interesting question whether or not T is definable, but for our purposes in this paper, we do not have to answer this question. However, note that not all definitions are compatible with condition $C-bdD$ (see definition 6) or with the proposition that there are possible worlds that are acceptable to some individual in some possible world. Suppose, for example, that y is acceptable to c iff the utility of y for c is positive and that there are no alethically accessible worlds from the world w in which the utility of y for c is positive. Then there is no world in w that is acceptable to c.

individuals accept or why they accept them (see footnote 22). So, the view is compatible with several different value theories and substantive normative theories.

Definition 5 does not entail that there is any world that everyone (who is perfectly rational) accepts. Accordingly, this definition does not guarantee that ought implies historical possibility. To guarantee this principle, we need to introduce another condition, namely the following:

Definition 6 *(C − bdD) The Accessibility Condition.* $\forall x \exists y (\mathfrak{R}xy \wedge \forall c\, Tcy)$.

The Accessibility Condition says the following: 'For every (possible world) x, there is a (possible world) y such that y is alethically accessible from x and for every (individual) c: c accepts y'; in other words, in every possible world there is at least one alethically accessible world that everyone accepts (or that is acceptable to everyone). Intuitively, this condition entails that no matter how good or bad things are in a given situation (possible world) everyone (who is perfectly rational) will agree that at least one possible outcome (world) is acceptable in this situation (world).[23] This definition (together with some other conditions) guarantees that ought implies historical possibility, that is, if we assume this condition, then everything that ought to be is historically possible and nothing historically impossible is obligatory (see Theorem 8 below).

Now we will investigate some consequences of these definitions.

Theorem 7 *Suppose that $\mathfrak{S}$ can be defined in terms of $\mathfrak{A}$ according to Definition 5 (Def $\mathfrak{S}$) and that the Harmony of the Wills holds (C − HW). Furthermore, suppose that $\mathfrak{A}$ is serial (C − bD), transitive (C − b4) and Euclidean (C − b5). Then $\mathfrak{A}$ is almost reflexive (C − bT′) and almost symmetric*

---

[23]   I do not suggest that The Accessibility Condition means that everyone '*consciously*' accepts at least one possible world. Ordinary people almost certainly do not have any conscious attitudes that involve whole possible worlds and they disagree about many things. But $\mathfrak{A}$ is only relevant for perfectly rational individuals; the truth values of sentences of the forms $\mathcal{W}_c B$ and $\mathcal{A}_c B$ when c is not perfectly rational do not depend on $\mathfrak{A}$. We are primarily interested in the structure of a perfectly rational will, not about the actual attitudes of ordinary people; T is an 'ideal' relation. So, this is not a problem for The Accessibility Condition.

$(C-bB')$, and $\mathfrak{S}$ is serial $(C-dD)$, transitive $(C-d4)$, Euclidean $(C-d5)$ and (hence) almost reflexive $(C-dT')$ and almost symmetric $(C-dB')$.

Theorem 8   *Suppose $\mathfrak{R}$ is an equivalence relation. (i) Then Def $\mathfrak{A}$ and Def $\mathfrak{S}$ entail the following conditions: d4, d5 and hence dT' and dB', b4, b5 and hence bT' and bB', MO, OC', MO', ad4, ad5, OMP, PMP, $M\mathcal{W}$, $\mathcal{W}C'$, $M\mathcal{W}'$, bd4, bd5, $\mathcal{W}MP$ and $\mathcal{A}MP$. (ii) Then Def $\mathfrak{A}$, Def $\mathfrak{S}$ and HW entail all conditions in (i), and all conditions in Table 6. (iii) Then Def $\mathfrak{A}$, Def $\mathfrak{S}$ and $C-bdD$ entail all the conditions in tables 1–5. (iv) Then Def $\mathfrak{A}$, Def $\mathfrak{S}$, HW and $C-bdD$ entail all the conditions in tables 1–6.*

Accordingly, if $C-HW$ and $C-bdD$ are plausible and the definitions in this section are reasonable (they do have significant intuitive appeal), then we have a good reason to accept all conditions in tables 1–6. The conditions in Table 7 and Table 8 might seem controversial. Nevertheless, I think one could make a good case for accepting them (at least every condition except $C-UR$[24]). However, space does not permit me to do this in the present paper. Whether or not we should accept all conditions in this section, clearly all of them are interesting enough to be worth discussing. (See sections 4.2.8 and 4.2.10 for more on some tableau rules that correspond to the semantic conditions in tables 7 and 8.)

3.5. Model classes and the logic of a class of models

The conditions mentioned in Section 3.3 can be used to obtain a categorisation of the set of all models into various kinds. We shall say that $\mathcal{M}(C_1, \ldots, C_n)$ is the class of (all) models that satisfy the conditions

---

[24] $C-UR$ is a theoretically important condition. Yet, there might be good reasons to reject it. Even though we, human beings, are not perfectly rational (see footnote 6), it seems interesting to consider what would be the case if we were. If all perfectly rational individuals necessarily are perfectly rational, we cannot do this, for then there are no individuals that are contingently perfectly rational—i.e. perfectly rational in some possible worlds and not perfectly rational in some other possible worlds. A being that is in fact not perfectly rational cannot then be perfectly rational in some other possible world.

$C_1, \ldots, C_n$. For example, $\mathcal{M}(C-bD, C-b4, C-b5)$ is the class of (all) models that satisfy the conditions $C-bD$, $C-b4$ and $C-b5$.

By imposing different conditions on our models we can obtain different logical systems. The set of all sentences in the language $L1$ ($L2$) that are valid in a class of models $\mathcal{M}$ is called the logical system of $\mathcal{M}$, or the system of $\mathcal{M}$, or the logic of $\mathcal{M}$, in symbols $\mathcal{S}(\mathcal{M})$. For example, $\mathcal{S}(\mathcal{M}(C-bD, C-b4, C-b5))$ (the system of $\mathcal{M}(C-bD, C-b4, C-b5)$) is the class of sentences in $L1$ ($L2$) that are valid in the class of (all) models that satisfy the conditions $C-bD$, $C-b4$ and $C-b5$.

By using this classification of model classes we can define a large set of systems. In the next section, I will develop semantic tableau systems that exactly correspond to these logics. I will consider four systems that seem especially philosophically interesting. The first is a pure boulesic system; the other three are boulesic-deontic systems (Section 4.3).

Definition 9 *(i) Let the class of all* strict *models be the class of models where $\mathfrak{R}$ is an equivalence relation and where $Def\,\mathfrak{A}$ holds. (ii) Let the class of all* strong *models be the class of all (supplemented) models where $\mathfrak{R}$ is an equivalence relation and where $Def\,\mathfrak{A}$, $Def\,\mathfrak{S}$, $C-HW$ and $C-\Sigma R$ hold. (iii) Let the class of all* strong+ *models be the class of all (supplemented) models where $\mathfrak{R}$ is an equivalence relation and where $Def\,\mathfrak{A}$, $Def\,\mathfrak{S}$, $C-bdD$, $C-HW$ and $C-\Sigma R$ hold. (iv) Let the class of all* almost complete *models be the class of all (supplemented) models where $\mathfrak{R}$ is an equivalence relation and where $Def\,\mathfrak{A}$, $Def\,\mathfrak{S}$, $C-bdD$, $C-HW$, $C-\Sigma R$ and $C-FTR$ hold.*

The first class in Definition 9 corresponds to strict boulesic logic, the second to strong boulesic-deontic logic, the third to strong+ boulesic-deontic logic, and the fourth to almost complete boulesic-deontic logic (Section 4.3, Definition 10). Hence, the system of the class of all strict models is the same as the set of all sentences provable in strict boulesic logic (see Section 4.3, Definition 10), etc. This follows from the soundness and completeness results in Section 6 and the results in Section 3.4.

3.6. An example of a valid formula

In the introduction, I mentioned the so-called hypothetical imperative. One of the most interesting readings of this principle was interpretation (6): $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$. Now I will show that this formula is valid in the class of all models that satisfy $C\text{--}M\mathcal{W}$ (Table 5).[25]

To establish this, assume that this sentence is not true in some possible world ω in some model $\mathcal{M}$ that satisfies $C\text{--}M\mathcal{W}$. Then there is some possible world ω' in $\mathcal{M}$ in which $\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$ is false. Hence, Rc, $\mathcal{W}_c A$ and $\Box(A \rightarrow B))$ are true in ω' in $\mathcal{M}$, while $\mathcal{W}_c B$ is false in ω' in $\mathcal{M}$ ('c' represents an arbitrary 'new' individual). Since c is perfectly rational in ω' in $\mathcal{M}$ and $\mathcal{W}_c B$ is false in ω' in $\mathcal{M}$, there is a possible world ω'' in $\mathcal{M}$ that is boulesically accessible to c from ω' in which B is false. c is perfectly rational in ω', ω'' is boulesically accessible to c from ω' and $\mathcal{W}_c A$ is true in ω' in $\mathcal{M}$. Hence, A is true in ω'' in $\mathcal{M}$. Since $\mathcal{M}$ satisfies $C\text{--}M\mathcal{W}$ and ω'' is boulesically accessible to c from ω' in $\mathcal{M}$, ω'' is alethically accessible from ω' in $\mathcal{M}$. Consequently, $A \rightarrow B$ is true in ω'' in $\mathcal{M}$, for ω'' is alethically accessible from ω' and $\Box(A \rightarrow B)$ is true in ω' in $\mathcal{M}$. Therefore, B is true in ω'' in $\mathcal{M}$ (by propositional logic). But this is absurd. Accordingly, our assumption cannot be true. In conclusion, $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$ is valid in $\mathcal{M}$. Since, ω and $\mathcal{M}$ were arbitrary, it follows that $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$ is valid in every model that satisfies $C\text{--}M\mathcal{W}$. Q.E.D.

4. Proof theory

4.1. Semantic tableaux

In Section 4, I will develop a set of tableau systems. The propositional part of these systems is similar to systems introduced by Raymond

---

[25]   In a strict sense, $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow \mathcal{W}_x B))$ is not a sentence but a schema. The argument in this section shows that every instance of this schema is valid in the class of all models that satisfy $C\text{--}M\mathcal{W}$.

Smullyan (1968) and Richard Jeffrey (1967), and the modal part is similar to systems discussed by Graham Priest (2008). For more information about the tableau method and various kinds of tableau systems, see, for example, (D'Agostino et al. 1999; and Fitting and Mendelsohn 1998).

The concepts of semantic tableau, branch, open and closed branch, etc. are essentially defined as in (Priest 2008).

4.2. Tableau rules

In this section, I will introduce a set of tableau rules that can be used to construct a large set of tableau systems (Section 4.3). They should be more or less self-explanatory. However, I will comment on some of the new rules.

4.2.1. Propositional rules

I will use the same propositional rules as in (Priest 2008). Let us call them $(\neg\neg)$, $(\wedge)$, $(\neg\wedge)$, $(\vee)$, $(\neg\vee)$, $(\rightarrow)$, $(\neg\rightarrow)$, $(\leftrightarrow)$ and $(\neg\leftrightarrow)$.

4.2.2. Basic alethic rules (ba-rules)

U	M	$\Box$	$\Diamond$
UA,i $\downarrow$ A,j for any j	MA,i $\downarrow$ A,j where j is new	$\Box A,i$ irj $\downarrow$ A,j	$\Diamond A,i$ $\downarrow$ irj A,j where j is new
$\neg U$	$\neg M$	$\neg\Box$	$\neg\Diamond$
$\neg UA,i$ $\downarrow$ $M\neg A,i$	$\neg MA,i$ $\downarrow$ $U\neg A,i$	$\neg\Box A,i$ $\downarrow$ $\Diamond\neg A,i$	$\neg\Diamond A,i$ $\downarrow$ $\Box\neg A,i$

Table 9

4.2.3. Basic boulesic and deontic rules (bb-rules and d-rules)

$\mathcal{W}$	$\mathcal{A}$	O	P
Rc,i $\mathcal{W}_cB,i$ $iAcj$ $\downarrow$ B,j	Rc,i $\mathcal{A}_cB,i$ $\downarrow$ $iAcj$ B,j where j is new	OB,i isj $\downarrow$ B,j	PB,i $\downarrow$ isj B,j where j is new
$\neg\mathcal{W}$	$\neg\mathcal{A}$	$\neg O$	$\neg P$
Rc,i $\neg\mathcal{W}_cB,i$ $\downarrow$ $\mathcal{A}_c\neg B,i$	Rc,i $\neg\mathcal{A}_cB,i$ $\downarrow$ $\mathcal{W}_c\neg B,i$	$\neg OB,i$ $\downarrow$ $P\neg B,i$	$\neg PB,i$ $\downarrow$ $O\neg B,i$

Table 10

Intuitively, 'Rc,i' in the boulesic rules says that the individual denoted by 'c' is perfectly rational in the possible world denoted by 'i,' and '$iAcj$' in the rules $\mathcal{W}$ and $\mathcal{A}$ says that the possible world denoted by 'j' is acceptable to the individual denoted by 'c' in the possible world denoted by 'i.' The basic boulesic rules hold for every constant c (i.e. c can be replaced by any constant in these rules).

4.2.4. Possibilist quantifiers

Π	Σ	$\neg\Pi$	$\neg\Sigma$
$\Pi xA,i$ $\downarrow$ $A[a/x],i$ for every constant a on the branch, a new if there are no constants on the branch	$\Sigma xA,i$ $\downarrow$ $A[c/x],i$ where c is new to the branch	$\neg\Pi xA,i$ $\downarrow$ $\Sigma x\neg A,i$	$\neg\Sigma xA,i$ $\downarrow$ $\Pi x\neg A,i$

Table 11

Note that a and c in the quantifier rules are rigid constants—we never instantiate with variables; a is any constant on the branch and c is a constant new to the branch.

4.2.5. Alethic accessibility rules (a-rules)

$T-aD$	$T-aT$	$T-aB$	$T-a4$	$T-a5$
i $\downarrow$ irj where j is new	i $\downarrow$ iri	irj $\downarrow$ jri	irj jrk $\downarrow$ irk	irj irk $\downarrow$ jrk

Table 12

4.2.6. Deontic accessibility rules (d-rules)

$T-dD$	$T-d4$	$T-d5$	$T-dT'$	$T-dB'$
i $\downarrow$ isj where j is new	isj jsk $\downarrow$ isk	isj isk $\downarrow$ jsk	isj $\downarrow$ jsj	isj jsk $\downarrow$ ksj

Table 13

4.2.7. Boulesic accessibility rules (b-rules)

$T-bD$	$T-b4$	$T-b5$	$T-bT'$	$T-bB'$
i $\downarrow$ $iAcj$ where j is new	$iAcj$ $jAck$ $\downarrow$ $iAck$	$iAcj$ $iAck$ $\downarrow$ $jAck$	$iAcj$ $\downarrow$ $jAcj$	$iAcj$ $jAck$ $\downarrow$ $kAcj$

Table 14

The boulesic accessibility rules hold for every constant c (i.e. c can be replaced by any constant in these rules). The b-rules in Table 14 correspond to the semantic conditions in Table 3.

4.2.8. Non-basic boulesic rules (nbb-rules)

$T-HW$	$T-\Sigma R$
$iAcj$ $\downarrow$ $iAdj$ for every c and d	i $\downarrow$ $\Sigma xRx,i$

Table 15

The rules in Table 15 correspond to the semantic conditions in Table 7. In every system that includes $T-HW$, we can prove the first six sentences in Table 25, and in every system that includes $T-HW$ and $T-bD$, we can prove all sentences in Table 25 (see Section 5). Therefore, in systems that include $T-HW$, we can prove that all perfectly rational individuals want and accept the same things. According to this rule, the idea of perfect rationality, or wisdom, includes a kind of *inter*personal consistency, not only a kind of *intra*personal consistency. The wills of perfectly rational individuals are consistent, they harmonise. If individual c wants it to be the case that B and individual d wants it to be the case that $\neg B$, then both cannot get what they want; either c's or d's desires will be frustrated: it is not possible to see to it that B and to see to it that $\neg B$. In systems that include $T-HW$ and $T-bD$, situations of this kind are ruled out. Hence, these conditions seem to be intuitively plausible (however, see footnote 14).

If we include $T-\Sigma R$ in our systems, we can prove that ΣxRx is necessarily true. Recall that ΣxRx says that there is something or someone, a possible individual that is perfectly rational. This does not entail that this individual exists.

Space does not permit me to discuss all philosophical arguments for and against these rules. However, it should be noted that $T-HW$ does not entail that all individuals that are *not* perfectly rational want and accept the same things, and it does not entail that everyone *should* have the same attitudes. Furthermore, it does not entail that everyone should act in the same way or be a certain kind of person, nor does it entail that if something is permitted for some person it is permitted for every person. Suppose that c and

d are perfectly rational. Even if this is the case, it is possible that both c and d want individual e to perform a certain action and that both want individual f not to perform this action. Situations of this kind are not inconsistent according to any system in this paper. If they were inconsistent, $T-HW$ would probably not be a philosophically reasonable rule.

4.2.9. The CUT-rule (CUT)

CUT
i
↙   ↘
A,i $\neg A,i$
for every A and i

Table 16[26]

4.2.10. Transfer-rules, etc.

$T-FTR$	$T-UR$
Rc,i	Rc,i
irj	↓
↓	Rc,j
Rc,j	for any j

Table 17[27]

The tableau rules in Table 17 correspond to the semantic conditions in Table 8.

---

[26] We could use a more restricted CUT rule, $CUTR$, where 'A' in CUT is replaced by 'Rc' where c is a constant (that occurs as an index to some boulesic operator) on the branch. In fact, in the completeness proofs we do not need CUT if our systems include $CUTR$. However, CUT is often more useful in proving theorems and deriving non-primitive rules. For more on the CUT rule, see, for example, (Rönnedal 2009).

[27] 'FT' in '$T-FTR$' is an abbreviation of 'Forward Transfer,' and 'R' in '$T-FTR$' and '$T-UR$' of 'Rationality.'

In every system that includes $T-UR$ or $T-FTR$ and $T-MW$ (Table 18), we can prove that the following sentence is a theorem: $\Pi x(Rx \rightarrow W_x Rx)$, which says that everyone who is perfectly rational wants to be perfectly rational.

In every system that includes $T-UR$ or $T-FTR$ and $T-MW$, and $T-bD$ (Table 14), we can prove that the following sentence is a theorem: $\Pi x(Rx \rightarrow A_x Rx)$, which says that everyone who is perfectly rational accepts that she is perfectly rational.

In every system that includes $T-UR$, we can prove the following sentence: $\Pi x(Rx \rightarrow URx)$, which says that every perfectly rational individual is necessarily perfectly rational.

We do not assume that the transfer rules (the rules in Table 17) are included in every system. Whether or not they should be added seems to be something of an open question.[28]

4.2.11. Alethic-boulesic accessibility rules (ab-rules)

$T-MW$	$T-MW'$	$T-WC$	$T-WC'$
$iAcj$ $\downarrow$ irj	$iAcj$ $jAck$ $\downarrow$ jrk	i $\downarrow$ $iAcj$ irj where j is new	$iAcj$ $\downarrow$ $jAck$ jrk where k is new
$T-ab4$	$T-ab5$	$T-AMP$	$T-WMP$
irj $jAck$ $\downarrow$ $iAck$	irj $iAck$ $\downarrow$ $jAck$	$iAcj$ irk $\downarrow$ jrl $kAcl$ where l is new	irj $jAck$ $\downarrow$ $iAcl$ lrk where l is new

Table 18

---

[28] See footnote 24 for some critique of $C-UR$, which is the semantic condition that corresponds to $T-UR$.

The ab-rules in Table 18 correspond to the semantic conditions introduced in Table 5.

4.2.12. Boulesic-deontic accessibility rules (bd-rules)

$T-\mathcal{OW}$	$T-\mathcal{WO}$
$iAbj$	isj
$\downarrow$	$\downarrow$
isj	$iAcj$
for any b	where c is new

Table 19

The rule $T-\mathcal{OW}$ in Table 19 corresponds to the semantic condition $C-\mathcal{OW}$ mentioned in Table 6, and the rule $T-\mathcal{WO}$ in Table 19 corresponds to the semantic condition $C-\mathcal{WO}$ in Table 6. In every system that includes $T-\mathcal{OW}$, we can prove that $OA \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$ is a theorem—i.e. if it ought to be the case that A, then everyone who is perfectly rational wants it to be the case that A. This is one version of a philosophically very interesting thesis often called 'existence internalism.' It follows from this theorem that if the individual c ought to do the action H, then if c is perfectly rational c wants to do H. However, if c is not perfectly rational, it is not necessary that she wants to do H. So, this kind of internalism is compatible with the existence of amoralists and with the phenomenon of weakness of will. Internalism can help explain the fact that we find utterances of the following kind puzzling: 'I know that I ought to do it, but I have no inclination whatsoever to do it' and 'You ought to do it, but by all means don't do it.' At the same time, the kind of internalism mentioned here avoids some of the common objections against this thesis.[29]

---

[29] For more information on internalism and various versions of internalism and arguments for and against this thesis, see, for example, (Björklund et.al. 2012; Björnsson et.al. 2015; and van Roojen 2013). It might be interesting to note that existence internalism entails the following version of 'knowledge internalism': $\mathcal{K}_c OA \rightarrow (Rc \rightarrow \mathcal{W}_c A)$, where '$\mathcal{K}_c A$' stands for '$c$ knows that A' (given that knowledge implies truth). '$\mathcal{K}_c OA \rightarrow (Rc \rightarrow \mathcal{W}_c A)$' says that if c knows that it ought to be the case that A, then if c is perfectly rational then c wants it to be the case that A.

In every system that includes $T-\mathcal{WO}$, $T-H\mathcal{W}$ and $T-\Sigma R$ we can prove that $\Pi x(Rx \rightarrow \mathcal{W}_x A) \rightarrow OA$ is a theorem—i.e. if everyone who is perfectly rational wants it to be the case that A, then it ought to be the case that A. This is the converse of $OA \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$. Together these theorems entail $OA \leftrightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$, which says that it ought to be the case that A iff everyone who is perfectly rational wants it to be the case that A. Similar equivalences hold for P (it is permitted that) and F (it is not permitted that) (see Table 30). These theorems can be seen as a part of a kind of ideal observer theory for normative propositions.[30]

Some might worry that the equivalences in Table 30 are too strong. If we accept those equivalences, do we not have to accept that, for example, 'Tom ought to go home' has the same meaning as 'Everyone who is perfectly rational wants Tom to go home' and isn't this unreasonable?[31] Personally, I do not think that we have to accept this. Let me explain why. Since $OA \leftrightarrow \Pi x(Rx \leftrightarrow \mathcal{W}_x A)$ holds in some systems, OA is in principle 'definable' in terms of $\Pi x(Rx \rightarrow \mathcal{W}_x A)$ in those systems (see Section 2.3). However, this fact does not entail that 'It ought to be the case that A' has the same meaning as 'Everyone who is perfectly rational wants it to be the case that A.' To say that OA is in principle 'definable' in terms of $\Pi x(Rx \rightarrow \mathcal{W}_x A)$ means that 'OA' can be replaced by '$\Pi x(Rx \rightarrow \mathcal{W}_x A)$' (and vice versa) in every 'extensional context,' but not necessarily in every 'intensional context,' for example, if 'OA' ('$\Pi x(Rx \rightarrow \mathcal{W}_x A)$') occurs within the scope of a boulesic operator. So, those systems do not entail that, for example, 'Tom ought to go home' says exactly the same thing as 'Everyone who is perfectly rational wants Tom to go home.' In this sense, our equivalences are similar

---

Furthermore, assume that every perfectly rational individual is infallible in the sense that everything she believes is true. Then existence internalism entails the following version of 'belief internalism': $\mathcal{B}_c OA \rightarrow (Rc \rightarrow \mathcal{W}_c A)$, where '$\mathcal{B}_c A$' stands for '$c$ believes that A.' '$\mathcal{B}_c OA \rightarrow (Rc \rightarrow \mathcal{W}_c A)$' says that if c believes that it ought to be the case that A, then if c is perfectly rational then c wants it to be the case that A.

[30]   For more on ideal observer theories, see, for example, (Firth 1952; and Kawall 2013).

[31]   An anonymous reviewer raised this worry.

to other equivalences in other branches of logic. In propositional logic, disjunction is in principle definable in terms of conjunction and negation since $(A \lor B) \leftrightarrow \neg(\neg A \land \neg B)$ is a tautology. This fact does not entail that 'Either London or Paris is the capital of France' means the same as 'It is not the case that it is not the case that London is the capital of France and it is not the case that Paris is the capital of France.' I suggest that the same thing is true of our equivalences.[32] In conclusion, the fact that 'Tom ought to go home' does not have the same meaning as 'Everyone who is perfectly rational wants Tom to go home' is not a serious problem for the systems that include the equivalence $OA \leftrightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$.

4.2.13. Alethic-deontic accessibility rules (ad-rules)

$T-MO$	$T-MO'$	$T-OC$	$T-OC'$
isj $\downarrow$ irj	isj jsk $\downarrow$ jrk	i $\downarrow$ isj irj where j is new	isj $\downarrow$ jsk jrk where k is new
$T-ad4$	$T-ad5$	$T-PMP$	$T-OMP$
irj jsk $\downarrow$ isk	irj isk $\downarrow$ jsk	isj irk $\downarrow$ jrl ksl where l is new	irj jsk $\downarrow$ isl lrk where l is new

Table 20

---

[32] In other words, meaning is stronger than necessary equivalence. The fact that A is necessarily equivalent with B does not entail that A and B have the same meaning; but if A has the same meaning as B, then A and B are necessarily equivalent.

4.2.14. Identity rules

$T-R=$	$T-S=$	$T-N=$	$T-A=$
$*$ $\downarrow$ $t=t,i$ for every t on the branch	$s=t,i$ $A[s/x],i$ $\downarrow$ $A[t/x],i$ where A is of a certain form (see below, 4.2.14)	$a=b,i$ $\downarrow$ $a=b,j$ for any j	$a=b,i$ $A\,ajk$ $\downarrow$ $A\,bjk$

Table 21[33]

$(T-S=)$ is applied only 'within worlds,' and we usually only apply the rule when A is atomic. However, we shall also allow applications of the following kind. Let M be a matrix where x_m is the first free variable in M and a_m is the constant in $M[a_1, \ldots, a, \ldots, a_n/\vec{x}]$ that replaces x_m. Furthermore, suppose we have $a=b,i$, $M[a_1, \ldots, a, \ldots, a_n/\vec{x}],i$ and $\neg Ra_m$ on the branch. Then we may apply $(T-S=)$ to obtain an extension of the branch that includes $M[a_1, \ldots, b, \ldots, a_n/\vec{x}],i$.

With the help of $(T-S=)$ and $(T-A=)$ we can prove the following theorems: $(\mathcal{W}_c B \land c=d) \to \mathcal{W}_d B$, $(\mathcal{A}_c B \land c=d) \to \mathcal{A}_d B$, $\Pi x \Pi y((\mathcal{W}_x B \land x=y) \to \mathcal{W}_y B)$ and $\Pi x \Pi y((\mathcal{A}_x B \land x=y) \to \mathcal{A}_y B)$. All of these theorems are intuitively plausible. By using $(T-N=)$, we can establish that all identities and non-identities are (absolutely and historically) necessary—i.e. we can prove all of the following theorems: $\Pi x \Pi y(x=y \to Ux=y)$, $\Pi x \Pi y(x=y \to \Box x=y)$, $\Pi x \Pi y(\neg x=y \to U\neg x=y)$, and $\Pi x \Pi y(\neg x=y \to \Box\neg x=y)$. This is plausible since every constant is treated as a rigid designator in this paper.

---

[33]   In the identity rules R stands for 'reflexive,' S for 'substitution (of identities),' N for 'necessary identity,' and A for '(boulesic) accessibility.'

4.3 Tableau systems and some basic proof-theoretical concepts

A tableau system is a set of tableau rules. I will consider two kinds of system in this paper: (pure) boulesic systems and boulesic-deontic systems.

A (pure) (alethic) boulesic system is a tableau system that includes the propositional rules, the basic alethic rules, the basic boulesic rules, the rules for the possibilist quantifiers, the *CUTR*-rule (or *CUT*) and the identity rules. The smallest boulesic system is called $\mathcal{V}$. By adding various transfer rules, and/or boulesic, alethic and/or alethic-boulesic accessibility rules to $\mathcal{V}$, we obtain a large class of stronger boulesic systems.

A (alethic) boulesic-deontic system is a tableau system that includes $\mathcal{V}$ and all basic deontic rules. The smallest boulesic-deontic system is called $\mathcal{BD}$. Every boulesic-deontic system that includes $T - HW$, $T - \Sigma R$, $T - OW$ and $T - WO$ will be called a normal boulesic-deontic system. The smallest normal boulesic-deontic system is called $\mathcal{NBD}$. By adding various tableau rules from Section 4.2 to $\mathcal{BD}$, we obtain extensions of this system. Our (normal) boulesic-deontic systems illustrate how deontic logic can be 'grounded' in boulesic logic in a certain sense.

Let $aA_1, \ldots, A_n bB_1, \ldots, B_n abC_1, \ldots, C_n TrD_1, \ldots, D_n$ be the boulesic system that includes the alethic accessibility rules $A_1, \ldots, A_n$, the boulesic accessibility rules $B_1, \ldots, B_n$, the alethic-boulesic accessibility rules $C_1, \ldots, C_n$, and the transfer rules $D_1, \ldots, D_n$. A boulesic-deontic system is defined in a similar way: $aA_1, \ldots, A_n bB_1, \ldots, B_n dC_1, \ldots, C_n abD_1, \ldots, D_n adE_1, \ldots, E_n TrF_1, \ldots, F_n$ is a boulesic-deontic system, where a, b, ab, and Tr are interpreted as in a boulesic system; $C_1, \ldots, C_n$ is a list (possibly empty) of deontic accessibility rules; and $E_1, \ldots, E_n$ is a list (possibly empty) of alethic-deontic rules.

Important proof theoretical concepts like the concepts of proof, theorem, derivation, consistency, inconsistency in a system, the logic of a tableau system, etc. are defined as usual (see, for example, Priest 2008).

I will now describe four different tableau systems that correspond to the four classes of models described in Definition 9. The first system is an example of a boulesic system; the other three are examples of boulesic-deontic systems.

Definition 10 *(i) Strict boulesic logic is the boulesic system that includes all a-rules and the tableau rules b4, b5, bT′, bB′, $\mathcal{MW}$, $\mathcal{WC}′$, $\mathcal{MW}′$, ab4, ab5, $\mathcal{WMP}$, and $\mathcal{AMP}$. (ii) Strong boulesic-deontic logic is the (normal) boulesic-deontic system that includes all a-rules and the tableau rules d4, d5, dT′, dB′, b4, b5, bT′, bB′, MO, OC′, MO′, ad4, ad5, OMP, PMP, MW, $\mathcal{WC}′$, $\mathcal{MW}′$, ab4, ab5, $\mathcal{WMP}$, $\mathcal{AMP}$. (iii) Strong+ boulesic-deontic logic is the (normal) boulesic-deontic system that includes all rules in tables 12–15 and 18–20. (iv) Almost complete boulesic-deontic logic is the (normal) boulesic-deontic system that includes all rules that are contained in Strong+ boulesic-deontic logic plus $T-FTR$.[34]*

Note that the following relations hold between these systems: Strict boulesic logic ⊆ Strong boulesic-deontic logic ⊆ Strong+ boulesic-deontic logic ⊆ Almost complete boulesic-deontic logic. As far as I can see, the following relations also hold: Strict boulesic logic ⊂ Strong boulesic-deontic logic ⊂ Strong+ boulesic-deontic logic ⊂ Almost complete boulesic-deontic logic. However, I will only offer the latter claim as a conjecture in the present paper.

5. Examples of theorems

In this section, I will mention some sentences that can be proved in various systems. The informal reading of the theses should be obvious. Every formula in Table 22 is a theorem in every system in this paper; every sentence in Table 23 is a theorem in every system that includes the tableau rule $T-bD$, etc.

All of the following sentences (schemas) are theorems in every system in this paper: $\Pi x(Rx \rightarrow (\mathcal{W}_x B \leftrightarrow \neg \mathcal{A}_x \neg B))$, $\Pi x(Rx \rightarrow (\neg \mathcal{W}_x B \leftrightarrow \mathcal{A}_x \neg B))$, $\Pi x(Rx \rightarrow (\mathcal{W}_x \neg B \leftrightarrow \neg \mathcal{A}_x B))$ and $\Pi x(Rx \rightarrow (\mathcal{A}_x B \leftrightarrow \neg \mathcal{W}_x \neg B))$. Note that universal necessity is stronger than historical necessity and that universal

---

[34]   Some of the rules in these systems are 'redundant,' and there are several 'weaker' systems that are deductively equivalent—i.e. they contain exactly the same theorems. 'Weaker system' here means a system with fewer primitive rules, not a system with fewer theorems.

possibility is weaker than historical possibility in every system in this paper. In other words, $UA \rightarrow \Box A$ and $\Diamond A \rightarrow MA$ are theorems in every system in this paper, while $\Box A \rightarrow UA$ and $MA \rightarrow \Diamond A$ are not theorems in any system in this paper. U and M behave as so-called $S5$-operators in every system in this paper and $\Box$ and $\Diamond$ behave as $S5$-operators in every system that includes every rule in Table 12 (note that not all rules have to be primitive).

Theorem	System
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \wedge B) \leftrightarrow (\mathcal{W}_xA \wedge \mathcal{W}_xB)))$	Every
$\Pi x(Rx \rightarrow ((\mathcal{W}_xA \vee \mathcal{W}_xB) \rightarrow \mathcal{W}_x(A \vee B)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{A}_x(A \wedge B) \rightarrow (\mathcal{A}_xA \wedge \mathcal{A}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{A}_x(A \vee B) \leftrightarrow (\mathcal{A}_xA \vee \mathcal{A}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \rightarrow B) \rightarrow (\mathcal{W}_xA \rightarrow \mathcal{W}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \rightarrow B) \rightarrow (\mathcal{A}_xA \rightarrow \mathcal{A}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \rightarrow B) \rightarrow (\mathcal{W}_x\neg B \rightarrow \mathcal{W}_x\neg A)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \leftrightarrow B) \rightarrow (\mathcal{W}_xA \leftrightarrow \mathcal{W}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \leftrightarrow B) \rightarrow (\mathcal{A}_xA \leftrightarrow \mathcal{A}_xB)))$	Every
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \leftrightarrow B) \rightarrow (\mathcal{W}_x\neg A \leftrightarrow \mathcal{W}_x\neg B)))$	Every

Table 22

Theorem	System
$\Pi x(Rx \rightarrow (\mathcal{W}_xB \rightarrow \mathcal{A}_xB))$	bD
$\Pi x(Rx \rightarrow \neg(\mathcal{W}_xB \wedge \mathcal{W}_x\neg B))$	bD
$\Pi x(Rx \rightarrow (\mathcal{A}_xB \vee \mathcal{A}_x\neg B))$	bD
$\Pi x(Rx \rightarrow \neg(\mathcal{W}_x(A \vee B) \wedge (\mathcal{W}_x\neg A \wedge \mathcal{W}_x\neg B)))$	bD
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \rightarrow B) \rightarrow (\mathcal{W}_xA \rightarrow \mathcal{A}_xB)))$	bD
$\Pi x(Rx \rightarrow (\mathcal{W}_x(A \rightarrow B) \rightarrow (\mathcal{W}_x\neg B \rightarrow \neg\mathcal{W}_xA)))$	bD

Table 23

Theorem	Systems
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow (\mathcal{W}_x B \rightarrow \mathcal{W}_x \mathcal{W}_x B))$	$b4$
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow (\mathcal{A}_x B \rightarrow \mathcal{W}_x \mathcal{A}_x B))$	$b5$
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow \mathcal{W}_x(\mathcal{W}_x B \rightarrow B))$	bT'
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow \mathcal{W}_x(\mathcal{A}_x \mathcal{W}_x A \rightarrow A))$	$bB'b4$
$\Pi x(Rx \rightarrow (\mathcal{W}_x B \rightarrow \mathcal{W}_x \mathcal{W}_x B))$	$b4\,UR$
$\Pi x(Rx \rightarrow (\mathcal{A}_x B \rightarrow \mathcal{W}_x \mathcal{A}_x B))$	$b5\,UR$
$\Pi x(Rx \rightarrow \mathcal{W}_x(\mathcal{W}_x B \rightarrow B))$	$bT'\,UR$
$\Pi x(Rx \rightarrow \mathcal{W}_x(\mathcal{A}_x \mathcal{W}_x A \rightarrow A))$	$bB'\,UR$

Table 24

Theorems	Systems
$\Pi x \Pi y((Rx \wedge Ry) \rightarrow (\mathcal{W}_x B \rightarrow \mathcal{W}_y B))$	HW
$\Pi x(Rx \rightarrow (\mathcal{W}_x B \rightarrow \Pi y(Ry \rightarrow \mathcal{W}_y B))$	HW
$\Pi x \Pi y((Rx \wedge Ry) \rightarrow (\mathcal{A}_x B \rightarrow \mathcal{A}_y B))$	HW
$\Pi x(Rx \rightarrow (\mathcal{A}_x B \rightarrow \Pi y(Ry \rightarrow \mathcal{A}_y B)))$	HW
$\Sigma x(Rx \wedge \mathcal{W}_x B) \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_x B)$	HW
$\Sigma x(Rx \wedge \mathcal{A}_x B) \rightarrow \Pi x(Rx \rightarrow \mathcal{A}_x B)$	HW
$\neg \Sigma x \Sigma y((Rx \wedge Ry) \wedge (\mathcal{W}_x B \wedge \mathcal{W}_y \neg B))$	$HWbD$
$\Pi x(Rx \rightarrow (\mathcal{W}_x B \rightarrow \Pi y(Ry \rightarrow \mathcal{A}_y B)))$	$HWbD$

Table 25

Theorems	Systems
$\Pi x(Rx \rightarrow (\Box A \rightarrow \mathcal{W}_x A))$	$abM\mathcal{W}$
$\Pi x(Rx \rightarrow (\mathcal{W}_x A \rightarrow \Diamond A))$	$ab\mathcal{W}C$
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow \mathcal{W}_x(\Box A \rightarrow \mathcal{W}_x A))$	$abM\mathcal{W}'$
$\Pi x((Rx \wedge \mathcal{W}_x Rx) \rightarrow \mathcal{W}_x(\mathcal{W}_x A \rightarrow \Diamond A))$	$ab\mathcal{W}C'$
$\Pi x(Rx \rightarrow (\mathcal{W}_x A \rightarrow \Box \mathcal{W}_x A))$	$ab4\,UR$

$\Pi x(Rx \to (\mathcal{A}_x B \to \Box \mathcal{A}_x B))$	$ab5\,UR$
$\Pi x(Rx \to (\mathcal{A}_x \Box B \to \Box \mathcal{A}_x B))$	$ab\mathcal{A}MPUR$
$\Pi x(Rx \to (\mathcal{W}_x \Box A \to \Box \mathcal{W}_x A))$	$ab\mathcal{W}MPUR$
$\Pi x(Rx \to \mathcal{W}_x(\Box A \to \mathcal{W}_x A))$	$abM\mathcal{W}'\,UR$
$\Pi x(Rx \to \mathcal{W}_x(\mathcal{W}_x A \to \Diamond A))$	$ab\mathcal{W}C'\,UR$

Table 26

Name	Theorem	Systems
MO	$\Box A \to OA$	$adMO$
OC	$OA \to \Diamond A$	$adOC$
OC'	$O(OA \to \Diamond A)$	$adOC'$
MO'	$O(\Box A \to OA)$	$adMO'$
$ad4$	$OA \to \Box OA$	$ad4$
$ad5$	$PA \to \Box PA$	$ad5$
PMP	$P\Box A \to \Box PA$	$adPMP$
OMP	$O\Box A \to \Box OA$	$adOMP$

Table 27

Theorem	System
$\Pi x(Rx \to (\Pi y \mathcal{W}_x B \leftrightarrow \mathcal{W}_x \Pi y B))$	Every
$\Pi x(Rx \to (\Sigma y \mathcal{A}_x B \leftrightarrow \mathcal{A}_x \Sigma y B))$	Every
$\Pi x(Rx \to (\mathcal{A}_x \Pi y B \to \Pi y \mathcal{A}_x B))$	Every
$\Pi x(Rx \to (\Sigma y \mathcal{W}_x B \to \mathcal{W}_x \Sigma y B))$	Every

Table 28

Theorem	System
$\Pi x(Rx \to (\Box(A \to B) \to (\mathcal{W}_x A \to \mathcal{W}_x B)))$	$M\mathcal{W}$
$\Pi x(Rx \to (\Box(A \to B) \to (\mathcal{A}_x A \to \mathcal{A}_x B)))$	$M\mathcal{W}$
$\Pi x(Rx \to (\Box(A \to B) \to (\mathcal{W}_x \neg B \to \mathcal{W}_x \neg A)))$	$M\mathcal{W}$

$\Pi x(Rx \rightarrow (\Box(A \leftrightarrow B) \rightarrow (\mathcal{W}_x A \leftrightarrow \mathcal{W}_x B)))$	$M\mathcal{W}$
$\Pi x(Rx \rightarrow (\Box(A \leftrightarrow B) \rightarrow (\mathcal{A}_x A \leftrightarrow \mathcal{A}_x B)))$	$M\mathcal{W}$
$\Pi x(Rx \rightarrow (\Box(A \leftrightarrow B) \rightarrow (\mathcal{W}_x \neg A \leftrightarrow \mathcal{W}_x \neg B)))$	$M\mathcal{W}$
$\Pi x(Rx \rightarrow (\Box(A \leftrightarrow B) \rightarrow (\neg \mathcal{W}_x A \leftrightarrow \neg \mathcal{W}_x B)))$	$M\mathcal{W}$

Table 29

Theorem	System
$OA \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$	$O\mathcal{W}$
$\Pi x(Rx \rightarrow \mathcal{W}_x A) \rightarrow OA$	$\mathcal{W}OHW\Sigma R$
$OA \leftrightarrow \Pi x(Rx \rightarrow \mathcal{W}_x A)$	$O\mathcal{W}\mathcal{W}OHW\Sigma R$
$PB \rightarrow \Pi x(Rx \rightarrow \mathcal{A}_x B)$	$\mathcal{W}OHW$
$\Pi x(Rx \rightarrow \mathcal{A}_x B) \rightarrow PB$	$O\mathcal{W}HW\Sigma R$
$PB \leftrightarrow \Pi x(Rx \rightarrow \mathcal{A}_x B)$	$\mathcal{W}OO\mathcal{W}HW\Sigma R$
$FA \rightarrow \Pi x(Rx \rightarrow \mathcal{W}_x \neg A)$	$O\mathcal{W}$
$\Pi x(Rx \rightarrow \mathcal{W}_x \neg A) \rightarrow FA$	$\mathcal{W}OHW\Sigma R$
$FA \leftrightarrow \Pi x(Rx \rightarrow \mathcal{W}_x \neg A)$	$O\mathcal{W}\mathcal{W}OHW\Sigma R$

Table 30

5.1. Examples: Valid arguments and valid and invalid formulas

In this section, I will consider one example of a valid argument, one example of a valid sentence and one example of an invalid sentence. I will show that *argument* 3 described in the introduction is valid (in the class of all models that satisfy $C - M\mathcal{W}$). This illustrates one of the possible applications of the systems that are introduced in this paper, namely as a tool in the analysis and evaluation of various arguments. *Argument* 3 is intuitively valid, but it seems impossible to prove this in any other systems in the literature. Nonetheless, we can prove that the conclusion is derivable from the premises in all systems in this paper that include $T - M\mathcal{W}$. Since the smallest boulesic system that includes $T - M\mathcal{W}$ is sound with respect to the class of all models that satisfy $C - M\mathcal{W}$, the argument is valid in the

class of all models that satisfy $C - M\mathcal{W}$. Hence, we seem to need systems of the kind developed in this paper.

$Argument$ 3 can be symbolised in the following way. $\Pi x(Px \rightarrow \mathcal{W}_x Mx)$ (for every x: if x is a person in the class, then x wants it to be the case that x passes the exam), Ps (Sandra is a person in the class), $\Box(Ms \rightarrow Ss)$ (it is necessary that Sandra passes the exam only if she studies hard), $Rs \rightarrow \mathcal{W}_s Ss$ (if Sandra is perfectly rational, she wants to study hard). To prove that the conclusion is derivable from the premises, we construct a semantic tableau that begins with all premises and the negation of the conclusion. Since this tableau is closed, it constitutes a derivation of the conclusion from the premises in the smallest boulesic system that includes $T - M\mathcal{W}$. Hence, the conclusion follows from the premises in the class of all models that satisfy $C - M\mathcal{W}$ (by the soundness theorems in Section 6). Here is our proof. ('MP' stands for the derived rule 'Modus Ponens.')

(1) $\Pi x(Px \rightarrow \mathcal{W}_x Mx)$, 0

(2) Ps, 0

(3) $\Box(Ms \rightarrow Ss)$, 0

(4) $\neg(Rs \rightarrow \mathcal{W}_s Ss)$, 0

(5) Rs, 0 $[4, \neg\rightarrow]$

(6) $\neg\mathcal{W}_s Ss$, 0 $[4, \neg\rightarrow]$

(7) $\mathcal{A}_s \neg Ss$, 0 $[5, 6, \neg\mathcal{W}]$

(8) $Ps \rightarrow \mathcal{W}_s Ms$, 0 $[1, \Pi \; [s/x]]$

(9) $\mathcal{W}_s Ms$, 0 $[2, 8, MP]$

(10) $0 As 1$ $[5, 7, \mathcal{A}]$

(11) $\neg Ss$, 1 $[5, 7, \mathcal{A}]$

(12) Ms, 1 $[5, 9, 10, \mathcal{W}]$

(13) $0 r 1$ $[10, T - M\mathcal{W}]$

(14) $Ms \rightarrow Ss$, 1 $[3, 13, \Box]$

(15) Ss, 1 $[12, 14, MP]$

(16) $*$ $[11, 15]$

Let us now turn to our valid sentence. In the introduction, we considered several interpretations of the so-called hypothetical imperative. One of the readings was (7): $U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \Box(A \rightarrow B)) \rightarrow OB))$. Intuitively,

this sentence says that it is absolutely necessary that if x is perfectly rational, then if x wants it to be the case that A and it is necessary that A only if B then it ought to be the case that B. Here is an instance of this schema: if x is perfectly rational and x wants to achieve end E and it is necessary that x achieves end E only if x does action A then x ought to do action A. Or more concretely, if x is perfectly rational then if x wants to become a doctor of philosophy (sometime in the future) and it is necessary that x will become a doctor of philosophy (sometime in the future) only if x writes a dissertation then x ought to write a dissertation. I will now show that (7) is a theorem in every boulesic-deontic system that includes the rules $T - \mathcal{W}O$, $T - HW$ and $T - M\mathcal{W}$. Here is our tableau proof:[35]

$$(1)\ \neg U\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \square(A \rightarrow B)) \rightarrow OB)),\ 0$$
$$(2)\ M\neg\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \square(A \rightarrow B)) \rightarrow OB)),\ 0\ [1, \neg U]$$
$$(3)\ \neg\Pi x(Rx \rightarrow ((\mathcal{W}_x A \wedge \square(A \rightarrow B)) \rightarrow OB)),\ 1\ [2, M]$$
$$(4)\ \Sigma x\neg(Rx \rightarrow ((\mathcal{W}_x A \wedge \square(A \rightarrow B)) \rightarrow OB)),\ 1\ [3, \neg\Pi]$$
$$(5)\ \neg(Rc \rightarrow ((\mathcal{W}_c A[c/x] \wedge \square(A[c/x] \rightarrow B[c/x])) \rightarrow OB[c/x])),\ 1\ [4, \Sigma]$$
$$(6)\ Rc,\ 1\ [5, \neg\rightarrow]$$
$$(7)\ \neg((\mathcal{W}_c A[c/x] \wedge \square(A[c/x] \rightarrow B[c/x])) \rightarrow OB[c/x]),\ 1\ [5, \neg\rightarrow]$$
$$(8)\ \mathcal{W}_c A[c/x] \wedge \square(A[c/x] \rightarrow B[c/x]),\ 1\ [7, \neg\rightarrow]$$
$$(9)\ \neg OB[c/x],\ 1\ [7, \neg\rightarrow]$$
$$(10)\ \mathcal{W}_c A[c/x],\ 1\ [8, \wedge]$$
$$(11)\ \square(A[c/x] \rightarrow B[c/x]),\ 1\ [8, \wedge]$$
$$(12)\ P\neg B[c/x],\ 1\ [9, \neg O]$$
$$(13)\ 1s2\ [12, P]$$
$$(14)\ \neg B[c/x],\ 2\ [12, P]$$
$$(15)\ 1Ad2\ [13, T - \mathcal{W}O]$$
$$(16)\ 1Ac2\ [15, T - HW]$$
$$(17)\ A[c/x],\ 2\ [6, 10, 16, \mathcal{W}]$$
$$(18)\ 1r2\ [16, T - M\mathcal{W}]$$
$$(19)\ A[c/x] \rightarrow B[c/x],\ 2\ [11, 18, \square]$$

[35]   In a strict sense, this is not a proof, but a proof schema. For it includes expressions such as $A[c/x]$. However, this schema shows that any proof of this form is correct.
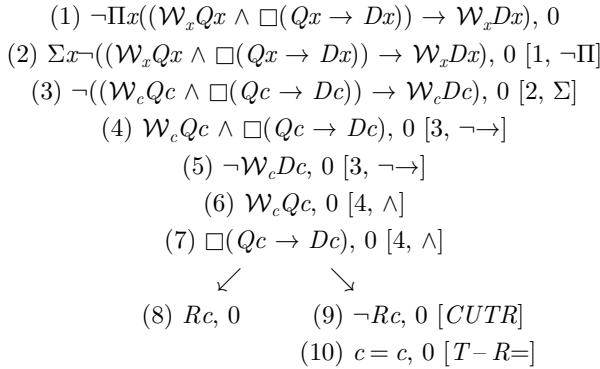
$$(20)\ B[c/x],\ 2\ [17,\ 19,\ MP]$$
$$(21)\ *\ [14,\ 20]$$

The tableau above is closed. Hence, $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to OB))$ is a theorem in every boulesic-deontic system that includes the rules $T - \mathcal{W}O$, $T - HW$ and $T - M\mathcal{W}$. It follows, by the soundness results in Section 6, that $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to OB))$ is valid in the class of all models that satisfy $C - \mathcal{W}O$, $C - HW$ and $C - M\mathcal{W}$. Even though $U\Pi x(Rx \to ((\mathcal{W}_x A \land \Box(A \to B)) \to OB))$ is valid in some systems, $U\Pi x((\mathcal{W}_x A \land \Box(A \to B)) \to OB)$ is not a theorem in any system in this paper (as I mentioned in the introduction). This is as it should be since this formula has countless counterintuitive consequences. Consider, for example, the following 'instance': 'If c wants to destroy the City Hall and it is necessary that c uses a bomb to destroy the City Hall, then c ought to use a bomb to destroy the City Hall.' Suppose that the antecedent is true. Then c ought to use a bomb to destroy the City Hall. But this is absurd.

Now I will show how it is possible to use semantic tableaux to prove that a sentence is not valid and how it is possible to use open complete branches to read off countermodels. Consider the following sentence:

> For every individual x, if x wants to quench her thirst and it is necessary that x quenches her thirst only if x drinks some water, then x wants to drink some water.

This sentence can be symbolised in the following way: $\Pi x((\mathcal{W}_x Qx \land \Box(Qx \to Dx)) \to \mathcal{W}_x Dx)$, where Qx says that x quenches her thirst and Dx says that x drinks some water. I will show that this formula is *not* valid in the class of all models. To establish this, I will show that the formula is not a theorem in our weakest system. By the completeness theorems in Section 6, it follows that the sentence is not valid in the class of all models. I will use an open branch in a complete tree for the formula to read off a countermodel and I will verify that this model is a countermodel to the formula. In fact, it is possible to prove that the sentence is not a theorem in any system in this paper. Consequently, it is possible to show that the formula is not valid in any class of models (in this paper). It is left to the reader to verify this claim. Here is our tableau:

(1) $\neg\Pi x((\mathcal{W}_x Qx \wedge \Box(Qx \to Dx)) \to \mathcal{W}_x Dx), 0$

(2) $\Sigma x \neg((\mathcal{W}_x Qx \wedge \Box(Qx \to Dx)) \to \mathcal{W}_x Dx), 0 \; [1, \neg\Pi]$

(3) $\neg((\mathcal{W}_c Qc \wedge \Box(Qc \to Dc)) \to \mathcal{W}_c Dc), 0 \; [2, \Sigma]$

(4) $\mathcal{W}_c Qc \wedge \Box(Qc \to Dc), 0 \; [3, \neg\to]$

(5) $\neg\mathcal{W}_c Dc, 0 \; [3, \neg\to]$

(6) $\mathcal{W}_c Qc, 0 \; [4, \wedge]$

(7) $\Box(Qc \to Dc), 0 \; [4, \wedge]$

$\swarrow \qquad \searrow$

(8) $Rc, 0 \qquad (9)$ $\neg Rc, 0 \; [CUTR]$

(10) $c = c, 0 \; [T\text{-}R{=}]$

It is possible to extend the left branch in this tree. Nevertheless, at this stage we cannot apply any more rules to the right branch, which is open (and complete). It follows that the whole tableau is open (and complete). Hence, $\Pi x((\mathcal{W}_x Qx \wedge \Box(Qx \to Dx)) \to \mathcal{W}_x Dx)$ is *not* a theorem in our weakest system. Consequently, the formula is *not* valid in the class of all models (by the completeness results in Section 6).

Let us verify this conclusion. We can use the right branch to read off a countermodel, $\mathcal{M}$, since this branch is open and complete. The matrix of $\mathcal{W}_c Qc$ is $\mathcal{W}_{x_1} Qx_2$ and the matrix of $\mathcal{W}_c Dc$ is $\mathcal{W}_{x_1} Dx_2$.

$W = \{\omega_0\}$, $D = \{[c]\}$, $v(c) = [c]$, and the extensions of Q and D are empty in ω_0. $\mathfrak{R}$, $\mathfrak{A}$ (and $\mathfrak{S}$) are empty. $v_{\omega_0}(\mathcal{W}_{x_1} Qx_2)$ is the extension of $\mathcal{W}_{x_1} Qx_2$ in ω_0 and $v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$ is the extension of $\mathcal{W}_{x_1} Dx_2$ in ω_0. If $\neg Ra_m, i$ is on the branch $\mathcal{B}$ and M is an n-place matrix with instantiations on the branch (where x_m is the first free variable in M and a_m is the constant in $M[a_1, \ldots, a_n/x_1, \ldots, x_n]$ that replaces x_m), then $\langle[a_1], \ldots, [a_n]\rangle$ is an element of $v_{\omega_i}(M)$ iff $M[a_1, \ldots, a_n/x_1, \ldots, x_n], i$ occurs on $\mathcal{B}$.

$\neg Rc, 0$ is on the branch, while $\mathcal{W}_{x_1} Dx_2[c, c/x_1, x_2], 0 \; (=\mathcal{W}_c Dc, 0)$ is not on the branch. x_1 is the first free variable in $\mathcal{W}_{x_1} Dx_2$ and c is the constant in $\mathcal{W}_{x_1} Dx_2[c, c/x_1, x_2]$ that replaces x_1. Consequently, $\langle[c], [c]\rangle$ is not an element in $v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$ ($v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$ is empty). Rc is false in ω_0 in $\mathcal{M}$, for $\neg Rc, 0$ is on $\mathcal{B}$. If Rc is false in ω_0 in $\mathcal{M}$, then $\mathcal{W}_{x_1} Dx_2[c, c/x_1, x_2]$ is true in ω_0 in $\mathcal{M}$ iff $\langle v(c), v(c)\rangle$ is in $v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$. Hence, $\mathcal{W}_{x_1} Dx_2[c, c/x_1, x_2]$ is true in ω_0 in $\mathcal{M}$ iff $\langle v(c), v(c)\rangle$ is in $v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$. $\langle v(c), v(c)\rangle$ is not in $v_{\omega_0}(\mathcal{W}_{x_1} Dx_2)$.

Consequently, it is not the case that $\mathcal{W}_{x_1}Dx_2[c,\ c/x_1,\ x_2]$ is true in ω_0 in $\mathcal{M}$. $\mathcal{W}_{x_1}Dx_2[c,\ c/x_1,\ x_2] = \mathcal{W}_cDc$. It follows that it is not the case that $\mathcal{W}_cDc$ is true in ω_0 in $\mathcal{M}$, that is, $\mathcal{W}_cDc$ is false in ω_0 in $\mathcal{M}$.

$\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2]$, 0 (that is, $\mathcal{W}_cQc$, 0) is on the branch. x_1 is the first free variable in $\mathcal{W}_{x_1}Qx_2$ and c is the constant in $\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2]$ that replaces x_1. Hence, $\langle[c],\ [c]\rangle$ is an element in $v_{\omega_0}(\mathcal{W}_{x_1}Qx_2)$. If Rc is false in ω_0 in $\mathcal{M}$, then $\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2]$ is true in ω_0 in $\mathcal{M}$ iff $\langle v(c),\ v(c)\rangle$ is in $v_{\omega_0}(\mathcal{W}_{x_1}Qx_2)$. Accordingly, $\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2]$ is true in ω_0 in $\mathcal{M}$ iff $\langle v(c),\ v(c)\rangle$ is in $v_{\omega_0}(\mathcal{W}_{x_1}Qx_2)$. $\langle v(c),\ v(c)\rangle$ is in $v_{\omega_0}(\mathcal{W}_{x_1}Qx_2)$. Therefore, $\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2]$ is true in ω_0 in $\mathcal{M}$. $\mathcal{W}_{x_1}Qx_2[c,\ c/x_1,\ x_2] = \mathcal{W}_cQc$. Consequently, $\mathcal{W}_cQc$ is true in ω_0 in $\mathcal{M}$.

Since no possible world is alethically accessible from ω_0 in $\mathcal{M}$, $\square(Qc \rightarrow Dc)$ is true in ω_0 in $\mathcal{M}$.

We have established that $\mathcal{W}_cQc$ is true in ω_0 in $\mathcal{M}$ and that $\square(Qc \rightarrow Dc)$ is true in ω_0 in $\mathcal{M}$. Accordingly, $\mathcal{W}_cQc \wedge \square(Qc \rightarrow Dc)$ is true in ω_0 in $\mathcal{M}$. Furthermore, we have shown that $\mathcal{W}_cDc$ is false in ω_0 in $\mathcal{M}$. It follows that $(\mathcal{W}_cQc \wedge \square(Qc \rightarrow Dc)) \rightarrow \mathcal{W}_cDc$ is false in ω_0 in $\mathcal{M}$. Since $[c]$ is an object in the domain, we conclude that $\Pi x((\mathcal{W}_xQx \wedge \square(Qx \rightarrow Dx)) \rightarrow \mathcal{W}_xDx)$ is false in ω_0 in $\mathcal{M}$. It follows that this formula is not valid in the class of all models. This result is intuitively plausible. If some individual is not perfectly rational, it is possible that she wants something, A, without wanting the necessary means to A. This is compatible with the proposition that several other versions of the hypothetical imperative are valid (in some models) (see above, the introduction and Section 3.6).

6. Soundness and completeness theorems

In this section, I will prove that every system in this essay is sound and complete with respect to its semantics. The concepts of soundness and completeness are defined as usual (see, for example, Priest, 2008). Many steps in the proofs are easy modifications of existing proofs. However, due to the presence of the boulesic operators in our language, some steps require some new techniques.

Lemma 11 (*Locality*): *Let* $\mathcal{M}_1 = \langle D, W, \mathfrak{R}, \mathfrak{A}, \mathfrak{S}, v_1 \rangle$ *and* $\mathcal{M}_2 = \langle D, W,$ $\mathfrak{R}, \mathfrak{A}, \mathfrak{S}, v_2 \rangle$ *be two supplemented models (the lemma for unsupplemented models is similar). The language of the two, which we call* $\mathcal{L}$, *is the same, for they have the same domain. Let A be any closed formula of* $\mathcal{L}$ *such that* v_1 *and* v_2 *agree on the denotations of all the predicates, constants and matrices in it. Then for all* $\omega \in W$: $v_{1\omega}(A) = v_{2\omega}(A)$.

Proof. The proof is by recursion on the sentences in our language. 'the IH' refers to the induction hypothesis.

Atomic formulas. $v_{1\omega}(Pa_1 \ldots a_n) = 1$ iff $\langle v_1(a_1), \ldots, v_1(a_n) \rangle \in v_{1\omega}(P)$ iff $\langle v_2(a_1), \ldots, v_2(a_n) \rangle \in v_{2\omega}(P)$ iff $v_{2\omega}(Pa_1 \ldots a_n) = 1$.

Suppose that $v_{1\omega}(Ra_m) = 0$, that M is a matrix where x_m is the first free variable in M and that a_m is the constant in $M[a_1, \ldots, a_n/\vec{x}]$ that replaces x_m. Then: $v_{2\omega}(Ra_m) = 0$ and $v_{1\omega}(M[a_1, \ldots, a_n/\vec{x}] = 1$ iff $\langle v_1(a_1), \ldots, v_1(a_n) \rangle$ $\in v_{1\omega}(M)$ iff $\langle v_2(a_1), \ldots, v_2(a_n) \rangle \in v_{2\omega}(M)$ iff $v_{2\omega}(M[a_1, \ldots, a_n/\vec{x}]) = 1$.

Truth-functional connectives. Straightforward.

($\square$). $v_{1\omega}(\square B) = 1$ iff for all ω' such that $\mathfrak{R}\omega\omega'$, $v_{1\omega'}(B) = 1$ iff for all ω' such that $\mathfrak{R}\omega\omega'$, $v_{2\omega'}(B) = 1$ [the IH] iff $v_{2\omega}(\square B) = 1$.

The cases for the other alethic and deontic operators are similar.

($\mathcal{W}_c B$). A is of the form $\mathcal{W}_c B$. Suppose $v_{1\omega}(\mathcal{W}_c B) = 1$. We have two cases: $v_{1\omega}(Rc) = 0$ or $v_{1\omega}(Rc) = 1$. Suppose $v_{1\omega}(Rc) = 0$. Then $v_{2\omega}(Rc) = 0$. Hence, $v_{2\omega}(\mathcal{W}_c B) = 1$. And vice versa. Suppose $v_{1\omega}(Rc) = 1$. Then for all ω' such that $\mathfrak{A}v_1(c)\omega\omega'$: $v_{1\omega'}(B) = 1$. Accordingly, for all ω' such that $\mathfrak{A}v_2(c)\omega\omega'$: $v_{2\omega'}(B) = 1$ [by assumption and the IH]. Furthermore, $v_{2\omega}(Rc)$ $= 1$. Hence, $v_{2\omega}(\mathcal{W}_c B) = 1$. And vice versa. Consequently, $v_{1\omega}(\mathcal{W}_c B) = 1$ iff $v_{2\omega}(\mathcal{W}_c B) = 1$.

The case for $\mathcal{A}_c B$ is similar.

(Π). $v_{1\omega}(\Pi x B) = 1$ iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_{1\omega}(B[k_d/x]) = 1$ iff for all $k_d \in$ $\mathcal{L}(\mathcal{M})$, $v_{2\omega}(B[k_d/x]) = 1$ [by the IH, and the fact that $v_{1\omega}(k_d) = v_{2\omega}(k_d) = d$] iff $v_{2\omega}(\Pi x B) = 1$.

The case for the particular quantifier is similar. ∎

Lemma 12 (*Denotation*): *Let* $\mathcal{M} = \langle D, W, \mathfrak{R}, \mathfrak{A}, \mathfrak{S}, v \rangle$ *be any supplemented model (the lemma for unsupplemented models is similar). Let A be any formula of* $\mathcal{L}(\mathcal{M})$ *with at most one free variable, x, and a and b be any two*

constants such that $v(a) = v(b)$. Then for any $\omega \in W$: $v_\omega(A[a/x]) = v_\omega(A[b/x])$.

Proof. The proof is by induction on the complexity of A.

Atomic formulas. (To illustrate, we assume that the formula has one occurrence of 'a,' distinct from each a_i.) $v_\omega(Pa_1 \ldots a \ldots a_n) = 1$ iff $\langle v(a_1), \ldots, v(a), \ldots, v(a_n) \rangle \in v_\omega(P)$ iff $\langle v(a_1), \ldots, v(b), \ldots, v(a_n) \rangle \in v_\omega(P)$ iff $v_\omega(Pa_1 \ldots b \ldots a_n) = 1$.

Suppose $v_\omega(Ra_m) = 0$, that M is a matrix where x_m is the first free variable in M and that a_m is the constant in $M[a_1, \ldots, a, \ldots, a_n/\vec{x}]$ $(M[a_1, \ldots, b, \ldots, a_n/\vec{x}])$ that replaces x_m. (To illustrate, we assume that the formula has one occurrence of 'a' distinct from each a_i and that a_m is not a (b).) Then: $v_\omega(M[a_1, \ldots, a, \ldots, a_n/\vec{x}]) = 1$ iff $\langle v(a_1), \ldots, v(a), \ldots, v(a_n) \rangle \in v_\omega(M)$ iff $\langle v(a_1), \ldots, v(b), \ldots, v(a_n) \rangle \in v_\omega(M)$ iff $v_\omega(M[a_1, \ldots, b, \ldots, a_n/\vec{x}]) = 1$.

Truth-functional connectives. Straightforward.

($\square$). $v_\omega(\square B[a/x]) = 1$ iff for all ω' such that $\mathfrak{R}\omega\omega'$, $v_{\omega'}(B[a/x]) = 1$ iff for all ω' such that $\mathfrak{R}\omega\omega'$, $v_{\omega'}(B[b/x]) = 1$ [the IH] iff $v_\omega(\square B[b/x]) = 1$.

The arguments for the other primitive alethic and deontic operators are similar.

($\mathcal{W}_t$). A is of the form $\mathcal{W}_t B$. Either $v_\omega(Rt) = 1$ or $v_\omega(Rt) = 0$. We have already shown that the result holds if $v_\omega(Rt) = 0$. Accordingly, suppose that $v_\omega(Rt) = 1$. Since x is the only free variable, t cannot be a variable distinct from x. So, t is either x or a constant. Suppose t is x. Then $v_\omega(\mathcal{W}_x B[a/x]) = 1$ iff $v_\omega(\mathcal{W}_a B[a/x]) = 1$ iff for all ω' such that $\mathfrak{A}v(a)\omega\omega'$, $v_{\omega'}(B[a/x]) = 1$ iff for all ω' such that $\mathfrak{A}v(b)\omega\omega'$, $v_{\omega'}(B[b/x]) = 1$ [by the fact that $v(a) = v(b)$ and the IH] iff $v_\omega(\mathcal{W}_b B[b/x]) = 1$ iff $v_\omega(\mathcal{W}_x B[b/x]) = 1$. Suppose t is a constant, say c. Then $v_\omega(\mathcal{W}_c B[a/x]) = 1$ iff for all ω' such that $\mathfrak{A}v(c)\omega\omega'$, $v_{\omega'}(B[a/x]) = 1$ iff for all ω' such that $\mathfrak{A}v(c)\omega\omega'$, $v_{\omega'}(B[b/x]) = 1$ [by the IH] iff $v_\omega(\mathcal{W}_c B[b/x]) = 1$.

The case for $\mathcal{A}_t$ is similar.

(Π). Let A be of the form $\Pi y B$. If $x = y$, then $A[a/x] = A[b/x] = A$, so the result is trivial. Accordingly, suppose that x and y are distinct. Then, $(\Pi y B)[b/x] = \Pi y(B[b/x])$ and $(B[b/x])[a/y] = (B[a/y])[b/x]$. $v_\omega((\Pi y B)[a/x])$

$= 1$ iff $v_\omega(\Pi y(B[a/x])) = 1$ iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega((B[a/x])[k_d/y]) = 1$ iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega((B[k_d/y])[a/x]) = 1$ iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega((B[k_d/y])[b/x]) = 1$ [the IH] iff for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_\omega((B[b/x])[k_d/y]) = 1$ iff $v_\omega(\Pi y(B[b/x])) = 1$ iff $v_\omega((\Pi yB)[b/x]) = 1$.

The case for the particular quantifier (Σ) is similar. ∎

6.1. Soundness theorem

Let $\mathcal{M} = \langle D, W, \mathfrak{R}, \mathfrak{A}, \mathfrak{S}, v \rangle$ be any (supplemented) model and $\mathcal{B}$ any branch of a tableau. Then $\mathcal{B}$ is satisfiable in $\mathcal{M}$ iff there is a function f from $0, 1, 2, \ldots$ to W such that

(i) A is true in $f(i)$ in $\mathcal{M}$, for every node A,i on $\mathcal{B}$,

(ii)   if irj is on $\mathcal{B}$, then $\mathfrak{R}f(i)f(j)$ in $\mathcal{M}$,

(iii)   if isj is on $\mathcal{B}$, then $\mathfrak{S}f(i)f(j)$ in $\mathcal{M}$, and

(iv)   if $iAcj$ is on $\mathcal{B}$, then $\mathfrak{A}v(c)f(i)f(j)$ in $\mathcal{M}$.

If these conditions are fulfilled, we say that f shows that $\mathcal{B}$ is satisfiable in $\mathcal{M}$.

Lemma 13 (*Soundness Lemma*): *Let $\mathcal{B}$ be any branch of a tableau and $\mathcal{M}$ be any model. If $\mathcal{B}$ is satisfiable in $\mathcal{M}$ and a tableau rule is applied to it, then there is a model $\mathcal{M}'$ and an extension of $\mathcal{B}$, $\mathcal{B}'$, such that $\mathcal{B}'$ is satisfiable in $\mathcal{M}'$.*

Proof. The proof is by induction on the height of the derivation. Let f be a function that shows that the branch $\mathcal{B}$ is satisfiable in $\mathcal{M}$.

Connectives and the modal operators. Straightforward.

($\mathcal{W}$). Suppose that Rc,i, $\mathcal{W}_cB,i$, and $iAcj$ are on $\mathcal{B}$, and that we apply the $\mathcal{W}$-rule. Then we get an extension of $\mathcal{B}$ that includes B,j. Since $\mathcal{B}$ is satisfiable in $\mathcal{M}$, $\mathcal{W}_cB$ is true in $f(i)$ and Rc is true in $f(i)$. Moreover, for any i and j such that $iAcj$ is on $\mathcal{B}$, $\mathfrak{A}v(c)f(i)f(j)$. Hence by the truth conditions for $\mathcal{W}_cB$, B is true in $f(j)$.

($\mathcal{A}$). Suppose that Rc,i, $\mathcal{A}_cB,i$ are on $\mathcal{B}$ and that we apply the $\mathcal{A}$-rule to get an extension of $\mathcal{B}$ that includes nodes of the form $iAcj$ and B,j. Since $\mathcal{B}$ is satisfiable in $\mathcal{M}$, $\mathcal{A}_cB$ is true in $f(i)$ and Rc is true in $f(i)$. Hence, for some w in W, $\mathfrak{A}v(c)f(i)w$ and B is true in w [by the truth conditions for

$\mathcal{A}_cB$ and the fact that Rc is true in $f(i)$]. Let f' be the same as f except that $f'(j) = w$. Since f and f' differ only at j, f' shows that $\mathcal{B}$ is satisfiable in $\mathcal{M}$. Moreover, by definition $\mathfrak{A}v(c)f'(i)f'(j)$, and B is true in $f'(j)$.

($\neg\mathcal{W}$) and ($\neg\mathcal{A}$). Similar.

(Π). Suppose that $\Pi xA,i$ is on $\mathcal{B}$ and that we apply the Π-rule to get an extension of $\mathcal{B}$ that includes a node of the form $A[a/x],i$. $\mathcal{M}$ makes ΠxA true in $f(i)$. For $\mathcal{B}$ is satisfiable in $\mathcal{M}$. Hence, $A[k_d/x]$ is true in $f(i)$ in $\mathcal{M}$, for all $k_d \in \mathcal{L}(\mathcal{M})$. Let d be such that $v(a) = v(k_d)$. By the Denotation Lemma, $A[a/x]$ is true in $f(i)$ in $\mathcal{M}$. Accordingly, we can take $\mathcal{M}'$ to be $\mathcal{M}$.

(Σ). Suppose that $\Sigma xA,i$ is on $\mathcal{B}$ and that we apply the Σ-rule to get an extension of $\mathcal{B}$ that includes a node of the form $A[c/x],i$ (where c is new). Since $\mathcal{B}$ is satisfiable in $\mathcal{M}$, ΣxA is true in $f(i)$ in $\mathcal{M}$. Hence, there is some $k_d \in \mathcal{L}(\mathcal{M})$ such that $\mathcal{M}$ makes $A[k_d/x]$ true in $f(i)$. Let $\mathcal{M}' = \langle D,\ W,\ \mathfrak{R}, \mathfrak{A},\ \mathfrak{S},\ v' \rangle$ be the same as $\mathcal{M}$ except that $v'(c) = d$. Since c does not occur in $A[k_d/x]$, $A[k_d/x]$ is true in $f(i)$ in $\mathcal{M}'$, by the Locality Lemma. By the Denotation Lemma and the fact that $v'(c) = d = v'(k_d)$, $A[c/x]$ is true in $f(i)$ in $\mathcal{M}'$. Furthermore, $\mathcal{M}'$ makes all other formulas on the branch true at their respective worlds as well, by the Locality Lemma. For c does not occur in any other formula on the branch.

($\neg\Pi$) and ($\neg\Sigma$). Straightforward.

Accessibility rules. I will go through three examples to illustrate the method.

($T - M\mathcal{W}$). Suppose we have $iAcj$ on $\mathcal{B}$, and that we apply ($T - M\mathcal{W}$) to obtain an extension of $\mathcal{B}$ that includes irj. Since $\mathcal{B}$ is satisfiable in $\mathcal{M}$, $\mathfrak{A}v(c)f(i)f(j)$. It follows that $\mathfrak{R}f(i)f(j)$, since $\mathcal{M}$ satisfies the condition $C - M\mathcal{W}$.

($T - \mathcal{W}C$). Suppose that i is on $\mathcal{B}$, and that we apply ($T - \mathcal{W}C$) to give an extended branch containing $iAcj$ and irj, where j is new. Since $\mathcal{B}$ is satisfiable in $\mathcal{M}$, $f(i)$ is in W and $v(c)$ is in D. Hence, for some w in W, $\mathfrak{A}v(c)f(i)w$ and $\mathfrak{R}f(i)w$, since $\mathcal{M}$ satisfies condition $C - \mathcal{W}C$. Let f' be the same as f except that $f'(j) = w$. Since j does not occur on $\mathcal{B}$, f' shows that $\mathcal{B}$ is satisfiable in $\mathcal{M}$. Moreover, $\mathfrak{A}v(c)f'(i)f'(j)$ and $\mathfrak{R}f'(i)f'(j)$ by construction. Hence, f' shows that the extension of $\mathcal{B}$ is satisfiable in $\mathcal{M}$.

($T - A=$). Suppose we have $a = b,i$ and $jAak$ on a branch and that we apply ($T - A=$) to obtain an extension that includes $jAbk$. Since f shows

that the branch is satisfiable in $\mathcal{M}$, $a = b$ is true in $f(i)$ and $\mathfrak{A}v(a)f(j)f(k)$ in $\mathcal{M}$. Accordingly, $v(a) = v(b)$. Hence, $\mathfrak{A}v(b)f(j)f(k)$, and we may take $\mathcal{M}'$ to be $\mathcal{M}$. ∎

Theorem 14 (*Soundness Theorem*): *Every system S in this paper is sound with respect to its semantics.*

Proof. Suppose that B does not follow from Γ in $\mathbf{M}$, where $\mathbf{M}$ is the class of models that corresponds to S. Then every premise in Γ is true and the conclusion B false at some world w in some model in $\mathbf{M}$. Consider an S-tableau whose initial list consists of $A,0$ for every $A \in \Gamma$ and $\neg B,0$, where '0' refers to w. Then the initial list is satisfiable in $\mathbf{M}$. Every time we apply a rule to this list it produces at least one extension that is satisfiable in $\mathbf{M}$ (by the Soundness Lemma). Hence, we can find a whole branch such that every initial section of this branch is satisfiable in $\mathbf{M}$. This branch cannot be closed, for then some sentence would be both true and false in some possible world in some model in $\mathbf{M}$. Accordingly, the tableau is open. Consequently, B is not derivable from Γ in S. In conclusion, if B is derivable from Γ in S, then B follows from Γ in $\mathbf{M}$. ∎

6.2. Completeness theorem

In this section, I will show that every system in this paper is complete with respect to its semantics. However, first we must define some important concepts.

Informally, a complete tableau is a tableau where every rule that can be applied has been applied. Since different systems include different rules, a tableau can be complete in one system even though it is not complete in another system. Furthermore, since the tableau rules may be applied in different orders, there may be several different (complete) tableaux for the same sentence or set of sentences in one and the same system, some longer than others, some shorter. To produce a complete tableau (in our completeness proofs) we shall use the following method.[36] (1) For every open branch

---

[36]   Note that it is often possible to produce shorter proofs by using some more intuitive method instead.

on the tree, we shall do the following. We shall begin at its root and move towards its tip. We shall apply any rule that produces something that has not appeared on the branch before. For example, there is no point in applying Σ more than once to a node of the form $\Sigma x A, i$. We shall not apply any rules to a branch that is already closed. Some rules can be applied more than once, for example, $\square$ and Π. When we arrive at a node of the form $\square A, i$ and it is possible to apply $\square$ several times, then we shall make all applications at once and we shall do the same for all similar nodes. (2) When we have extended all open branches on the tree in this way, we shall repeat the procedure. Some rules introduce new 'possible worlds,' for example $T - aD$ and $T - \mathcal{W}C$. If a rule introduces a new possible world, then we shall apply it once at the tip of every open branch at the end of every cycle (that is, when we have gone through all nodes). If a system includes several different rules that introduce new possible worlds $(R1, R2, \ldots)$, we shall alternate between them. The first time, we shall use $R1$ once; the second time we shall use $R2$ once, etc. Before we conclude a cycle and start to move through all nodes again, we shall apply $CUTR$. We shall split the end of every open branch in the tree and add Rc, i to the left node and $\neg Rc, i$ to the right node, for every constant c that occurs as an index to some boulesic operator on the tree and i on the branch. The tableau is *incomplete* precisely when there is still something to do according to this method. A tableau is *complete* iff it is not incomplete.

Definition 15 (*Induced Model*): *Let $\mathcal{B}$ be an open complete branch of a tableau, let i, j, k, etc. be numbers on $\mathcal{B}$, and let I be the set of numbers on $\mathcal{B}$. Furthermore, let C be the set of all constants on $\mathcal{B}$. Define $a \sim b$ to mean that $a = b, 0$ is on the branch. $a \sim b$ is obviously an equivalence relation. Let $[a]$ be the equivalence class of a under $\sim$. The (supplemented) model, $\mathcal{M} = \langle D, W, \mathfrak{R}, \mathfrak{S}, \mathfrak{A}, v \rangle$, induced by $\mathcal{B}$ is defined as follows. $D = \{[a]: a \in C\}$ (or, if $C = \emptyset$, $D = \{o\}$ for an arbitrary o). (o is not in the extension of anything.) $W = \{\omega_i: i \text{ occurs on } \mathcal{B}\}$, $\mathfrak{R}\omega_i\omega_j$ iff irj occurs on $\mathcal{B}$, $\mathfrak{S}\omega_i\omega_j$ iff isj occurs on $\mathcal{B}$, $\mathfrak{A}v(a)\omega_i\omega_j$ iff iAaj occurs on $\mathcal{B}$. $v(a) = [a]$, and $\langle [a_1], \ldots, [a_n] \rangle \in v_{\omega_i}(P)$ iff $Pa_1 \ldots a_n, i$ is on $\mathcal{B}$, given that P is any n-place predicate other than identity. If $\neg Ra_m, i$ occurs on $\mathcal{B}$ and M is an n-place matrix with*

instantiations on the branch (where x_m is the first free variable in M and a_m is the constant in $M[a_1, \ldots, a_n/\vec{x}]$ that replaces x_m), then $\langle [a_1], \ldots, [a_n] \rangle \in v_{\omega_i}(M)$ iff $M[a_1, \ldots, a_n/\vec{x}], i$ occurs on $\mathcal{B}$. (Due to the identity rules this is well defined.) When we have $a = b, 0$, $b = c, 0$, etc. we choose one single object for all constants to denote.

Lemma 16 (*Completeness Lemma*): *Let $\mathcal{B}$ be an open branch in a complete tableau and let $\mathcal{M}$ be a (supplemented) model induced by $\mathcal{B}$. Then, for every formula A:*

(i)   *if A, i is on $\mathcal{B}$, then $v_{\omega_i}(A) = 1$, and*

(ii)   *if $\neg A, i$ is on $\mathcal{B}$, then $v_{\omega_i}(A) = 0$.*

    Proof. The proof is by induction on the complexity of A.

    Atomic formulas.

    $Pa_1 \ldots a_n, i$ is on $\mathcal{B} \Rightarrow \langle [a_1], \ldots, [a_n] \rangle \in v_{\omega_i}(P) \Rightarrow \langle v(a_1), \ldots, v(a_n) \rangle \in v_{\omega_i}(P) \Rightarrow v_{\omega_i}(Pa_1 \ldots a_n) = 1$.

    $\neg Pa_1 \ldots a_n, i$ is on $\mathcal{B} \Rightarrow Pa_1 \ldots a_n, i$ is not on $\mathcal{B}$ ($\mathcal{B}$ open) $\Rightarrow \langle [a_1], \ldots, [a_n] \rangle \notin v_{\omega_i}(P) \Rightarrow \langle v(a_1), \ldots, v(a_n) \rangle \notin v_{\omega_i}(P) \Rightarrow v_{\omega_i}(Pa_1 \ldots a_n) = 0$.

    $a = b, i$ is on $\mathcal{B} \Rightarrow a \sim b$ ($T-N=$) $\Rightarrow [a] = [b] \Rightarrow v(a) = v(b) \Rightarrow v_{\omega_i}(a = b) = 1$.

    $\neg a = b, i$ is on $\mathcal{B} \Rightarrow a = b, 0$ is not on $\mathcal{B}$ ($\mathcal{B}$ open) $\Rightarrow$ it is not the case that $a \sim b \Rightarrow [a] \neq [b] \Rightarrow v(a) \neq v(b) \Rightarrow v_{\omega_i}(a = b) = 0$.

    Suppose that M is a matrix where x_m is the first free variable and a_m is the constant in $M[a_1, \ldots, a_n/\vec{x}]$ that replaces x_m and that $v_{\omega_i}(Ra_m) = 0$. Then: $M[a_1, \ldots, a_n/\vec{x}], i$ occurs on $\mathcal{B} \Rightarrow \langle [a_1], \ldots, [a_n] \rangle \in v_{\omega_i}(M) \Rightarrow \langle v(a_1), \ldots, v(a_n) \rangle \in v_{\omega_i}(M) \Rightarrow v_{\omega_i}(M[a_1, \ldots, a_n/\vec{x}]) = 1$.

    $\neg M[a_1, \ldots, a_n/\vec{x}], i$ occurs on $\mathcal{B} \Rightarrow M[a_1, \ldots, a_n/\vec{x}], i$ is not on $\mathcal{B}$ ($\mathcal{B}$ open) $\Rightarrow \langle [a_1], \ldots, [a_n] \rangle \notin v_{\omega_i}(M) \Rightarrow \langle v(a_1), \ldots, v(a_n) \rangle \notin v_{\omega_i}(M) \Rightarrow v_{\omega_i}(M[a_1, \ldots, a_n/\vec{x}]) = 0$.

    Other truth-functional connectives and modal operators. Straightforward.

    Boulesic operators. ($\mathcal{A}$). Suppose $\mathcal{A}_c B, i$ is on $\mathcal{B}$. Furthermore, suppose that Rc, i is not on $\mathcal{B}$. Then $\neg Rc, i$ is on $\mathcal{B}$ [by *CUTR* (or *CUT*)]. Hence, $\mathcal{A}_c B$ is true in ω_i by definition and previous steps. Suppose Rc, i is on $\mathcal{B}$. Then the $\mathcal{A}$-rule has been applied to $\mathcal{A}_c B, i$, since the branch is complete. So, for some new j, $iAcj$ and B, j occur on $\mathcal{B}$. By the induction hypothesis,

$\mathfrak{A}v(c)\omega_i\omega_j$, and B is true in ω_j. Since Rc,i is on $\mathcal{B}$, $v(c)$ is perfectly rational in ω_i. Hence, $\mathcal{A}_cB$ is true in ω_i, as required. Suppose $\neg\mathcal{A}_cB,i$ is on $\mathcal{B}$. Furthermore, suppose that Rc,i is not on $\mathcal{B}$. Then $\neg Rc,i$ is on $\mathcal{B}$ [by $CUTR$ (or CUT)]. Consequently, $\mathcal{A}_cB$ is false in ω_i by definition and previous steps. Suppose Rc,i is on $\mathcal{B}$. Then the $\neg\mathcal{A}$-rule has been applied, and $\mathcal{W}_c\neg B$, i is on $\mathcal{B}$ since the branch is complete. Again, since Rc,i is on $\mathcal{B}$ and the branch is complete, the $\mathcal{W}$-rule has been applied and for every j such that $iAcj$ is on $\mathcal{B}$, $\neg B,j$ is on $\mathcal{B}$. By the induction hypothesis, B is false in every ω_j such that $\mathfrak{A}v(c)\omega_i\omega_j$. Since Rc,i is on $\mathcal{B}$, $v(c)$ is perfectly rational in ω_i. It follows that $\mathcal{A}_cB$ is false in ω_i, as required.

($\mathcal{W}$). Similar as for ($\mathcal{A}$).

Quantifiers. (Σ). Suppose that $\Sigma xA,i$ is on the branch. Since the tableau is complete (Σ) has been applied. Accordingly, for some c, $A[c/x],i$ is on the branch. Hence, $v_{\omega_i}(A[c/x]) = 1$, by (IH). For some $k_d \in \mathcal{L}(\mathcal{M})$, $v(c) = d$, and $v(k_d) = d$. Consequently, $v_{\omega_i}(A[k_d/x]) = 1$, by the Denotation Lemma. It follows that $v_{\omega_i}(\Sigma xA) = 1$. Suppose that $\neg\Sigma xA,i$ is on the branch. Since the tableau is complete ($\neg\Sigma$) has been applied. So, $\Pi x\neg A,i$ is on the branch. Again, since the tableau is complete (Π) has been applied. Thus, for all $c \in C$, $\neg A[c/x],i$ is on the branch. Consequently, $v_{\omega_i}(A[c/x]) = 0$ for all $c \in C$ [by the induction hypothesis]. If $k_d \in \mathcal{L}(\mathcal{M})$, then for some $c \in C$, $v(c) = v(k_d)$. By the Denotation Lemma, for all $k_d \in \mathcal{L}(\mathcal{M})$, $v_{\omega_i}(A[k_d/x]) = 0$. Consequently, $v_{\omega_i}(\Sigma xA) = 0$.

The case for Π is similar. ∎

Theorem 17 (*Completeness Theorem*): *Every system in this paper is complete with respect to its semantics.*

Proof. First we prove that the theorem holds for our weakest system $\mathcal{V}$. Then we extend the theorem to all extensions of this system. Let $\mathbf{M}$ be the class of models that corresponds to $\mathcal{V}$.

Suppose that B is not derivable from Γ in $\mathcal{V}$: then it is not the case that there is a closed $\mathcal{V}$-tableau whose initial list comprises $A,0$ for every A in Γ and $\neg B,0$. Let t be a complete $\mathcal{V}$-tableau whose initial list comprises $A,0$ for every A in Γ and $\neg B,0$. Then t is not closed—i.e. it is open. Since t is open, there is at least one open branch in t. Let $\mathcal{B}$ be an open branch in t. The

model induced by $\mathcal{B}$ makes all the premises in Γ true and B false in ω_0. Hence, it is not the case that B follows from Γ in $\mathbf{M}$. Consequently, if B follows from Γ in $\mathcal{M}$, then B is derivable from Γ in $\mathcal{V}$.

To prove that all extensions of $\mathcal{V}$ are complete with respect to their semantics, we have to check that the model induced by the open branch $\mathcal{B}$ is of the right kind. To do this we first check that this is true for every single semantic condition. Then we combine each of the individual arguments. I will go through some steps to illustrate the method.

$C-bD$. Suppose that ω_i is in W. Then i occurs on $\mathcal{B}$ [by the definition of an induced model]. Since $\mathcal{B}$ is complete $(T-bD)$ has been applied. Hence, for some j, $iAcj$ is on $\mathcal{B}$. Accordingly, for some ω_j, $\mathfrak{A}v(c)\omega_i\omega_j$, as required [by the definition of an induced model].

$C-b4$. Suppose that $\mathfrak{A}v(c)\omega_i\omega_j$ and $\mathfrak{A}v(c)\omega_j\omega_k$. Then $iAcj$ and $jAck$ occur on $\mathcal{B}$ [by the definition of an induced model]. Since $\mathcal{B}$ is complete, $(T-b4)$ has been applied and $iAck$ occurs on $\mathcal{B}$. It follows that $\mathfrak{A}v(c)\omega_i\omega_k$, as required [by the definition of an induced model].

$C-HW$. Suppose that $\mathfrak{A}v(c)\omega_i\omega_j$. Then $iAcj$ occurs on $\mathcal{B}$ [by the definition of an induced model]. Since $\mathcal{B}$ is complete, $(T-HW)$ has been applied and $iAdj$ occurs on $\mathcal{B}$. Consequently, $\mathfrak{A}v(d)\omega_i\omega_j$, as required [by the definition of an induced model].

$C-M\mathcal{W}$. Suppose that $\mathfrak{A}v(c)\omega_i\omega_j$. Then $iAcj$ occurs on $\mathcal{B}$ [by the definition of an induced model]. Since $\mathcal{B}$ is complete, $(T-M\mathcal{W})$ has been applied and irj occurs on $\mathcal{B}$. Consequently, $\mathfrak{R}\omega_i\omega_j$, as required [by the definition of an induced model].

$C-\mathcal{W}C$. Suppose that ω_i is in W. Then i occurs on $\mathcal{B}$ [by the definition of an induced model]. Since $\mathcal{B}$ is complete $(T-\mathcal{W}C)$ has been applied. Accordingly, for some j, $iAcj$ and irj are on $\mathcal{B}$. Thus, for some ω_j, $\mathfrak{A}v(c)\omega_i\omega_j$ and $\mathfrak{R}\omega_i\omega_j$, as required [by the definition of an induced model]. ∎

7. Conclusion

In this paper, I have developed a set of boulesic and boulesic-deontic tableau systems and I have investigated some possible connections between

boulesic logic and deontic logic. Boulesic logic is a new kind of logic that deals with 'boulesic' concepts and expressions, such as wanting and accepting, and 'boulesic' sentences, arguments and systems. I have shown how deontic logic, the logic of norms, might be grounded in boulesic logic. I have used a kind of possible world models to define the systems semantically and I have shown that all systems are sound and complete with respect to their semantics. Intuitively, we can think of our semantics as a description of the structure of a perfectly rational will. Finally, I have mentioned some interesting theorems that can be proved in our systems, including some versions of the so-called *hypothetical imperative.*

The deontic fragments of the systems in this paper are pretty standard monadic deontic systems. For a long time, systems of this kind have been criticised and various deontic 'paradoxes' have been introduced, for example, Ross's paradox, the paradox of derived obligations, the contrary-to-duty paradox, the good Samaritan paradox, the paradox of epistemic obligation and the free choice permission paradox.[37] Some think that these puzzles show that normal deontic logic is seriously defective. However, I am inclined to believe that most of the so-called 'deontic paradoxes' can be 'solved' and that they do not show that we have to abandon classical deontic logic. Of course, some of the puzzles are quite serious, for example, the contrary-to-duty paradox. It does not seem to be possible to solve this puzzle adequately in normal monadic deontic systems. This does not necessarily imply that we have to abandon classical deontic logic, but it indicates that the systems in this paper should be expanded or supplemented.[38]

I would now like to mention two ways in which the systems in this paper can be improved.

---

[37] For more on deontic paradoxes, see, for example, (Åqvist 1967; Castañeda 1981; Chisholm 1963; Hilpinen and McNamara 2013; Prior 1954, 1958; Ross 1941, 1944; and von Wright 1968).

[38] In (Rönnedal 2018), I discuss the contrary-to-duty paradox and suggest a solution. This solution, which is attractive in many respects, does not require that we abandon normal monadic deontic logic. The systems in the present paper are compatible with this solution. For more on the contrary-to-duty paradox and various possible solutions, see Rönnedal (forthcoming).

First, the systems in this paper can be combined with temporal logic. In a quantified temporal alethic boulesic deontic system, it is possible to investigate 'diachronistic' rationality and the relationships between temporal, alethic, boulesic and deontic concepts. I am currently trying to develop a set of quantified temporal alethic boulesic deontic systems.

Second, there appears to be a close connection between the logic of wishing/not accepting and the logic of good/bad. Good and bad are usually strongly connected in formal systems to the logic of preference (see, for example, Chisholm and Sosa 1966; Lenzen 1983; and Hansson 1990). In future work, I hope that I will be able to combine boulesic logic with the logic of preference and construct a set of boulesic-preference systems. Such systems might be used to overcome some of the shortcomings with the kind of monadic systems that I have investigated in this paper. Systems of this kind might, for example, perhaps be used to solve the contrary-to-duty paradox.

No doubt there are other possible extensions, but these examples seem to me to be among the most interesting ones. I hope to return to these topics in future work.

Acknowledgements

References

Åqvist, Lennart. 1967. "Good Samaritans, Contrary-to-duty Imperatives, and Epistemic Obligations." *Noûs* 1 (4): 361–79. https://doi.org/10.2307/2214624

Åqvist, Lennart. 1987. *Introduction to Deontic Logic and the Theory of Normative Systems*. Naples: Bibliopolis.

Åqvist, Lennart. 2002. "Deontic Logic." In *Handbook of Philosophical Logic*, 2nd Edition, vol. 8, edited by D. M. Gabbay and F. Guenthner, 147–264. Dordrecht/Boston/London: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0387-2_3

Åqvist, Lennart, and Hoepelman, Jaap. 1981. "Some Theorems about a 'Tree' System of Deontic Tense Logic." In *New Studies in Deontic Logic: Norms, Actions, and the Foundation of Ethics*, edited by R. Hilpinen, 187–221. Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-8484-4_9

Aristotle. 1992. *The Nicomachean Ethics*. Translated by Sir David Ross. Oxford and New York: Oxford University Press.

Barcan (Marcus), Ruth. C. 1946. "A Functional Calculus of First Order Based on Strict Implication." *Journal of Symbolic Logic* 11 (1): 1–16. https://doi.org/10.2307/2269159

Bedke, Matthew S. 2009. "The Iffiest Oughts: A Guise of Reasons Account of End-Given Conditionals." *Ethics* 119 (4): 672–98. https://doi.org/10.1086/600130

Björklund, Fredrik, Björnsson, Gunnar, Eriksson, John, Francén Olinder, Ragnar, and Strandberg, Caj. 2012. "Recent Work on Motivational Internalism." *Analysis* 72 (1): 124–37. https://doi.org/10.1093/analys/anr118

Björnsson, Gunnar, Strandberg, Caj, Francén Olinder, Ragnar, Eriksson, John, and Björklund, Fredrik. Eds. 2015. *Motivational Internalism*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199367955.001.0001

Blackburn, Patrick, de Rijke, Maarten, and Venema, Yde. 2001. *Modal Logic*. Cambridge University Press. https://doi.org/10.1017/CBO9781107050884

Blackburn, Patrick, van Benthem, Johan, and Wolter, Frank. Eds. 2007. *Handbook of Modal Logic*. Elsevier.

Bratman, Michael E. 1999. *Intention, Plans, and Practical Reason*. CSLI Publications.

Broersen, Jan M. 2011. "Making a Start with the *stit* Logic Analysis of Intentional Action." *Journal of Philosophical Logic* 40 (4): 499–530. https://doi.org/10.1007/s10992-011-9190-6

Broersen, Jan M., Dastani, Mehdi, and van der Torre, Leendert. 2001. "Resolving Conflicts between Beliefs, Obligations, Intentions, and Desires." In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, edited by Salem Benferhat, and Philippe Besnard, 568–79. Springer. https://doi.org/10.1007/3-540-44652-4_50

Broome, John. 1999. "Normative Requirements." *Ratio* (new series) 12 (4): 398–419. https://doi.org/10.1111/1467-9329.00101

Broome, John. 2013. *Rationality through Reasoning*. Wiley-Blackwell. https://doi.org/10.1002/9781118609088

Brunero, John. 2010. "Self-Governance, Means-Ends Coherence, and Unalterable Ends." *Ethics* 120 (3): 579–91. https://doi.org/10.1086/652448

Carnap, Rudolf. 1946. "Modalities and Quantification." *Journal of Symbolic Logic* 11 (2): 33–64. https://doi.org/10.2307/2268610

Castañeda, Héctor-Neri. 1981. "The Paradoxes of Deontic Logic: the Simplest Solution to all of them in one Fell Swoop." In *New Studies in Deontic Logic: Norms, Actions, and the Foundation of Ethics*, edited by R. Hilpinen, 37–85. Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-8484-4_2

Chellas, Brian. F. 1969. *The Logical Form of Imperatives*. Stanford: Perry Lane Press.

Chellas, Brian. F. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621192

Chisholm, Roderick M. 1963. "Contrary-to-duty Imperatives and Deontic Logic." *Analysis* 24 (2): 33–36. https://doi.org/10.1093/analys/24.2.33

Chisholm, Roderick M. and Sosa, Ernest. 1966. "On the Logic of 'Intrinsically Better.'" *American Philosophical Quarterly* 3 (3): 244–49.

Cohen, Philip R. and Levesque, Hector J. 1990. "Intention is Choice with Commitment." *Artificial Intelligence* 42 (2–3): 213–61. https://doi.org/10.1016/0004-3702(90)90055-5

D'Agostino, Marcello, Gabbay, Dov M., Hähnle, Reiner, and Posegga, Joachim. Eds. 1999. *Handbook of Tableau Methods*. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-017-1754-0

Downie, Robin S. 1984. "The Hypothetical Imperative." *Mind* (New Series) 93 (372): 481–90. https://doi.org/10.1093/mind/XCIII.372.481

Feldman, Fred. 1986. *Doing the Best We Can: An Essay in Informal Deontic Logic*. Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-4570-8

Feldman, Fred. 2004. *Pleasure and the Good Life*. Oxford, New York: Oxford University Press. https://doi.org/10.1093/019926516X.001.0001

Firth, Roderick. 1952. "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12 (3): 317–45. https://doi.org/10.2307/2103988

Fitting, Melvin, and Mendelsohn, Richard L. 1998. *First-Order Modal Logic*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-5292-1

Foot, Philippa. 1972. "Morality as a System of Hypothetical Imperatives." *The Philosophical Review* 81 (3): 305–16. https://doi.org/10.2307/2184328

Gabbay, Dov M. 1976. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Dordrecht: Reidel. https://doi.org/10.1007/978-94-010-1453-3

Gabbay, Dov M., Horty, John, Parent, Xavier, van der Meyden, Ron, and van der Torre, Leendert. Eds. 2013. *Handbook of Deontic Logic and Normative Systems*. London: College Publications.

Garson, James W. 1984. "Quantification in Modal Logic." In *Handbook of Philosophical Logic* 2 (2nd edition 3, 2001), edited by D. M. Gabbay and F. Guenthner, 249–307. Dordrecht: Springer. https://doi.org/10.1007/978-94-009-6259-0_5

Garson, James W. 2006. *Modal Logic for Philosophers*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511617737

Gensler, Harry J. 1985. "Ethical Consistency Principles." *The Philosophical Quarterly* 35 (139): 156–70. https://doi.org/10.2307/2219341

Gensler, Harry J. 2002. *Introduction to Logic.* London and New York: Routledge.

Greenspan, Patricia S. 1975. "Conditional Oughts and Hypothetical Imperatives." *The Journal of Philosophy* 72 (10): 259–76. https://doi.org/10.2307/2024734

Hansson, Sven Ove. 1990. "Defining 'Good' and 'Bad' in Terms of 'Better.'" *Notre Dame Journal of Formal Logic* 31 (1): 136–49. https://doi.org/10.1305/ndjfl/1093635338

Harsanyi, John C. 1958. "Ethics in Terms of Hypothetical Imperatives." *Mind* 67 (267): 305–16. https://doi.org/10.1093/mind/LXVII.267.305

Hill, Jr. Thomas E. 1973. "The Hypothetical Imperative." *The Philosophical Review* 82 (4): 429–50. https://doi.org/10.2307/2183709

Hill, Jr. Thomas E. 1989. "Kant's Theory of Practical Reason." *The Monist* 72 (3): 363–83. https://doi.org/10.5840/monist198972320

Hilpinen, Risto. Ed. 1971. *Deontic Logic: Introductory and Systematic Readings.* Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-010-3146-2

Hilpinen, Risto. Ed. 1981. *New Studies in Deontic Logic: Norms, Actions, and the Foundation of Ethics.* Dordrecht: D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-8484-4

Hilpinen, Risto, and McNamara, Paul. 2013. "Deontic Logic: A Historical Survey and Introduction." In *Handbook of Deontic Logic and Normative Systems*, edited by D. M. Gabbay, J. Horty, X. Parent, R. van der Meyden and L. van der Torre, 3–136. London: College Publications.

Hintikka, Jaakko. 1961. "Modality and Quantification." *Theoria* 27 (3): 117–28. https://doi.org/10.1111/j.1755-2567.1961.tb00020.x

Horty, John F. 2015. "Requirements, Oughts, Intentions." *Philosophy and Phenomenological Research* 91 (1): 220–29. https://doi.org/10.1111/phpr.12204

Hughes, George E. and Cresswell, Max J. 1968. *An Introduction to Modal Logic.* London: Routledge. https://doi.org/10.4324/9780203028100

Jeffrey, Richard C. 1967. *Formal Logic: Its Scope and Limits.* New York: McGraw-Hill. https://doi.org/10.2307/2271990

Kant, Immanuel. 1991. *Grundlegung zur Metaphysik der Sitten.* In *The Moral Law: Kant's Groundwork of the Metaphysics of Morals.* Translated and analysed by

H. J. Paton. London and New York: Routledge (Reprinted 1991; originally published 1785).

Kawall, Jason. 2013. "Ideal Observer Theories." In *The International Encyclopedia of Ethics*, edited by H. LaFollette, 2523–30. Blackwell Publishing. https://doi.org/10.1002/9781444367072.wbiee548

Knuuttila, Simo. 2004. *Emotions in Ancient and Medieval Philosophy*. Oxford: Oxford University Press. https://doi.org/10.1093/0199266387.001.0001

Korsgaard, Christine M. 2008. "The Normativity of Instrumental Reason." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford/New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552733.003.0002

Kracht, Marcus. 1999. *Tools and Techniques in Modal Logic*. Amsterdam: Elsevier.

Lenzen, Wolfgang. 1983. "On the Representation of Classificatory value Structures." *Theory and Decision* 15 (4): 349–69. https://doi.org/10.1007/BF00162113

Lewis, Clarence I., and Langford, Cooper H. 1932. *Symbolic Logic*. New York: The Century Company.

Lorini, Emiliano, and Herzig, Andreas. 2008. "A Logic of Intention and Attempt." *Synthese* 163 (1): 45–77. https://doi.org/10.1007/s11229-008-9309-7

Mally, Ernst. 1926. *Grundgesetze des Sollens: Elemente der Logik des Willens*. Leuschner and Lubensky.

Marra, Alessandra, and Klein, Dominik. 2015. "Logic and Ethics: An Integrated Model for Norms, Intentions and Actions." In *International Workshop on Logic, Rationality and Interaction*, edited by Wiebe van der Hoek, Wesley H. Holliday, and Wen-fang Wang, 268–81. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-48561-3_22

Marshall, John. 1982. "Hypothetical Imperatives." *American Philosophical Quarterly* 19 (1): 105–14.

Mele, Alfred R. Ed. 2004. *The Oxford Handbook of Rationality*. Oxford: Oxford University Press. https://doi.org/10.1093/0195145399.001.0001

Parks, Zane. 1976. "Investigations into Quantified Modal Logic I." *Studia Logica* 35 (2): 109–25.

Paton, Herbert J. 1948. *The Moral Law: Kant's Groundwork of the Metaphysics of Morals*. London and New York: Routledge (Reprinted 1991).

Priest, Graham. 2005. *Towards Non-Being*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198783596.001.0001

Priest, Graham. 2008. *An Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511801174

Prior, Arthur N. 1954. "The Paradoxes of Derived Obligation." *Mind* 63: 64–65.

Prior, Arthur N. 1958. "Escapism: The Logical Basis of Ethics." In *Essays in Moral Philosophy*, edited by Abraham I. Melden, 135–46. Seattle: University of Washington Press.

Rönnedal, Daniel. 2009. "Counterfactuals and Semantic Tableaux." *Logic and Logical Philosophy* 18 (1): 71–91. https://doi.org/10.12775/LLP.2009.006

Rönnedal, Daniel. 2012. "Bimodal Logic." *Polish Journal of Philosophy.* VI (2): 71–93. https://doi.org/10.5840/pjphil20126214

Rönnedal, Daniel. 2018. "Temporal Alethic Dyadic Deontic Logic and the Contrary-to-Duty Obligation Paradox." *Logic and Logical Philosophy* 27 (1): 3–52. https://doi.org/10.12775/LLP.2017.012

Rönnedal, Daniel. Forthcoming. "Contrary-to-Duty Obligations and the Contrary-to-Duty (Obligation) Paradox." *The Internet Encyclopedia of Philosophy.*

Ross, Alf. 1941. "Imperatives and Logic." *Theoria* 7: 53–71.

Ross, Alf. 1944. "Imperatives and Logic." *Philosophy of Science* 11 (1): 30–46. https://doi.org/10.2307/2268025

Schroeder, Mark. 2004. "The Scope of Instrumental Reason." *Philosophical Perspectives* 18 (1): 337–64. https://doi.org/10.1111/j.1520-8583.2004.00032.x

Schroeder, Mark. 2005. "The Hypothetical Imperative?" *Australasian Journal of Philosophy* 83 (3): 357–72. https://doi.org/10.1080/00048400500191958

Schroeder, Mark. 2009. "Means-End Coherence, Stringency, and Subjective Reasons." *Philosophical Studies* 143 (2): 223–48. https://doi.org/10.1007/s11098-008-9200-x

Schroeder, Mark. 2015. "Hypothetical Imperatives." In *Reason, Value, and Respect: Kantian Themes from the Philosophy of Thomas E. Hill, Jr.*, edited by Mark Timmons and Robert N. Johnson, Chapter 4. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199699575.003.0005

Semmling, Caroline, and Wansing, Heinrich. 2008. "From BDI and stit to bdi-stit Logic." *Logic and Logical Philosophy* 17 (1–2): 185–207. https://doi.org/10.12775/LLP.2008.011

Shaver, Robert. 2006. "Korsgaard on Hypothetical Imperatives." *Philosophical Studies* 129 (2): 335–47. https://doi.org/10.1007/s11098-004-1646-x

Smullyan, Raymond M. 1968. *First-Order Logic.* Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-86718-7

Sumner, Leonard W. 1996. *Welfare, Happiness, and Ethics.* Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198238782.001.0001

Sumner, Leonard W. 2000. "Something in Between." In *Well-Being and Morality: Essays in Honour of James Griffin*, edited by Roger Crisp and Brad Hooker, 1–19. Oxford: Clarendon Press.

van Roojen, Mark. 2013. "Internalism, Motivational." In *International Encyclope-dia of Ethics*, edited by H. LaFollette, 2693–706. Malden, MA: Wiley Black-well.

von Wright, Georg H. 1968. *An Essay in Deontic Logic and the General Theory of Action.* Amsterdam: North-Holland.

Wallace, R. Jay. 2001. "Normativity, Commitment, and Instrumental Reason." *Philosophers' Imprint* 1 (3): 1–26.

Way, Jonathan. 2010. "Defending the Wide-Scope Approach to Instrumental Rea-son." *Philosophical Studies* 147 (2): 213–33. https://doi.org/10.1007/s11098-008-9277-2