Deontic Modality in Rationality and Reasoning

Marra, Alessandra

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Deontic Modality in Rationality and Reasoning

Cover Image: *Hommage to the square, new gate* by Josef Albers

Printed by: Print Service Ede — The Netherlands

Deontic Modality in Rationality and Reasoning

Proefschrift
ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op maandag 20 mei 2019 om 13.30 uur
door

Alessandra Marra
geboren te Padova, Italië

Promotor:	Prof. dr. A.P. Thomas
Copromotor:	Dr. R.A. Muskens
Promotiecommissie:	Prof. dr. J.F. Horty
	Dr. A. Silk
	Dr. A.M. Tamminga
	Prof. dr. F.J.M.M. Veltman
	Dr. N.W. Wildman

Acknowledgments

I am indebted to many for discussions on the ideas contained in this thesis. Thank you to my supervisors, Reinhard Muskens and Alan Thomas, who gave me the opportunity to pursue such a project and guided me through it. Many thanks for sharing with me their enthusiasm and expectations on what formal philosophy should be. They have been a source of encouragement and inspiration.

I am grateful to the members of my PhD committee, Jeff Horty, Alex Silk, Allard Tamminga, Frank Veltman, and Nathan Wildman. Thank you for their detailed comments, insightful questions, sage (and truly supererogatory) advice on this project.

I am fortunate to have had the opportunity to write this dissertation surrounded by great colleagues. Thank you to those with whom I crossed paths at TiLPS: they made it a place of learning and sharing, where I always felt heard. Special thanks to Amanda Cawston, who kindly accepted the role of junior research supervisor in the final year of my doctorate studies. Collaborations have been a source of growth and great fun: many thanks to Dominik Klein, Olivier Roy and Naftali Weinberger. I have learned a lot. I should thank Frank Veltman a second time: he endured many hours of conversation with me in the last years, since my time as a master student in Amsterdam. Thank you for such a generosity.

I owe thanks to the Philosophy Department in Tilburg for graciously supporting my participation in conferences and other events abroad, including my visit to UC Berkeley. And thank you to the Philosophy Department at UC Berkeley, and especially the graduate students, for making my stay such a wonderful experience. Finally, many thanks to the Philosophy Department in Bayreuth for welcoming me over the last two years.

I owe much to Chiara, Beira, Matteo, Silvia, Colin, Lasha, Jun and Machteld.

Thank you for making Tilburg home. And for showing me the beauty in the experiments of living.

Thank you to Anneliese and Hans, and to Margherita for traveling all the way to Tilburg to share this moment with me.

Thank you to my mother Emanuela, my father Roberto, and my brother Stefano. Thank you for their trust and loving support. Grazie per avermi sempre lasciato libera, accompagnandomi per mano.

My heartfelt thank you goes to Dominik. Thank you for all the care, encouragement, and great laughs.

Contents

Chapter 1

Introduction

This thesis seeks to investigate certain facets of the logical structure of *oughts* — where "ought" is used as a noun, roughly meaning obligation. We do so by following two lines of inquiry. The first part of the thesis places oughts in the context of *practical rationality*. The second part of the thesis concerns the rules of inference governing deontic arguments, and specifically the rule of *Reasoning by Cases*. These two lines of inquiry, together, aim to expound upon oughts in rationality and reasoning — thereby the title.

1.1 Oughts, Goals, and Enkrasia

In the first part of the thesis, we investigate oughts in the context of practical rationality. Specifically, we address the rational balance between *oughts*, on the one hand, and practical notions such as *plans*, on the other. The relation between those concepts can be construed in different ways. Suppose an agent believes sincerely and with conviction that she ought to X. For Gibbard, this amounts, fundamentally, to planning to X. "As a first approximation [...]" — he writes — "ought thoughts are like plans. Thinking what I ought to do amounts to thinking what to do." (Gibbard, 2008, p.19). There are further complexities in Gibbard's view. However, for our purposes, what is relevant is that it constitutes one account of how oughts that are believed by an agent relate to her plans. For Gibbard, believing one ought to X identifies with (certain aspects of) planning to X.

As an implication, Gibbard's view rules out the possibility of believing one ought to X while planning to do something else. This point has already been brought to attention by Bratman and Broome in their comments to Gibbard

(2008) (included in Gibbard, 2008, pp.95,106). Suppose I believe sincerely and with conviction that today I ought to repay my friend Ann the 10 euro she lent me. Yet, thinking about what to do today, I decide to spend all my money on going to the movies. That appears defective but still possible — contrary to Gibbard's view. This observation suggests an alternative way in which the relation between believed oughts and plans can be understood. Following Broome, we can say that believing that one ought to X without planning to X is not impossible, but rather irrational (Broome, 2013, p.175). It is *rationality* that requires certain relations between believed oughts and plans. The principle of rationality governing such relation is called *Enkrasia*, and is the focus of the first part of this thesis.

Let us say that X-ing is a *goal* in an agent's plan if X is something the agent plans for in itself. Drawing on Bratman (1987), we understand goals in plans as indicating what the agent is committed to achieve. Such a commitment shapes the agent's deliberation already before the agent starts to act, and is independent to the question of whether the agent will in fact succeed in achieving her goals.[1]

The starting point of our inquiry is Enkrasia in the following interpretation, first suggested by Horty (2015): rationality requires that *if* an agent sincerely and with conviction believes she ought to X, *then* X-ing is a goal in the agent's plan. Suppose I believe sincerely and with conviction that today I ought to repay my friend Ann 10 euro. Rationality demands that if this is the case, then repaying my friend is a goal in my plan, something I am committed to. The first part of the present thesis is devoted to the analysis of the structure of Enkrasia from a logical point of view. This is, to the best of our knowledge, a largely novel project. We show that it can provide a conceptual and formal contribution to the understanding of the logic of oughts in the context of practical rationality.

Specifically, we address the following questions:

- What is the logical relation between believed oughts and goals?

[1] Specifically, Bratman (1987) identifies two relevant dimensions of commitment: (i) the so-called *volitional dimension*, which is related to affecting the agent's conduct and motivating her to take the first steps to fulfill her goal; and (ii) the so called *reasoning-centered dimension* that concerns the role that goals in plans, thanks to their characteristic stability, play in the agent's deliberation in the period between their formation and their eventual execution. Here we are mainly concerned with the reasoning-centered dimension of commitment (cf. Bratman, 1987, pp. 15-18, 107-110). Moreover, a logical analysis of the (relative) stability of goals in plans will be the focus of Chapter 3 of this thesis.

- Do all believed oughts *potentially* correspond to goals?
- What is the dynamic relation between believed oughts and goals?

Our answers are summarized below.

What is the logical relation between believed oughts and goals? *Chapter 2* provides an answer to this question. To a first approximation, it is shown that the relation between the oughts believed by the agent and her goals is logically non-trivial. Our investigation begins by considering two further requirements of rationality governing goals in plans: the principles of Internal and Strong Consistency. These principles demand that goals are both logically consistent, and consistent with what the agent believes to be possible. These are standardly endorsed principles, ultimately grounded on the dimension of commitment that goals carry. Oughts, on the other hand, are not necessarily constrained by similar consistency demands. If we admit the possibility that believed oughts can conflict, and thus do not obey internal or strong consistency requirements, it follows that not all the oughts believed by the rational agent correspond to her goals. Thus, there is a tension between Enkrasia, on the one hand, and Internal and Strong Consistency for goals on the other. In Chapter 2, we solve this tension by proposing a logic for oughts and goals in which the principles of Internal and Strong Consistency are valid, while Enkrasia is generally not. Importantly, it is formally shown that Enkrasia is a principle of *bounded logical validity.*

Do all believed oughts *potentially* correspond to goals? Our second question is also addressed in *Chapter 2*. In brief, we argue for a reply in the negative. The key to such an reply lies in the conceptual distinction between *basic* and *derived* oughts. Let us suppose that an episode of the agent's deliberation begins with a given set of oughts the agent believes in. We think of the oughts in such a set as basic oughts. Derived oughts are those implied — in a sense to be made more precise — from basic oughts. While such distinction between basic and derived oughts is not immediately related to Enkrasia, we argue that it is (at least) with respect to Enkrasia that such a distinction becomes non-trivial. In fact, Enkrasia applies to one but not the other. Derived oughts, even if believed by the agent, cannot correspond to goals the agent plans for in themselves. The logic developed in Chapter 2 captures such conceptual distinction both at the syntax (by having two different operators for oughts) and at the semantic level (via neighborhood semantics).

The remainder of Chapter 2 is devoted to the illustration of one further case in which the distinction between basic and derived oughts can be put to work. In particular, we show that such a distinction can be fruitfully applied to the debate surrounding the validity of *deontic closure*. Deontic closure is a family of logical principles indicating that oughts are closed under implication: from an ought to X, and from X implies Y, it derives an ought to Y. Some instances of deontic closure appear to capture crucial features of deontic reasoning: this is the case for the so-called Practical Inference, which allows one to reason from one ought to an ought instrumental to it (Von Wright, 1963). Other instances of deontic closure are more problematic and appear to lead to unacceptable results: the so-called Ross' Paradox is an emblematic example (Ross, 1941; Hilpinen and McNamara, 2013). Until now, the trade-off appeared to be between an outright rejection of deontic closure (and thus a "too thin" logic) and the unrestricted validity of deontic closure (and thus a "too thick" logic). We argue that there is an intermediate way. The distinction between basic and derived oughts indeed helps us discriminate between valid and invalid instances of deontic closure. Specifically, we argue that deontic closure is valid whenever the ought inferred is a derived ought, but not if it is a basic ought.

What is the dynamic relation between believed oughts and goals? *Chapter 3* explores one specific dynamic relation between believed oughts and goals. There exist several relevant kinds of dynamics: for instance, oughts and goals may change with the passing of time, or the agent might come to believe different basic oughts, or the agent learns new facts of the world. We restrict our attention to the latter kind of dynamics. In particular, we focus on the changes that concern only the courses of events the agent believes are open to her. We call this *practical dynamics*. Suppose the agent has a certain plan for the day. To paraphrase Bratman (1987), an unanticipated obstacle may restrict the agent's options of action, while unexpected opportunities may open new and better venues. Either case may trigger significant replanning. Practical dynamics bears significant conceptual connections to the debate on the *stability* of goals. Standardly, goals in plans are said to be relatively stable: although not irrevocable, they tend to resist reconsideration (Bratman, 1987, pp. 16,67). This implies that a goal will, in general, not be discarded with every piece of incoming information. We formally investigate the conditions of stability of goals under practical dynamics. It is shown that goals are in fact more stable than derived oughts. However, goals may vary, and they do so in a non-monotonic way. This closes the first part of this thesis.

1.2 Oughts, Reasoning by Cases, and the Miners' Puzzle

Which inference rules should guide deontic arguments? This question can be understood in various ways. For instance, we can ask which inference rules best guide our deontic reasoning, given that we have limited cognitive powers and that we are prone to reasoning errors. Or, starting from a certain ethical conception of oughts, we can ask about the logical structure of appropriate inferences involving those oughts. Alternatively, we can ask which rules govern acceptable arguments involving deontic modals in natural language. These are, in principle, different perspectives; here, we focus on the last two. In fact, recent literature has shown that certain classical inference rules become philosophically and linguistically problematic whenever applied to the deontic domain — even supposing that those who draw such inferences are cognitively ideal and fully rational agents. In the second part of the thesis, we investigate one of those problematic classical rules: the rule of *Reasoning by Cases.* Our investigation is carried out under certain idealized assumptions. Specifically, we proceed by abstracting away from issues related to agents' bounded cognitive and rational capacities. Furthermore, we focus on a narrow conception of reasoning. We use the term to indicate a certain deductive connection between premises and conclusion, rather than a non-monotonic process from the former to the latter (cf. Harman, 1986). "Reasoning", "argument" and "inference" are here used interchangeably.

Schematically, the inference rule of Reasoning by Cases moves from the premises $\ulcorner\varphi_1$ or $\varphi_2\urcorner$, $\ulcorner$if φ_1 then $\psi_1\urcorner$ and $\ulcorner$if φ_2 then $\psi_2\urcorner$ to the conclusion $\ulcorner\psi_1$ or $\psi_2\urcorner$. For the sake of illustration, let us consider the following scenario — originally credited to Levesque, and appearing in (Brachman et al., 1992, pp.25-26) and (Stanovich, 2011, p.106). We describe the scenario as reported in *The Guardian* newspaper (Bellos, 2016): *Jack is looking at Anne, but Anne is looking at George. Jack is married, but George is not. Is a married person looking at an unmarried person?*

Anne's marital status in unknown. Yet, the correct answer to the question is *yes.* The following is an instance of Reasoning by Cases:

1 | Anne is married or Anne is unmarried
2 | If Anne is married, then a married person looks at George
3 | If Anne is unmarried, then Jack looks at an unmarried person
4 | A married person looks at George or Jack looks at an unmarried person

No matter whether Anne is married or not, in either case Reasoning by Cases leads to the (correct) conclusion that a married person is looking at an unmarried person. In fact, Reasoning by Cases appears to be a crucial inference rule especially in contexts of *partial information*, that is, in contexts in which there is not enough information to determine which one of the disjuncts in the first premise is true (and so to rule out the other disjunct, whenever exclusive).

Reasoning by Cases, however, has received compelling criticism. Kolodny and MacFarlane (2010) discuss the following example, known as the *Miners' Puzzle*. They describe the scenario as follows:[2]

> Ten miners are trapped either in shaft A or in shaft B, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If you block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.
>
> (Kolodny and MacFarlane, 2010, p.115)

You do not know in which shaft the miners are in. On the one hand, blocking the shaft the miners are in results in saving all ten miners. On the other, blocking the shaft the miners are not in results in killing all the miners. Blocking neither shaft, finally, guarantees that nine miners are saved. What to do? You can reason from the following premises:

P1: The miners are in A or they are in B

[2] Kolodny and MacFarlane (2010) report to have taken the example from Parfit (1988) who, in turn, refers back to Donald Regan. A structurally identical scenario is also known in the literature as the *Jackson's Example*. See (Mason, 2013, p.2, ft.4) for further references. A similar scenario, known as the *Gambling Problem*, can be found in (Horty, 2001, p.55). We will come back to (a variation of) Horty's scenario in Chapter 4 of the thesis.

P2: If they are in A, I ought to block A

P3: If they are in B, I ought to block B

The premises of the above argument appear to all be acceptable. Let us consider them in turn. Premise P1 simply expresses a feature of the scenario: you assume the miners are in one of the two shafts A or B. Premises P2 and P3 register that, under the supposition that the miners are in a specific shaft (in turn, shaft A or shaft B), the right thing to do is indeed blocking that shaft — as this allows to save all ten miners.

But here is the puzzle. From the above premises, it follows via Reasoning by Cases:

C: I ought to block A or I ought to block B

But this conclusion strikes us as unacceptable. In fact, given your *uncertainty* about the position of the miners, neither of the disjuncts in the conclusion hold. It is not the case that, unconditionally, you ought to block shaft A — as you cannot rule out that the miners are actually in shaft B. Nor is it the case that, unconditionally, you ought to block shaft B — again, you cannot rule out that the miners are actually in shaft A. Following Kolodny and MacFarlane (2010); Willer (2012); Carr (2015); Cariani et al. (2013); Bledin (2015), you might even say that:

P4: I ought to block neither of the shafts

As this guarantees that nine miners are saved. P4 contradicts C, thus the unacceptability of C remains.

Let us summarize the main aspects of the Miners' Puzzle. It is a characteristic feature of the scenario that you have only partial information about where the miners actually are. The position of the miners is already *settled*, but you do not know where. Hence, we can think of the above argument as an example of reasoning under *uncertainty*. Importantly, reasoning by case distinction might go wrong under uncertainty. The above argument is an example in which the application of the inference rule of Reasoning by Cases leads from acceptable premises to an unacceptable conclusion.

How could the Miners' Puzzle be solved? Some possible reactions consist in (i) accepting the conclusion, (ii) rejecting the acceptability of one (or more) of the premises, or (iii) taking issue with the inference rule of Reasoning by

Cases.[3] Let us illustrate them.

One possible reaction consists in taking issue with the apparent unacceptability of the conclusion C. In fact, there is an ethical conception of oughts for which C turns out to be acceptable. These are the so-called objective oughts (see Gibbard, 2005). Crucially, objective oughts are not tied to the information an agent possesses. Thus, in determining what you objectively ought to do, your uncertainty about the miners' position plays no role. Imagine, for the sake of illustration, that the miners are actually in shaft A. In light of this fact, what you objectively ought to do is block shaft A. Hence, under some plausible assumptions on the meaning of disjuction, it indeed follows that you (objectively) ought to block A or you (objectively) ought to block B. Hence the puzzle is (dis)solved: under the objectivist reading of oughts, C is acceptable (as are P1, P2, and P3). Dowell (2012) develops a Kratzerian semantics for oughts in natural language that accounts for such an objectivist conception of oughts. Other ways to incorporate the objectivist reading in a semantics for oughts in natural language are discussed by Silk (2014a). We will come back to the scope of the objectivist solution to the Miners' Puzzle in Chapter 4.

A second possible way out of the Miners' Puzzle consists in challenging the acceptability of the argument's premises, and in particular of premises P2 and P3. von Fintel (2012), for instance, denies that P2 and P3 are acceptable, and explains away their apparent acceptability by means of an enthymematic interpretation. According to such interpretation, if a sentence like "If the miners are in X, then I ought to block X" seems acceptable in the Miners' scenario, it is because it is elliptical for "If the miners are in X *and I know it*, then I ought to block X". Therefore, P2 and P3 are not literally acceptable. Under the elliptical (and acceptable) reading, on the other hand, the form of premises P1, P2 and P3 would not license the application of the inference rule of Reasoning by Cases. Criticisms to this line of response to the Miners' Puzzle can be found in Carr (2015) . Furthermore, the treatment of oughts as informational modals (defended by Kolodny and MacFarlane (2010); Bledin (2015); Carr (2015), and adopted in Chapter 5 of this thesis) can predict the truth of sentences like "If the miners are in X and I know it, then I ought to

[3]This list is not complete. For instance, it could be argued that the *logical form* of the premises P1, P2 and P3 is not such to warrant the application of the inference rule of Reasoning by Cases. Critical responses to such solution can be found in Kolodny and MacFarlane (2010); Charlow (2013) and, more generally, in Silk (2014b). We will come back to this in Chapter 5.

block X" in the Miners' scenario. It does so, however, via semantic mechanisms and without postulating any elliptical, non-literal reading of P2 and P3.

Finally, the third possible response to the Miners' Puzzle takes issue with the inference rule of Reasoning by Cases via which the conclusion C is derived from premises P1, P2 and P3. Kolodny and MacFarlane (2010); Willer (2012); Carr (2015); Cariani et al. (2013); Bledin (2015), among others, take the Miners' Puzzle as a genuine counterexample to Reasoning by Cases, and use it to draw conclusions about the logical validity of such an inference rule. The Miners' Puzzle, it is argued, shows that Reasoning by Cases is an invalid inference rule in the deontic domain: for oughts in natural language as well as in ethics.[4]It is this debate the present thesis seeks to contribute to.

In particular, we set out to investigate the following questions:

- Is the invalidity of Reasoning by Cases in the deontic domain limited to situations of reasoning under uncertainty?
- When is Reasoning by Cases valid?

Below we outline our answers.

Is the invalidity of Reasoning by Cases in the deontic domain limited to situations of reasoning under uncertainty? An answer to this question is provided in *Chapter 4*. The answer is, briefly put, a negative one. More specifically, we show that the invalidity of Reasoning by Cases is *not* ultimately rooted in the epistemic nature of the Miners' scenario. There exists indeed an interpretation of *objective* oughts for which a Miners-like puzzle emerges. Where the original Miners' Puzzle is an example of reasoning under uncertainty, the new puzzle is an example of reasoning under *indeterminacy*. The relevant indeterminacy here is indeterminacy of the future, and the objective oughts for which the new puzzle emerges are future-dependent objective oughts. The rest of the chapter is devoted to investigate the semantics of those special objective oughts.

When is Reasoning by Cases valid? *Chapter 5* answers this second ques-

[4]For what concerns the invalidity of Reasoning by Cases for deontic modals in natural language, see for instance Cariani et al. (2013). For what concerns oughts in ethics, Carr (2015) focuses on an interpretation of oughts as subjective, and argues that Reasoning by Cases is invalid for such ethical interpretation. Finally, Kolodny and MacFarlane (2010); Willer (2012); Bledin (2015) appear to draw conclusions about the invalidity of Reasoning by Cases in natural language as well as in ethics. To our understanding, their frameworks are intended to capture a very general structure of oughts, which is common between natural language and ethics.

tion. In fact, counterexamples to Reasoning by Cases go beyond the deontic domain. Bledin (2014) proposes a counterexample to Reasoning by Cases involving certain epistemic modals, while Carr (2015) discusses a counterexample involving probabilistic modals. Chapter 5 brings all those counterexample together. Again, the contribution made here is to be situated *within* the literature that takes those counterexamples as genuine counterexamples to Reasoning by Cases. The chapter pursues two tasks. The first task is diagnostic. Although the counterexamples proposed in the literature thus far all involve modals (be they deontic, epistemic or probabilistic), we argue that the failure of Reasoning by Cases is *not* strictly speaking due to the presence of modals. The problem has a more general nature. To show this, we defend a novel counterexample involving only indicative conditionals. Afterwards, we turn to a more positive proposal. The second task of the chapter is indeed an investigation of the boundaries within which Reasoning by Cases *is* valid. We propose a sufficient criterion: Reasoning by Cases is valid if it involves sentences that are — in a sense to be specified — stable with respect to information loss. Our proposal predicts the validity of Reasoning by Cases in a wide array of contexts. This, if correct, is (largely) positive news for classical reasoning. This concludes Chapter 5 and, with it, the present thesis.

1.3 Sources of the Chapters

This thesis builds on earlier work, some co-authored, that has been published or is undergoing submission. The large majority of Chapter 2 and Chapter 3 is currently in the process of production in *Studia Logica* as Klein and Marra (2019), and is reproduced here with Springer's permission — which is gratefully acknowledged. Chapter 2 extends the first part of the paper Klein and Marra (2019), and relates to some ideas already appeared in Marra and Klein (2015). Chapter 3 extends the second part of the paper Klein and Marra (2019). In Klein and Marra (2019), authors are listed in alphabetic order, and contributed to the output equally. The standard disclaimer applies: all remaining errors are mine. Finally, Chapter 4 extends Marra (2016), while Chapter 5 is based on Marra (2018).

This thesis is divided into two main conceptual parts. Chapter 2 and Chapter 3, which together constitute the first part of the thesis, are the continuation of each other. In particular, Chapter 3 presupposes familiarity with the formal framework introduced in Chapter 2. Chapter 4 and Chapter 5 constitute the second part of the thesis. They address, from two different perspectives, the same general conceptual problem. They are, however, self-contained and can be read independently.

Chapter 2

From Oughts to Goals. Part I

This chapter focuses on (an interpretation of) the Enkratic principle of rationality, according to which rationality requires that if an agent sincerely and with conviction believes she ought to X, then X-ing is a goal in her plan. We analyze the logical structure of Enkrasia and its implications for deontic logic. To do so, we elaborate on the distinction between basic and derived oughts, and provide a multi-modal neighborhood logic with three characteristic operators: a non-normal operator for basic oughts, a non-normal operator for goals in plans, and a normal operator for derived oughts. We illustrate how this setting informs deontic logic by considering issues related to the filtering of inconsistent oughts and the restricted validity of deontic closure. The following chapter provides a dynamic extension of the logic by means of product updates, and investigates the stability of oughts and goals under dynamics.

2.1 Introduction

Suppose I believe sincerely and with conviction that today I *ought* to repay my friend Ann the 10 euro that she lent me. But I do not make any *plan for* repaying my debt: Instead, I arrange to spend my entire day at the local spa enjoying aromatherapy treatments. This seems wrong.

Enkrasia is the principle of rationality that rules out the above situation. The principle plays a central role within the domain of practical rationality, and has recently been receiving considerable attention in practical philosophy.[1] In its most general formulation, Enkrasia is the principle according to which

[1]See the works of Broome (2013); Kolodny (2005); Shpall (2013); Horty (2015). For a complementary account of the relation between oughts and plans, see Gibbard (2008).

rationality requires that if an agent sincerely and with conviction believes she ought to X, then she intends to X. There might be several ways in which such an intention to X is to be understood. Inspired by Bratman (1987), here we consider the agent's intention to X as indicating that the agent is committed to achieve X, and thus has, in some sense, a plan for X-ing. When this is the case, we say that X-ing is a goal in the agent's plan. Combining these aspects, we can understand Enkrasia as the principle of rationality requiring that if an agent sincerely and with conviction believes she ought to X, then X-ing is a goal in the agent's plan. Such interpretation of Enkrasia was first suggested by Horty (2015), and constitutes the starting point of the present chapter. To avoid confusion, we drop the term "intention" altogether.

This and the following chapters pursue two aims. Firstly, we want to analyze the logical structure of Enkrasia in light of the interpretation just described. This is, to the best of our knowledge, a largely novel project within the literature. Much existing work in modal logic deals with various aspects of practical rationality starting from Cohen and Levesque's seminal 1990 paper. The framework presented here aims to complement this literature by explicitly addressing Enkrasia. The principle, in fact, bears some non-trivial conceptual and formal implications — which might be of interest to the practical philosopher as well as the modal logician. This leads to our second aim. We want to address the repercussions that Enkrasia has for deontic logic. To this end, we elaborate on the distinction between so-called "basic oughts" and "derived oughts", and show how this distinction is especially meaningful in the context of Enkrasia. Moreover, we address issues related to the filtering of inconsistent oughts, the restricted validity of deontic closure, and the stability of oughts and goals under dynamics.

In pursuit of these two aims, we introduce a multi-modal neighborhood logic for Enkrasia. The logic has three characteristic operators: a non-normal operator for basic oughts, a non-normal operator for goals in plans, and a normal operator for derived oughts. Finally, we provide a dynamic extension of the logic by means of product updates.

This chapter proceeds along the following general lines. First, we clarify its philosophical foundations by introducing Enkrasia's main characteristics and its connection with two principles of rationality requiring goals in plans to be consistent (Section 2.2). We then introduce two challenges that illustrate the relevance of Enkrasia for deontic logic (Section 2.3). After discussing some core features and design choices of our approach (Section 2.4), we present a static logic for Enkrasia (Sections 2.5–2.7). This concludes the first part of our investigation. The second part will be devoted to a dynamic extension of the logic for Enkrasia, and will constitute the focus of the next chapter.

2.2 Enkrasia and the Consistency of Goals

The starting point of our investigation is the Enkratic principle of rationality, in the following interpretation:

> ENKRASIA. If an agent believes she ought to X, then X-ing is a goal in the agent's plan.

Such interpretation is inspired by Horty (2015). This section introduces ENKRASIA's main components and emphasizes its connection with two principles of rationality governing goals in plans. Let us stress before continuing that the aim pursued here is not to engage in a direct defense of ENKRASIA (for this, the interested reader can consult Broome, 2013 and Horty, 2015). Rather, this section is meant to lay the groundwork for our formal analysis of ENKRASIA's structure and of its position within the domain of practical rationality.

Let us begin with the oughts to which ENKRASIA applies — where "oughts" is used as a noun, roughly meaning obligations. It should be stressed that ENKRASIA does not take as antecedents all possible oughts. For one, ENKRASIA applies only to those oughts that are **believed by the agent** — in fact, this straightforwardly follows from the above formulation of the principle. However, further constraints are in place. We take inspiration from Broome (2013), and require that the oughts that fall within the scope of ENKRASIA have at least two further properties: They are **normative** and **ascribed to the agent herself**. These constraints are better illustrated via examples, so let us briefly consider them in turn.

One constraint limits the scope of ENKRASIA to normative oughts. These are the oughts that have to do, for instance, with morality, law or prudence. "I ought to repay my friend (as morality demands me so)" is an illustrative example of a normative ought. Contrariwise, examples of non-normative oughts are often to be found where oughts are used to express what is typically expected to be the case (see Yalcin, 2016), as in "I ought to have heard from the landing module ten minutes ago" (Broome, 2013, p.9). It would make little sense to say that hearing from the landing module is something I plan for. Indeed, ENKRASIA does not apply there.

The other constraint demands that the agent ascribes the oughts to herself. We can put this point in various ways: We can say this constraint demands that the agent believes the ought is required of her, that she recognizes it is her job to bring about the ought, or that she believes she is the "owner" of the ought (cf. Broome, 2013, p. 22). Examples of oughts ascribed to the agent herself are "I ought to get a sun hat" (Broome, 2013, p.12), and "I ought to see to it that the kids are alright". An ought that is *not* ascribed to the agent herself is "I ought to get a punishment", in a (natural) context where it is not on me to ensure that I receive this punishment. As long as getting a punishment is not my job, it would be incorrect to say that I fall short of rationality if getting punished is not a goal in my plan. This is why we demand ENKRASIA to apply only to oughts that are ascribed to the agent herself.

We have just identified a way in which ENKRASIA is constrained: It applies *only* to the oughts that enjoy the three properties above, namely, that are believed by the agent, normative, and ascribed to the agent herself. In our formal framework, we will implicitly assume that the oughts of ENKRASIA are of that kind. This is not to mean, however, that *all* oughts with those properties will correspond, via ENKRASIA, to goals in the agent's plans. In fact, in the next section, we will suggest that ENKRASIA needs to be further weakened.

So much for oughts. Let us now turn to another crucial component of ENKRASIA: Goals in plans. Drawing from Bratman (1987), when saying that X-ing is a goal in the agent's plan, we mean that the agent is *committed* to achieve X, which includes figuring out (to an appropriate degree) how to do so. To put it more succinctly, we mean that the agent has a plan *for* X-ing. For instance, repaying my friend is a goal in my plan only if I am committed to do so: I have a plan for repaying my friend which, minimally, for me rules out all the options (such as spending all my money, leaving the country, etc.) that I believe would make it impossible to achieve my goal. Those options become, given my commitment to repay my friend, no longer **admissible**. In this context, goals in plans differ from mere desires or wishes, which lack such a dimension of commitment (Thomason, 2000; Cohen and Levesque, 1990). Those notions should be kept apart here.

Furthermore, goals in plans are **future-directed**: The most natural reading of "X-ing is a goal" is the one in which X is something that still has to happen (see Bratman, 1987, p.4). Indeed, when talking about having a goal, we generally refer to something we are committed to do in the future (by the end of today, tomorrow, next month, etc.). In line with these considerations, we will assume X to include an element of futurity.

The literature imposes constraints on goals in plans. For instance, Broome suggests a property that — paraphrased in our own terms — amounts to requiring that the agent has the ability, via forming the goal to X, to have an impact on X-ing (Broome, 2013, pp.162-163). Although we find such a suggestion worth further (formal) analysis, we do not follow this direction here. Rather, we focus our attention on two minimal principles of rationality governing goals in plans. These principles of rationality require goals in plans to be **consistent**, in the following two senses of the term:

> INTERNAL CONSISTENCY. If X-ing, Y-ing, ... are goals in an agent's plans, then it is *logically* consistent to X and Y and ... .
>
> STRONG CONSISTENCY. If X-ing, Y-ing, ... are goals in an agent's plans, then the agent *believes* it is *possible* to X and Y and ... .

Both principles reflect the idea that it should be possible for goals in plans to be successfully achieved: INTERNAL CONSISTENCY demands that goals in plans be jointly logically consistent, while STRONG CONSISTENCY requires that goals in plans be jointly consistent with respect to the agent's beliefs. Their motivation is ultimately rooted in the dimension of commitment that goals in plans have: I could not truly be committed to repaying my friend and, at the same time, be committed to spending all my money to see the movies, while believing that these two things are incompatible — let alone jointly logically impossible (see Bratman, 1987; Cohen and Levesque, 1990; Horty, 2015).

Straightforward consequences of the above consistency principles are that if X-ing is a goal in a plan, then not X-ing is not a goal in a plan (from INTERNAL CONSISTENCY), and that if X-ing is a goal in a plan, then the agent believes it is possible to X (from STRONG CONSISTENCY). That is to say, goals in plans should neither be contradictory, nor believed to be impossible to achieve.

This is perhaps the right moment to mention some aspects of the current debate surrounding ENKRASIA — and the principles of rationality, more generally — that we will *not* address. The first has to do with the debate on whether principles of rationality are of wide or narrow scope. Consider ENKRASIA. Under the narrow scope, if the agent believes she ought to X, *rationality requires that* X-ing is a goal in the agent's plan. Under the wide scope, on the other hand, *rationality requires that* if the agent believes she ought to X, then X-ing is a goal in her plan. The two readings lead to different pictures of rationality.

Under the narrow scope reading, rationality requires a particular *attitude* of the agent. Under the wide scope, rationality only requires a particular *relation* between the agent's attitudes, typically leaving the rational agent leeway to either adopt X-ing as a goal in her plan or to revise her belief that she ought to X.[2] Since the focus of the project is not on operators akin to "rationality requires that", we take our contribution to be largely independent of the question whether ENKRASIA is a narrow or wide scope principle of rationality.

A second issue to which the present work does not contribute is whether principles of rationality are synchronic or diachronic. Consider again ENKRASIA, now enriched with time-indexes: If the agent believes at t that she ought to X, then X-ing is a goal in the agent's plan at t'. Diachronically, t precedes t'. Synchronically, t and t' refer to the same time. Thus, under the diachronic reading, believed oughts can be thought of *generating* corresponding goals; while, under the synchronic reading, believed oughts and goals *coexist* at the same time. For reasons of simplicity, we follow Broome (2013) and focus on the synchronic interpretation of ENKRASIA. We hold, however, that both interpretations have a certain appeal, especially from a logical perspective.

2.3 Challenges

We now introduce two (of the three) challenges surrounding ENKRASIA that are apt to illustrate the relevance such a principle holds for deontic logic. The third challenge will be considered in the following chapter.

2.3.1 Challenge I: From Inconsistent Oughts to Consistent Goals

There is a potential *tension* between ENKRASIA and the principles of INTERNAL and STRONG CONSISTENCY for goals in plans. Consider the following:

Example 2.1. *Suppose I believe I ought to repay 10 euro to my friend Ann. I also believe I ought to go to the movies with Barbara (I have promised her so). However, money is scarce, and I believe it is impossible to do both.*

It is safe to suppose that the oughts in Example 2.1 are of the kind to which

[2] See, among others, Broome (2013); Kolodny (2005) and Shpall (2013). Broome (2013) defends the wide scope reading, and Kolodny (2005) the narrow scope one, while Shpall (2013) proposes a "conciliatory view" between the two camps of the debate.

ENKRASIA may apply (i.e., they enjoy all three properties introduced in Section 2.2). Now if ENKRASIA were in fact applied to those oughts, I would need to plan for both repaying the money to Ann and for going to the movies with Barbara — ending up with two goals I believe to be *inconsistent*, and so violating STRONG CONSISTENCY in this specific case.

How to solve this tension? One way is to assume oughts are always consistent, both from a logical viewpoint and from the perspective of the agent's beliefs (see Broome, 2013). This assumption certainly solves the problem. But consider again the example above. Especially when oughts originate from different sources, it seems a viable possibility that these may end up being jointly inconsistent.

In what follows, we investigate another strategy to solve the tension between ENKRASIA, INTERNAL and STRONG CONSISTENCY. In a nutshell, this strategy is not to rule out the possibility of inconsistent oughts, nor to abandon the consistency principles for goals in plans, but rather to weaken ENKRASIA. The rationale for maintaining both INTERNAL and STRONG CONSISTENCY is rather pragmatic: In the face of a normative conflict about how to act, the least I can do is to assure that whatever I commit to is achievable.

Allowing for oughts, but not goals, to be inconsistent has several major consequences. Firstly, since oughts are possibly inconsistent but goals are not, it straightforwardly follows that not *all* oughts can correspond to goals in plans. In fact, this makes ENKRASIA a logically *invalid* principle. Secondly, it is natural to ask if not all, then *which* oughts do correspond to goals in plans. The challenge consists then in formally determining how oughts can be filtered out, in order to move from inconsistent oughts to consistent goals.

2.3.2 Challenge II: Basic Oughts and Derived Oughts

The second challenge revolves around a family of logical principles and inference rules that goes under the name of "deontic closure under implication" — for short: *deontic closure*. A longstanding tradition in deontic logic rejects the validity of deontic closure, arguing that it leads to unacceptable conclusions. An example is given by Ross' Paradox (Ross, 1941; Hilpinen and McNamara, 2013): Suppose *I ought to mail the letter*; now, since mailing the letter logically implies mailing the letter or burning it, deontic closure would imply that *I ought to mail the letter or burn it* — which is intuitively implausible.

The issue is that *even if* we accept that deontic closure is in fact problematic and should not be generally valid, an outright rejection of deontic closure would not constitute an adequate solution. For one, it would lead to miss out also on deontic inferences that *are* intuitively plausible.

To see this, consider the following example, which we owe to Horty (2015). For this example, let us forget about my promise to go to the movies with Barbara, and simply assume that going to the movies is something I like:

Example 2.2. *Suppose that I ought to repay Ann 10 euro. Now suppose that I would also like to go to the movies, but I do not have a lot of money. In fact, I believe that unless I refrain from going to the movies it is impossible to repay Ann. So, I conclude, I ought not go to the movies.*

Such a conclusion strikes us as impeccable. Following Von Wright (1963), we call the above piece of reasoning *practical inference*, and schematically represent it as:

(P1) I ought to repay Ann

(P2) Necessarily, repaying Ann implies not going to the movies

(C) Therefore, I ought not go to the movies

Practical inference is the cornerstone of instrumental reasoning.[3] Yet, practical inference — just as Ross' Paradox— is a variant of deontic closure (specifically, deontic closure under necessary implication). An outright rejection of deontic closure would have the effect of also blocking the above derivation.

The challenge then takes the following shape: Even assuming that deontic closure is not generally valid, a deontic logic should be "thick" enough to license crucial deontic inferences — including those instances of deontic closure that are valid. In the remainder of this section, we explore the boundaries between valid and invalid instances of deontic closure, and show that ENKRASIA provides us with the conceptual tools to do so.

All we need is to fix one *set* of oughts to start with. This set functions as input for the agent's deliberation. We do not impose any requirements on this set other than demanding that all oughts enjoy the three properties described in Section 2.2, i.e., being believed by the agent, normative, and ascribed to the agent herself. It follows, hence, that these are oughts to which ENKRASIA *may* apply. We call the oughts in this set **basic oughts**. Apart from what we

[3]Typically (P2) expresses a *practical necessity*, which might vary with the circumstances or the agent's beliefs thereof, cf. Von Wright (1963, p.161).

just said, there is nothing intrinsically special about these.[4] We do not assume basic oughts to have any particular surface grammar, nor do we assume they share any further commonalities. In fact, we even admit the possibility that basic oughts are jointly inconsistent.[5] Once the set of basic oughts is fixed, we call **derived oughts** those oughts that are *implied* by basic oughts.

The distinction between basic and derived oughts is crucially meaningful in relation to ENKRASIA, and helps us to discern valid from invalid instances of deontic closure. Let us take practical inference as a case study. The central observation — originally noticed by Horty (2015) — is that the oughts in (P1) and in (C) interact differently with ENKRASIA. Suppose I deliberate about my day and I take as input that *I ought to repay Ann* (P1). In the absence of conflicts, this *basic ought* leads via ENKRASIA to the goal of repaying Ann. Something that I plan for in itself. From there, via deontic closure, I do well in deriving that *I ought not go to the movies* (C). However this *derived ought* does not interact with ENKRASIA in the same way: refraining from going to the movies is *not* a goal *in its own right*. Rather, it is something I necessarily have to do in order to fulfill my goal of repaying Ann. In other terms, the derived ought registers the *necessary* (though possibly not sufficient) *conditions* for the fulfillment of such a goal (see also Brown, 2004).

It is with respect to ENKRASIA that the different roles played by basic and derived oughts become evident. This motivates taking basic and derived oughts as *two* separate kinds of oughts in this context. Once these are understood as two separate oughts, having different logical meanings, it becomes non-trivial to say that there are instances of deontic closure that move from basic oughts to derived oughts. These instances will be valid in our logic. As elaborated above, this bears crucial implications for practical inference. Similar considerations apply to Ross' Paradox. Acknowledging the different roles of the oughts involved, I do well in deriving that *I ought to mail the letter or burn it* only to the extent that this expresses no more than the (logically) necessary — but not sufficient — conditions for the fulfillment of my goal of mailing the letter. In other terms, my inference is only valid to the extent that *I ought to mail*

[4]Various interpretations can be imposed on the set of basic oughts. Horty (2015) thinks of basic oughts as those oughts *directly* generated by normative requirements. Alternatively, one may think of basic oughts as those *explicitly* believed by the agent.These interpretations are compatible with our characterization of basic oughts. An alternative characterization of basic oughts is provided by Nair (2014).

[5]This is why we have stressed that basic oughts are oughts to which ENKRASIA *may* apply. Since basic oughts are possibly inconsistent, while goals are not, it follows that not all basic oughts correspond to goals. See our discussion in Section 2.3.1.

the letter or burn it is a derived ought.

2.4 Introducing the Framework

We can turn now to the first aim of this work: providing a logical framework for ENKRASIA. For the analysis, we posit a set of **minimal requirements** about basic oughts, goals in plans, and derived oughts. The reader may find these incomplete. However, our aim here is not to reveal the full logical principles governing oughts or goals. Rather, we aim for a minimal set of axioms strong enough to identify relations between basic oughts, goals and derived oughts that result from our analysis of ENKRASIA. Working towards a more complete logic of oughts and goals, additional axioms could be added in the future, validating further theorems. In such a stronger logic, the relationships identified here would continue to hold.

Despite the intended minimalism, devising a logic for ENKRASIA requires a variety of conceptual and formal choices. Some of these are core features of the framework developed. Others are mere design choices that could be altered easily. The following discussion details both.

2.4.1 Core Features

Basic oughts, goals and derived oughts. The framework's first core component is three main logical operators: A modal operator for *basic oughts*, one for *goals* in plans, and finally one for *derived oughts.* We implicitly assume basic oughts to satisfy the three conditions identified in Section 2.2: they are believed by the agent, normative, and ascribed to the agent herself. Moreover, we take oughts and goals to be future-looking, referring to future states of affairs to be brought about.

Information states. The framework focuses on a single moment in time, specifically, where the agent deliberates on what to do. The choice options represented in the logic are those believed possible by the agent; they form, in some sense, her *information state.*[6] In fact, the framework with its various components is fully relative to the agent's beliefs, and so can do without any

[6]Unlike in most epistemic frameworks, this information state does not list epistemic possibilities the agent cannot distinguish between, but a set of possible options the agent can choose from.

explicit doxastic operators.

A thin logic for basic oughts and goals. The starting point of the framework is a set of basic oughts. At present, the logic governing basic oughts remains thin. We do not assume basic oughts to have any logical structure such as being closed under implication or pairwise intersections, nor do we require the content of a basic ought to be satisfiable, even in principle. The only requirement made is that basic oughts are independent of their exact description, i.e., the agent's set of basic oughts is closed under replacement of logically equivalent formulas.[7] The logic of basic oughts, hence, will turn out *weaker than normal* in the logical sense: it will be a neighborhood modal logic (cf. Pacuit, 2017). Similar considerations apply to goals. The set of goals in the agent's plans will be a *consistent* subsets of her basic oughts. Hence, also the goal modality will turn out to be a non-normal neighborhood operator.

A thicker logic for derived oughts. While assuming the logic of basic oughts and goals to be thin, the resulting neighborhood logic is strong enough to license crucial deontic inferences. Derived oughts play a central role in such reasoning. To illustrate how these are represented in the framework, we first note that the agent may be committed to multiple goals in parallel. Following the principles of INTERNAL CONSISTENCY and STRONG CONSISTENCY, these goals are required to be jointly consistent. Put formally, this means that there must exist some possible course of events that satisfies all of the agent's goals. We call such courses of events *admissible*. Derived oughts, then, denote those properties that all admissible courses of events have in common. In other words, derived oughts indicate the necessary (but possibly not sufficient) conditions for the fulfillment of *all* the agent's goals. Derived oughts, unlike basic oughts, hence follow a *normal* modal logic.

2.4.2 Design Choices

Branching temporal trees. Oughts and goals, we have said, are future-looking. Correspondingly, the agent's relevant choices when deliberating on what to do are between possible future courses of events. In the present frame-

[7]We are arguably omitting certain structural properties of basic oughts. For instance, a plausible further requirement to impose on basic oughts could be what Cariani (2016) calls "weakening": $O\varphi \wedge O\psi \vDash O(\varphi \vee \psi)$. Nevertheless, we will show that the present framework contains sufficient structure for a logical analysis of ENKRASIA.

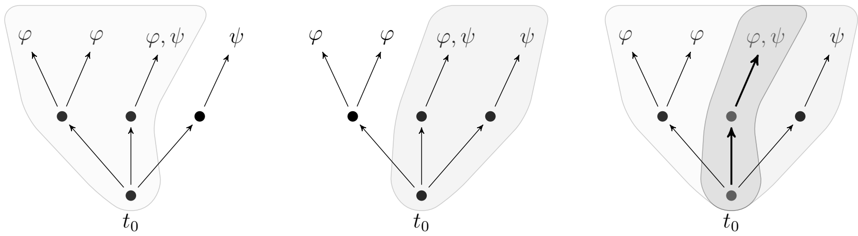


Figure 2.1: Left: The subtree compatible with the satisfaction of the agent's basic ought and goal that φ (gray). Middle: The subtree compatible with the satisfaction of agent's basic ought and goal that ψ (gray). Right: Interaction of both basic oughts and goals (dark gray). Bold arrows denote the admissible subtree, i.e., the courses of events compatible with the satisfaction of both goals φ and ψ.

work a fine-grained perspective on such future courses of events is assumed, representing the relevant temporal structure explicitly. To this end, all possible future unfoldings of the world are recorded in a temporally branching tree, where each maximal branch — each *history* — corresponds to a possible future course of events. For an illustration of a branching time setting, see Figure 2.1.

In accordance with this fine-grained perspective, oughts and goals need to be expressed in an adequate formal language rich enough to capture their temporal structure. To this end, the framework involves a temporal logic that can express, for instance, that certain states of affairs should always be avoided, reached at least once or maintained throughout.

Notably, representing possible courses of events as temporally extended histories is not strictly necessary. For the static part of the logic (Section 2.7), it would suffice to treat each possible course of events as a single state, giving rise to a more classic neighborhood logic. It is only in the dynamic extension of Section 3.3 that the temporal structure becomes relevant.

From basic oughts to goals. An agent's goals, we have said, form a subset of her basic oughts. The latter, however, are potentially inconsistent which goals are not. A central component for the transition from basic oughts to goals will hence be (maximally) consistent subsets of basic oughts, as these guarantee the principles of INTERNAL CONSISTENCY and STRONG CONSISTENCY to be satisfied.[8] Note, however, that there can be multiple maximally consistent

[8] A competing notion of consistency, which we might call *free choice consistency*, is dis-

subsets of basic oughts. So how are goals related to maximally consistent sets of basic oughts? There exist at least two viable ways of approaching this:

- In a strict reading, a basic ought is adopted as a goal if it is contained in *every* maximally consistent set of basic oughts.
- In a more tolerant approach, a basic ought is adopted as goal if it is contained in *some specific* maximally consistent set of basic oughts.

The tolerant approach will, in general, lead to more goals than the strict approach. In fact, by picking a single maximally consistent subset of basic oughts, it guarantees the agent to do the best she can in terms of adopting a multitude of goals without violating consistency.[9] The following analysis follows the tolerant approach.[10] We are hence in need of a mechanism for selecting which maximally consistent set of basic oughts corresponds to goals.

Linear priority on basic oughts. For selecting a maximally consistent subset of basic oughts, we assume the latter to be ordered linearly.[11] By means of the lexicographic order (cf. Definition 2.11), this linear order extends to a priority ordering among sets of basic oughts. The agent then adopts the highest ranked maximally consistent subset of her basic oughts as goals.[12] The

cussed in Veltman (2011). Veltman would consider $\mathrm{O}(\neg\varphi)$ and $\mathrm{O}(\varphi \vee \psi)$ inconsistent, as the former violates the free choice expressed by the latter. We do not deal with free choice, and hence we limit ourselves to a classic account of consistency. Free choice in the context of planning is considered in Marra and Klein (2015).

[9]This mirrors, from a formal point of view, the question on how *reasons* accrue to support all-things-considered oughts (cf. Horty, 2012; Nair, 2014, 2016).

[10]The strict approach is prominently pursued by Kratzer (2012c) in her seminal approach to the semantics of deontic operators. There, a possibly inconsistent set of normative requirements N creates an ideality ordering on a set W of possible worlds. To define the ordering, let $N(w)$ for a world w be the set of normative requirements from N satisfied at w. The ordering is then defined by $w > v$ (read "w is more ideal than v") if $N(w) \supset N(v)$. A deontic necessity statement $\Box_d\varphi$, finally, holds true in the framework if φ is satisfied in *all* $>$-maximal worlds. Notably, $>$-maximality is tightly related to maximally consistent subsets. More specifically, world w is $>$-maximal iff no M with $N(w) \subset M \subseteq N$ is satisfiable in any $v \in W$, i.e., iff $N(w)$ is maximally W-consistent. It follows that $\Box_d\varphi$ is true iff φ holds in *all* intersections of maximally consistent subsets of norms. This is exactly the above strict reading.

In fact, various aspects of Kratzer's approach have counterparts in the present framework. To make these explicit: normative requirements N and possible worlds correspond to basic oughts and histories of tree $\mathcal{T}$ respectively. The deontic necessity operator, finally, corresponds to our modality for derived oughts.

[11]Hence, although we do not rule out the possibility of having both $\mathrm{O}\varphi$ and $\mathrm{O}\neg\varphi$ as basic oughts, we exclude irresolvable dilemmas. One basic ought must take priority over the other.

[12]As mentioned above, the framework is relative to the agent's beliefs. Hence, we do not

current framework is, however, modular in this respect. Any other mechanism for picking out one element from any given set of maximally consistent set of oughts would function just as well. In fact, the choice of selection mechanism does not have any impact of the static analysis of Sections 2.6 and 2.7. In particular, the assumption of oughts being ordered *linearly* is non-substantial for the present purpose.

2.4.3 Towards a Logic for Enkrasia

The construction of our formal framework proceeds as follows. Sections 2.5–2.7 define two static logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$. Having modalities for basic oughts, goals and derived oughts, these already incorporate ENKRASIA through a number of axioms regulating the relationship between the three components. The second of these logics offers an additional global modality $\Box$ allowing the agent to reason about which options are available to her.

2.5 The Language

To begin, let us specify the logical language used. The construction proceeds in several steps. First, we define two languages $\mathcal{L}_0$ and $\mathcal{L}_1$ to talk about present and future states of affairs. This language will serve to express *the content* of oughts and goals. Afterwards, we introduce language $\mathcal{L}_2$ that allows to reason about basic and derived oughts, goals and their interaction.

Definition 2.3. *Let* At *be a finite or countable set of atomic propositions. The* **basic language** $\mathcal{L}_0$ *is given by the standard language of propositional logic combined with a future-tensed operator* F. *It is defined by the following BNF:*

$$\psi := p|\neg\psi|\psi \wedge \psi|F\psi$$

for $p \in \mathsf{At}$. The intended reading of modal expressions $F\psi$ is "ψ is true at least once in the future". We denote the dual of F by G. $G\psi$ hence reads as "ψ is always true in the future". Operators $\to$ and $\vee$, finally, are defined as usual.

exclude the possibility that the agent has quite extravagant beliefs on which basic oughts take the precedence over the rest. This is perhaps a case in which the wide-scope reading of ENKRASIA fits better than the narrow-scope, as it does not lead to the conclusion that rationality requires of the agent to adopt as a goal something ultimately stemming from her misled priority relation.

It is convenient to consider the future looking fragment of $\mathcal{L}_0$:

Definition 2.4. *The* **language** $\mathcal{L}_1$ *is the fragment of* $\mathcal{L}_0$ *containing only future-tensed formulas where every atomic proposition is in the scope of a temporal operator. Formally,* $\mathcal{L}_1$ *is defined as follows:*

$$\varphi := F\psi|\neg\varphi|\varphi \wedge \varphi$$

for $\psi \in \mathcal{L}_0$. Building on $\mathcal{L}_1$, the modal language for reasoning about basic oughts, goals in plans, and derived oughts can be defined.

Definition 2.5. *The modal* **language** $\mathcal{L}_2$ *is given by the following BNF:*

$$\varphi := p|O\psi|Goal\psi|D\psi|\neg\varphi|\varphi \wedge \varphi$$

for $p \in \mathsf{At}$ and $\psi \in \mathcal{L}_1$. The intended reading of the three modal operators is the following: $O\varphi$ reads as "φ is a basic ought", $Goal\varphi$ as "φ is a goal in a plan", and finally $D\varphi$ reads as "φ is a derived ought". Again, operators $\rightarrow$ and $\vee$ are defined as usual.

Two observations about $\mathcal{L}_2$ are in order. Firstly, the language does not allow for iterated modalities. This is a feature shared with several other systems of deontic logic. Secondly, being built over the temporal fragment $\mathcal{L}_1$ of $\mathcal{L}_0$, the modal language $\mathcal{L}_2$ only allows for basic oughts, goals and derived oughts to scope over future-tensed formulas. Our oughts and goals are, as we have said, future-looking.

2.6 Semantics

Before introducing logical principles on the above languages, we specify the intended semantic structures for basic oughts, goals, and derived oughts. Section 2.7 then provides an axiomatization that is sound and complete with respect to the semantics introduced here. We begin our analysis by introducing trees, delineating how the agent envisages the possible unfoldings of future events.

Definition 2.6. *A* ***tree*** *is an ordered set* $\mathcal{T} = \langle T, \prec_\mathcal{T} \rangle$ *where* T *is a set of moments and* $\prec_\mathcal{T}$ *a tree-order on* T. *We make two additional assumptions about* $\prec_\mathcal{T}$. *First, the tree order is assumed to have a* **root**, *i.e., a minimal element* t_0 *satisfying* $t_0 \prec_\mathcal{T} t$ *for all* $t \neq t_0$. *Second,* $\prec_\mathcal{T}$ *is also serial, i.e.,* *every moment must have at least one successor.*[13] *A* ***history*** *h, finally, is a maximal linearly ordered subset of $\mathcal{T}$.*

Intuitively, t_0 indicates the current time step, i.e, the moment at which the agent ponders what to do. Notably, a tree is the union of its histories, i.e., $\mathcal{T} = \bigcup\{h \subseteq \mathcal{T} \mid h \text{ history}\}$. We will make heavy use of this later. To ease terminology, we will use the term **subtree** for any tree $\mathcal{T}'$ that is of the form $\bigcup_{h \in Hist} h$ with $Hist$ a set of histories of $\mathcal{T}$. We will denote the set of subtrees of $\mathcal{T}$ by $\mathcal{P}(\mathcal{T})$. Lastly, let $\mathcal{T}'$ and $\mathcal{T}''$ be subtrees of $\mathcal{T}$ given by $\mathcal{T}' = \bigcup_{h \in Hist'} h$ and $\mathcal{T}'' = \bigcup_{h \in Hist''} h$ respectively. Then define the **intersection subtree** $\mathcal{T}' \Cap \mathcal{T}''$ of $\mathcal{T}$ as the subtree generated by $Hist' \cap Hist''$, i.e., $\mathcal{T}' \Cap \mathcal{T}'' := \bigcup_{h \in Hist' \cap Hist''} h$.[14]

Based on the definition of a tree, we can define a tree model for our temporal language $\mathcal{L}_0$.

Definition 2.7. *A* **pointed tree model** *is a tuple $\mathcal{M} = \langle \mathcal{T}, t_0, v \rangle$ where $\mathcal{T} = \langle T, \prec_{\mathcal{T}} \rangle$ is a tree, t_0 the distinguished time, i.e., the root of $\mathcal{T}$, and $v : \mathsf{At} \to \mathcal{P}(T)$ is a valuation function that maps each atomic proposition of the background language into a set of moments of $\mathcal{T}$.*

A pointed tree models provides a semantics for language $\mathcal{L}_0$:

Definition 2.8. *Let $\mathcal{M}$ be a pointed tree model. The* **evaluation** *of formulas of $\mathcal{L}_0$ on time-history pairs t/h with $t \in h$ of $\mathcal{M}$ is defined as follows:*

- $\mathcal{M}, t/h \models p$ *iff* $t \in v(p)$ *for* p *atomic*
- $\mathcal{M}, t/h \models \neg\varphi$ *iff* $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ *iff* $\mathcal{M}, t/h \models \varphi$ *and* $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models F\varphi$ *iff there is a* $t' \in h$ *such that* $t \prec_{\mathcal{T}} t'$ *and* $\mathcal{M}, t'/h \models \varphi$

Finally, we say that a formula is true at t simpliciter iff it is true at t/h' for all histories h' passing through t.

Definition 2.9. *Let $\varphi \in \mathcal{L}_0$ and $t \in \mathcal{T}$. The proposition expressed by φ at t, i.e., the* **truth subtree** *$[\![\varphi]\!]^t$, is defined as follows:*

$$[\![\varphi]\!]^t = \bigcup\{h | t \in h \text{ and } M, t/h \models \varphi\}$$

[13] This definition remains silent about the exact shape of a tree. It allows for finite as well as infinite branchings and also for discrete as well as dense orders.

[14] Note that $\mathcal{T}' \Cap \mathcal{T}'' \subseteq \mathcal{T} \cap \mathcal{T}$. In general, however, $\mathcal{T}' \Cap \mathcal{T}''$ is a proper subset of $\mathcal{T} \cap \mathcal{T}$.

Towards developing a semantics for the language $\mathcal{L}_2$, we finally extend tree models with neighborhoods representing the agent's basic oughts. A central component of these extended models will be sets of the form $[\![\varphi]\!]^{t_0}$, representing the truth set of φ as seen from the moment of deliberation t_0.

Definition 2.10. *An* **enkratic model** *is a tuple* $\mathcal{M} = \langle \mathcal{T}, t_0, v, N_O, \succ_O \rangle$ *where* $\langle \mathcal{T}, t_0, v \rangle$ *is a pointed tree model, and* $N_O \subseteq \mathcal{P}(\mathcal{T}) \times \mathcal{L}_1$ *is a neighborhood with the additional condition that* $(\mathcal{T}', \varphi) \in N_O$ *implies that* $\mathcal{T}' = [\![\varphi]\!]^{t_0}$*. Finally* $\succ_O$ *is a conversely well-founded linear order on the set of all* φ *such that* $([\![\varphi]\!]^{t_0}, \varphi) \in N_O$*.* [15]

Presently, we are only interested in the agent's basic oughts at the time of reasoning t_0. We can represent these with a set of treelike neighborhoods N_O listing all the basic oughts the agent is exposed to at t_0.[16] It might seem counterintuitive to represent a basic ought by a subset-formula pair $([\![\varphi]\!]^{t_0}, \varphi)$ rather than simply a subtree $[\![\varphi]\!]^{t_0}$. The reason for this will become clear in the next chapter (Section 3.3) where dynamics enters the picture. Briefly, two propositions φ and ψ may be co-extensional in the current tree, but might cease to be so once new information about the world is acquired. For this case, it is necessary to keep track of whether the basic ought prescribes that φ or ψ.

On a given enkratic model, we can construct additional structures related to the semantics of goals and derived oughts. The first is the **goal-neighborhood** $N_G \subseteq \mathcal{P}(\mathcal{T})$. For the construction we recall the definition of a lexicographic order.

Definition 2.11. *Let* $\succ_O$ *be a conversely well-founded linear order on a set of formulas* $\Psi \subseteq \mathcal{L}_1$*. Then the* **lexicographic order** $\succ_{Lex}$ *on the power set* $\mathcal{P}(\Psi)$ *is defined by* $X \succ_{Lex} Y$ *iff there is some* $x \in X$*,* $x \notin Y$ *such that*

$$\{z \in X \mid z \succ_O x\} = \{z \in Y \mid z \succ_O x\}.$$

In other words, x *is the* $\succ_O$*-most important element on which* X *and* Y *disagree.*

[15] Where $\succ_O$ is a conversely well-founded linear order if and only if it is antisymmetric, transitive, total and every subset $B \subseteq \{\varphi \mid ([\![\varphi]\!]^{t_0}, \varphi) \in N_O\}$ has a $\succ_O$-maximal element.

[16] The approach could be extended to include the agent's basic oughts along all moments of a tree. Such an extension requires additional conceptual work, as basic oughts may, for instance, get discarded once they have been satisfied. Also, an extension will need to specify what happens to basic oughts in future moments where they have become unsatisfiable for pragmatic or principal reasons. Technically, such an extension would work by replacing the neighborhood N_O with a neighborhood function $n_O : \mathcal{T} \to \mathcal{P}(\mathcal{P}(\mathcal{T}) \times \mathcal{L}_1)$.

The goal neighborhood $N_G \subseteq \mathcal{P}(\mathcal{T})$ is determined by three conditions. First, the goals in an agent's plan must be derived from basic oughts. Second, the set of goals in a plan should be consistent. The third condition, finally, expresses that the set of goals is chosen optimally, given the agent's priority relation $\succ_O$ between her basic oughts. Formally, the conditions on N_G are:

i) $N_G \subseteq \{[\![\varphi]\!]^{t_0} \mid ([\![\varphi]\!]^{t_0}, \varphi) \in N_O\}$.

ii) N_G is maximally consistent, i.e.,

  a) there is some history h of $\mathcal{T}$ with $h \subseteq [\![\varphi]\!]^{t_0}$ for all $[\![\varphi]\!]^{t_0} \in N_G$ and

  b) whenever $N_G \subset Y \subseteq \{[\![\varphi]\!]^{t_0} \mid ([\![\varphi]\!]^{t_0}, \varphi) \in N_O\}$ there is no history h' with $h' \subseteq [\![\varphi]\!]^{t_0}$ for all $[\![\varphi]\!]^{t_0} \in Y$.

iii) N_G is $\succ_O$-maximal, i.e., whenever Y satisfies i) and ii) then

$$\{\varphi \mid [\![\varphi]\!]^{t_0} \in N_G\} \succ_{Lex} \{\varphi \mid [\![\varphi]\!]^{t_0} \in Y\};$$

where $\succ_{Lex}$ is the lexicographic order on $\mathcal{P}(\{\varphi \mid ([\![\varphi]\!]^{t_0} \varphi) \in N_O\})$ induced by $\succ_O$. (Cf. Definition 2.11).

Note that the three conditions uniquely determine the neighborhood N_G which is therefore well-defined.

From N_G the third central component of enkratic models —besides basic oughts and goals— can be defined. Let us begin by introducing what we call the **admissible subtree $\mathcal{T}_{\mathbf{Goal}}$**. The admissible subtree $\mathcal{T}_{Goal}$, briefly, is the intersection of the various subtrees corresponding to the agent's goals. Hence, it consists of all those histories that guarantee *all* of the agent's goals to be satisfied. It is from this admissible subtree that the agent's derived oughts are determined. Derived oughts indicate what holds in $\mathcal{T}_{Goal}$, and therefore can be thought of as expressing the necessary conditions for the fulfillment of all the agent's goals. To state things formally, the admissible subtree is defined as

$$\mathcal{T}_{\mathbf{Goal}} := \bigcap_{[\![\varphi]\!]^{t_0} \in N_G} [\![\varphi]\!]^{t_0}.$$

From the properties of N_G, it follows that $\mathcal{T}_{Goal}$ is non-empty. Having defined $\mathcal{T}_{Goal}$, we can give the semantic conditions turning enkratic model into models for language $\mathcal{L}_2$. Unlike $\mathcal{L}_0$, the language $\mathcal{L}_2$ is evaluated on moments t rather than time-history pairs t/h. We take this to be a natural condition, as $\mathcal{L}_2$ represents the agent's oughts and goals at a moment in time t_0 where she has not yet acted on any particular course of events, i.e., any history h. The following definition builds on the evaluation of $\mathcal{L}_0$ (and hence $\mathcal{L}_1$) on pointed tree models, cf. Definition 2.8.

Definition 2.12. *The* **evaluation** *of $\mathcal{L}_2$ on an enkratic model $\mathcal{M}$ is given by the following clauses:*

- $\mathcal{M}, t \models p$ *iff* $t \in v(p)$ *for* p *atomic*
- $\mathcal{M}, t \models \neg\varphi$ *iff* $\mathcal{M}, t \not\models \varphi$
- $\mathcal{M}, t \models \varphi \wedge \psi$ *iff* $\mathcal{M}, t \models \varphi$ *and* $\mathcal{M}, t \models \psi$
- $\mathcal{M}, t \models O\varphi$ *iff* $([\![\varphi]\!]^{t_0}, \varphi) \in N_O$
- $\mathcal{M}, t \models Goal\varphi$ *iff* $[\![\varphi]\!]^{t_0} \in N_G$ *and* $([\![\varphi]\!]^{t_0}, \varphi) \in N_O$.
- $\mathcal{M}, t \models D\varphi$ *iff* $\mathcal{T}_{Goal} \subseteq [\![\varphi]\!]^{t_0}$

Notably, the semantics of operators $O, Goal$ and D does not depend on the moment t of evaluation, but only on the initial time t_0. These modalities, hence, are meant to represent the agent's basic oughts, goals and derived oughts *at the time of deliberation* t_0.

In sum, the semantics of all three modalities supervenes on two components of the model: The neighborhood N_O and the priority ordering $\succ_O$. While the semantics of O, the basic ought modality, is directly given by N_O, the $Goal$ modality's neighborhood is derived by having $\succ_O$ pick a maximally consistent subset of N_O. This goal neighborhood, in turn, defines the D derived ought modality's admissible subtree by means of intersection.

2.7 Syntax: Axioms and Results

In this section, we provide an axiomatization for the various languages introduced in Section 2.5. We start with axioms for the temporal languages $\mathcal{L}_0$ and $\mathcal{L}_1$.

K_G	$G(\varphi \to \psi) \to (G\varphi \to G\psi)$
4	$G\varphi \to GG\varphi$
L	$F\varphi \wedge F\psi \to (F(\varphi \wedge \psi) \vee F(\varphi \wedge F\psi) \vee F(\psi \wedge F\varphi))$
D_G	$\neg G\bot$

These are accompanied by the classic necessitation rule:

$$\frac{\vdash \varphi}{\vdash G\varphi}\ \textsc{Nec}_G$$

The first two axioms are the standard K and 4 axioms, expressing that G is a normal modal operator and that the 'later' relation is transitive. The third axiom L reflects the fact that histories are linear, expressing that two future events φ and ψ will either be simultaneous, or that one comes after the other. Finally, the D-style axiom D_G expresses that time never ends, as there always is a future moment. We denote by Λ_{temp} the temporal logic over language $\mathcal{L}_0$ generated by K_G,4,L, D_G and G-necessitation $\textsc{Nec}_G$.

Next, we turn to the extended language $\mathcal{L}_2$. Operators O and *Goal* only have a limited logical structure. Reflecting the actual content of oughts issued by a normative source, we do not presuppose any logical requirements on basic oughts other than being invariant under replacement with logical equivalents. This is the content of:

$$\frac{\vdash \varphi \leftrightarrow \psi}{\vdash O\varphi \leftrightarrow O\psi}\ \textsc{Int}_O$$

The corresponding intensionality condition for the *Goal* operator also holds, as is shown in Lemma 2.15. While goals and basic oughts are not closed under logical reasoning, derived oughts are. In particular, the D-operator is normal and non-trivial, as expressed by the following axioms:

$$\begin{array}{ll} \mathrm{K}_D & D(\varphi \to \psi) \to (D\varphi \to D\psi) \\ \mathrm{D}_D & \neg D\bot \end{array} \qquad \frac{\vdash \varphi}{\vdash D\varphi}\ \textsc{Nec}_D$$

Lastly, and most importantly, the logic is guided by three interaction axioms describing the interplay between goals, basic and derived oughts. It is these principles that embody the ENKRASIA principle in the logic.

$$\begin{array}{ll} \mathrm{GO} & Goal\varphi \to O\varphi \\ \mathrm{GD} & Goal\varphi \to D\varphi \\ \textsc{Max} & O\varphi \wedge \neg Goal\varphi \to D\neg\varphi \end{array}$$

The first of these expresses that basic oughts are the only admissible sources of goals in the agent's plan. Every *G*oal follows from a basic *O*ught. The second axiom, GD, is a weak converse, saying that every *G*oal gives rise to a corresponding *D*erived ought. Most importantly, the third axiom, MAX, embodies the **bounded validity** of ENKRASIA. This can best be seen from its counterpositive $\neg D\neg\varphi \to (O\varphi \to Goal\varphi)$: if it is not the case that already $\neg\varphi$ is a derived ought, then if φ is a basic ought, φ is also a goal. Hence, in combination with K_D and D_D, MAX states that every basic ought has a corresponding goal *unless* this causes a violation of consistency. Put semantically, MAX expresses that the set of goals is a *Max*imally consistent subset of the agent's basic oughts.[17]

Definition 2.13. *The* **Enkrasia logic** Λ_{Enkr} *on language* $\mathcal{L}_2$ *is defined by all propositional tautologies together with the axioms* K_G,4,L,$\mathrm{D}_G, \mathrm{K}_D, \mathrm{D}_D$, GO, GD, MAX *and the rules* $\mathrm{INT}_O, \mathrm{NEC}_G$ *and* NEC_D *(cf. Table 2.1).*

Before moving on to completeness, let us take a moment to derive a number of consequences of the above axioms. First, we note that whenever an agent has a goal to φ, her derived oughts contain all logical consequences of φ. This follows immediately from axioms K_D and GD together with NEC_D.

Fact 2.14.

$$\frac{\vdash \varphi \to \psi}{\vdash Goal\varphi \to D\psi}$$

Second, we note that the *Goal* operator is closed under replacement with logical equivalents:

Lemma 2.15.

$$\frac{\vdash \varphi \leftrightarrow \psi}{\vdash Goal\varphi \leftrightarrow Goal\psi}$$

Proof. Assume $\vdash \varphi \leftrightarrow \psi$. For a contradiction, also assume that $Goal\varphi$ but $\neg Goal\psi$. By GO we have $O\varphi$ and hence by INT_O also $O\psi$. Hence we have $O\psi \wedge \neg Goal\psi$ which implies $D\neg\psi$ by MAX. On the other hand, GD implies $D\varphi$. By NEC_D and K_D this implies $D(\varphi \wedge \neg\psi)$ which, again by K_D, implies $D\bot$ contradicting D_D. □

Next, note that the logic does not demand an agent's basic oughts to be jointly consistent. Our agent may, for instance, believe both $O\varphi$ and $O\neg\varphi$ simultaneously. The set of goals, however, is required to be internally consistent.

Lemma 2.16. *Let* $\Lambda \subseteq \mathcal{L}_2$ *be a consistent set and let* $S = \{\varphi \in \mathcal{L}_1 \mid Goal\varphi \in \Lambda\}$ *Then* $S \nvdash_{\Lambda_{temp}} \bot$.

Proof. Assume for a contradiction that $S \vdash_{\Lambda_{temp}} \bot$. Since all its axioms correspond to first order expressible frame conditions, Λ_{temp} is compact (cf. Blackburn et al., 2001, Chapter 2.4). Hence there is a finite $S_0 \subseteq S$ such that $S_0 \vdash_{\Lambda_{temp}} \bot$. By Fact 2.14, we have $\{Goal\varphi | \varphi \in S_0\} \vdash_{\Lambda_{Enkr}} \bigwedge_{\varphi\in S_0} D\varphi$. By K_D

[17] If we had instead chosen the strict principle of translating basic oughts into goals, (cf. Section 2.4.2) i.e., only taking those basic oughts that are contained in all maximally consistent subsets instead, MAX would need to be replaced by the weaker $O\varphi \wedge \neg Goal\varphi \to \neg D\varphi$

we then get $\{Goal\varphi | \varphi \in S\} \vdash_{\Lambda_{Enkr}} D \bigwedge_{\varphi \in S_0} \varphi$, i.e., $\{Goal\varphi | \varphi \in S\} \vdash_{\Lambda_{Enkr}} D\bot$ — contradicting D_D. □

An immediate consequence is that the *Goal* operator satisfies the *D*-axiom, i.e.,

$$\vdash Goal\varphi \rightarrow \neg Goal\neg\varphi$$

In fact, this consistency requirement is solely responsible for discrepancies between basic oughts and goals. By MAX, whenever $O\varphi \wedge \neg Goal\varphi$ hold at some state w, this is because $Goal\varphi$ could not have been consistently added to the set of present goals, as it would require both $D\varphi$ and $D\neg\varphi$ to hold simultaneously.

Having specified our treatment of ENKRASIA, it is now time to present a general characterization result. However, before being able to do so, we need to make an extra assumption about enkratic models. We assume the neighborhood N_O to be closed under logical equivalence. That is, if φ and ψ are logically equivalent in Λ_{temp} and $(\llbracket\varphi\rrbracket^{t_0}, \varphi) \in N_O$ then also $(\llbracket\psi\rrbracket^{t_0}, \psi) \in N_O$. It follows immediately that also N_G is closed under Λ_{temp} logical equivalence. With this assumption, we can show the following characterization result, which is proved in the appendix.

Theorem 2.17. *The logic Λ_{Enkr} is sound and complete with respect to the class of enkratic models.*

2.7.1 Enriching the Language: A Global Modality

Note that language $\mathcal{L}_2$ suffers from what might be perceived as a lack of expressive power. So far, $\mathcal{L}_2$ can express whether the agent is under a certain basic ought that φ and whether this ought translates into a goal. What $\mathcal{L}_2$ cannot yet express is whether the agent considers φ possible in the first place, i.e., whether she believes her basic ought that φ to be satisfiable. To remedy this, we add a new modal operator $\Box$, where $\Box\varphi$ for some $\varphi \in \mathcal{L}_1$ is to express that φ holds in all possible histories. As usual, $\Diamond$ stands for the dual of $\Box$. So $\Diamond\psi$ expresses that there is a possible ψ-history or, at least, one the agent considers possible. To incorporate $\Box$, we expand language $\mathcal{L}_2$ to $\mathcal{L}_\Box$ given by the BNF:

$$\varphi := p | O\psi | Goal\psi | D\psi | \Box\psi | \neg\varphi | \varphi \wedge \varphi$$

for $p \in \mathsf{At}$ and $\psi \in \mathcal{L}_1$. The semantics of $\mathcal{L}_\Box$ on an enkratic model is given by the semantics of $\mathcal{L}_2$ extended with the clause

$$\mathcal{M}, t \vDash \Box\varphi \text{ iff } \mathcal{M}, t/h' \vDash \varphi \text{ for all histories } h' \text{ with } t \in h'$$

On the axiomatic side, the new modality $\Box$ is governed by the axioms and rules below. The first, ND, expresses that the agent is under a derived ought to φ whenever φ is unavoidable to her. $\mathrm{K}_\Box$ is the K-axiom for $\Box$.

ND	$\Box\varphi \to D\varphi$	$\dfrac{\vdash \varphi}{\vdash \Box\varphi}$ $\textsc{Nec}_\Box$
$\mathrm{K}_\Box$	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	

We denote by $\Lambda_{Enkr,\Box}$ the extension of Λ_{Enkr} with ND, $\mathrm{K}_\Box$ and $\textsc{Nec}_\Box$. That is, $\Lambda_{Enkr,\Box}$ is the logic on $\mathcal{L}_\Box$ defined by all propositional tautologies together with the axioms K_G, 4, L, D_G, K_D, D_D, GO, GD, $\textsc{Max}$, ND, $\mathrm{K}_\Box$ and the rules $\textsc{Int}_O$, $\textsc{Nec}_D$, $\textsc{Nec}_G$ and $\textsc{Nec}_\Box$. See Table 2.1 for an overview. As we will see below, ND covers the full interaction between $\Box$ and the other operators.

Axioms of Λ_{temp} over $\mathcal{L}_0$		
4	$G\varphi \to GG\varphi$	$\dfrac{\vdash \varphi}{\vdash G\varphi}$ $\textsc{Nec}_G$
D_G	$\neg G\bot$	
K_G	$G(\varphi \to \psi) \to (G\varphi \to G\psi)$	
L	$F\varphi \wedge F\psi \to (F(\varphi \wedge \psi) \vee F(\varphi \wedge F\psi) \vee F(\psi \wedge F\varphi))$	
Axioms of Λ_{Enkr} over $\mathcal{L}_2$		
K_D	$D(\varphi \to \psi) \to (D\varphi \to D\psi)$	$\dfrac{\vdash \varphi}{\vdash D\varphi}$ $\textsc{Nec}_D$
D_D	$\neg D\bot$	
GD	$Goal\varphi \to D\varphi$	
GO	$Goal\varphi \to O\varphi$	$\dfrac{\vdash \varphi \leftrightarrow \psi}{\vdash O\varphi \leftrightarrow O\psi}$ $\textsc{Int}_O$
$\textsc{Max}$	$O\varphi \wedge \neg Goal\varphi \to D\neg\varphi$	
Λ_{temp} for $\mathcal{L}_1$ formulas inside $O, Goal, D$		
Additional Axioms for $\Lambda_{Enkr,\Box}$ over $\mathcal{L}_\Box$		
ND	$\Box\varphi \to D\varphi$	$\dfrac{\vdash \varphi}{\vdash \Box\varphi}$ $\textsc{Nec}_\Box$
$\mathrm{K}_\Box$	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	

Table 2.1: Axioms and rules of Λ_{temp}, Λ_{Enkr} and $\Lambda_{Enkr,\Box}$

Having the expressive resources of $\Lambda_{Enkr,\Box}$, we are finally in a position to show that goals satisfy STRONG CONSISTENCY, as desired.

Lemma 2.18. *Let $\Lambda \subseteq \mathcal{L}_2$ be consistent, and let $S \subseteq \{\varphi \in \mathcal{L}_1 \mid Goal\varphi \in \Lambda\}$ be finite. Then $\{Goal\varphi \in \Lambda\} \vdash_{\Lambda_{Enkr,\Box}} \Diamond \bigwedge_{\varphi \in S} \varphi$.*

Proof. Assume $\bigwedge_{\varphi \in S} Goal\varphi$. By iterated application of GD, we can derive $\bigwedge_{Goal\varphi \in S} D\varphi$. By K_D and NEC_D this implies $D \bigwedge_{Goal\varphi \in S} \varphi$ By K_D and D_D this implies $\neg D \neg \bigwedge_{Goal\varphi \in S} \varphi$. The counterpositive of KD then allows us to derive $\Diamond \bigwedge_{Goal\varphi \in S} \varphi$.

□

An immediate consequence is:

$$\vdash Goal\varphi \to \Diamond\varphi$$

We turn now to a characterization result for $\Lambda_{Enkr,\Box}$:

Theorem 2.19. *Assume* At *is infinite. Then the logic* $\Lambda_{Enkr,\Box}$ *is sound and weakly complete with respect to the class of enkratic models.*

The proof can be found in the appendix. Note that a strengthening of this result is not valid. Unlike Λ_{Enkr}, the extended logic $\Lambda_{Enkr,\Box}$ is only weakly complete with respect to the class of enkratic trees, at least if At is infinite.[18]

2.7.2 Back to Challenge I: From Inconsistent Oughts to Consistent Goals

This is the right moment to return to the two challenges posed in Section 2.3. We consider them in turn. The first challenge concerned the potential tension between ENKRASIA and the two principles of INTERNAL and STRONG CONSISTENCY of goals in plans. Within the current framework, the tension was solved by weakening ENKRASIA. INTERNAL and STRONG CONSISTENCY of goals are logically valid principles, while ENKRASIA is only valid *within bounds*. INTERNAL and STRONG CONSISTENCY are, in fact, the only bounds to ENKRASIA's validity. While basic oughts are possibly inconsistent, and hence not *all* basic oughts can correspond to goals, the agent's set of goals is guaranteed to be a maximally consistent subset of her basic oughts. That is, the agent validates as many instances of ENKRASIA as is possible without violating INTERNAL and STRONG CONSISTENCY.

[18] To see this take some $\varphi \in \mathcal{L}_1$ that is neither a tautology nor a contradiction. Then the set $\{\Diamond\varphi \wedge D\neg\varphi\} \cup \{\neg O\psi \mid \psi \in \mathcal{L}_1\}$ is $\Lambda_{Enkr,\Box}$ consistent. In fact, every finite subset thereoff is realizable in a enkratic -model. However, for a enkratic -model $\mathcal{M}$ to satisfy $\{\neg O\psi \mid \psi \in \mathcal{L}_1\}$ we need that $N_O = \emptyset$. This, however, implies that the admissible subtree is all of $\mathcal{T}$, which yields that $\mathcal{M}, t_0 \models D\neg\varphi$ iff $\mathcal{M}, t_0 \not\models \Diamond\varphi$, i.e., it is impossible that $\mathcal{M}, t_0 \not\models \Diamond\varphi \wedge D\neg\varphi$.

Note that the bounds imposed on ENKRASIA do not, as such, fully determine for *which* basic oughts the principle is valid. There might be more than one maximally consistent subset of the agent's basic oughts, hence additional choices are necessary. To this end, the framework incorporates a selection mechanism, fueled by the agent's priority ordering $\succ_O$. This aspect is, however, less central for the resulting logic. Any alternative selection mechanism would validate the same logical principles.

To illustrate the choices made, let us return to **Example 2.1**: Suppose I believe I ought to repay 10 euro to my friend Ann. I also believe I ought to go to the movies with Barbara. However, money is scarce, and I believe it is impossible for me to do both.

Put formally, the basic oughts of Example 2.1 are $O(F\mathsf{r})$ and $O(F\mathsf{m})$ (we assume, for the sake of illustration, that no other basic oughts are in play). If ENKRASIA were applied unrestrictedly, it would follow that $Goal(F\mathsf{r})$ *and* $Goal(F\mathsf{m})$ which, given that $\neg\Diamond(F\mathsf{r} \wedge F\mathsf{m})$, would violate STRONG CONSISTENCY. Hence, ENKRASIA can only be applied to a maximally consistent subset of those basic oughts, i.e., either to $\{OF\mathsf{r}\}$ or to $\{OF\mathsf{m}\}$. Which one of the two depends on the lexicographic order induced by $\succ_O$. Suppose I believe that settling my debt with Ann takes precedence over going to the movies with Barbara, i.e., $F\mathsf{r} \succ_O F\mathsf{m}$. It follows that ENKRASIA applies only to $\{OF\mathsf{r}\}$. The only goal derived is $Goal(F\mathsf{r})$, and since $\Diamond(F\mathsf{r})$ holds, no violation of INTERNAL or STRONG CONSISTENCY occurs.

2.7.3 Back to Challenge II: Basic and Derived Oughts

The second challenge asked to distinguish valid from invalid logical inferences about oughts. Our focus was specifically on the notorious principle of deontic closure. Even if one accepts that deontic closure is not generally valid, we have argued that an outright rejection of the principle is a too strong, and ultimately unsatisfying, solution. The challenge, hence, is to provide a deontic logic that is thick enough to license those instances of deontic closure that are unproblematic.

A central step towards meeting this challenge was the distinction between two types of oughts: basic and derived. Building on this distinction, we can illustrate the main characteristics of those deontic inferences that are valid in our logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$. Let us begin by limiting the possible conclusions derivable from valid deontic inferences. Leaving aside axiom GO and INT$_O$, the logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$ can only produce derived oughts as conclusions. In valid instances of deontic closure, hence, the ought inferred as a conclusion is a derived ought D in our sense.

More positively, valid deontic inferences are of the following kinds. First, the axiom GD allows us to infer a derived ought $D\varphi$ from a corresponding basic ought, provided that $Goal\varphi$ also holds. This reflects the idea that derived oughts are necessary conditions for the fulfillment of goals. Second, since the derived ought operator D is a normal modality, classical reasoning within the scope of derived oughts is a valid mode of inference. Thus, new derived oughts can be derived by standard modal reasoning from old ones. Finally, within the extended logic $\Lambda_{Enkr,\Box}$, axiom ND can be used to infer derived oughts also from global facts about the space of available options.

To illustrate the strength of this approach, let us recall the main lines of **Example 2.2**: Suppose that I ought to repay Ann 10 euro. Moreover, I believe that unless I refrain from going to the movies it is impossible to repay Ann. So, I conclude, I ought not go to the movies.

We have called the inference in the above example *practical inference*, and argued that valid practical inferences move from basic oughts to derived oughts. In the specific case of Example 2.2, practical inference moves from $O(F\mathsf{p})$, indicating the basic ought to repay Ann (i.e., once), to $D(G\neg\mathsf{m})$, indicating the derived ought to refraining (i.e., always) from going to the movies. It is a characteristic of our approach that derived oughts indicate the necessary conditions for the fulfillment of the agent's goals. To derive that $D(G\neg\mathsf{m})$ we therefore need to require that repaying Ann is in fact a goal in the agent's plan. With this in place, we can apply the inference rules described in Section 2.7 to derive the desired conclusion:

(i)	$O(F\mathsf{r})$	(P1)
(ii)	$\Box(F\mathsf{r} \to G\neg\mathsf{c})$	(P2)
(iii)	$Goal(F\mathsf{r})$	(P3)
(iv)	$Goal(F\mathsf{r}) \to D(F\mathsf{r})$	Axiom GD
(v)	$D(F\mathsf{r})$	From (iii), (iv) and Modus Ponens
(vi)	$D(F\mathsf{r} \to G\neg\mathsf{c})$	From (ii) and ND
(vii)	$D(G\neg\mathsf{c})$	From (v)–(vi) and K_D

Let us conclude with some observations about Ross' Paradox. The distinction between basic and derived oughts allows us to disentangle different readings of Ross' Paradox. Some of these are problematic, others in fact are not. Suppose we start from a basic ought to mail the letter. From such a premise, the logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$ allow us to infer at best a *derived* ought to mail or burn the letter.[19] Such an inference is not paradoxical. Being a derived ought, mailing or burning the letter is *not* an ought to which ENKRASIA may apply: It cannot become a goal in its own right. Such a derived ought merely describes a necessary condition for the fulfillment of the goal of mailing the letter, not a sufficient one. In fact, burning the letter is not an admissible option (i.e., states of affairs in which the letter is burnt lie outside the admissible subtree). Hence, also a derived ought *not* to burn the letter can be inferred.

2.8 Conclusion

We can now take a preliminary stock. The focus of our project is (an interpretation of) the Enkratic principle of rationality, according to which rationality requires that if an agent sincerely and with conviction believes she ought to X, then X-ing is a goal in her plan. We have developed a framework aimed at investigating the logical structure of such a principle and its implications for deontic logic.

Main ingredients of our framework are basic oughts, goals and derived oughts. Basic oughts are the inputs of the agent's deliberation. They are the oughts to which ENKRASIA may apply, that is, those oughts that potentially correspond to goals. We say "potentially" because there is a tension between, on the one hand, allowing for (jointly) inconsistent basic oughts and, on the other, requiring INTERNAL and STRONG CONSISTENCY for goals. There are several ways in which such a tension can be solved. We pursued a tolerant approach: even if not *all* basic oughts can correspond to goals, the agent's set of goals is guaranteed to be a maximally consistent subset of her basic oughts. ENKRASIA is therefore a principle of bounded validity — whose only bounds are INTERNAL and STRONG CONSISTENCY.

From goals, derived oughts follow: these indicate the necessary (but possibly not sufficient) conditions for the fulfillment of all the agent's goals. Derived oughts are conceptually distinct for basic oughts, and such conceptual distinction emerges especially with respect to ENKRASIA: ENKRASIA applies to basic oughts but not to derived oughts. We showed that such conceptual distinction between oughts plays a significant logical role, specifically in delineating which instances of deontic closure are valid. Within the framework, deontic closure is valid whenever the ought inferred is a derived ought, but not if it is a basic

[19] And even that only if also the goal to mail the letter is present.

ought. This allowed us to validate inferences (such as practical inference) that intuitively strike as plausible, but also to distinguish between problematic and unproblematic readings of Ross' Paradox (and to validate the latter).

The next chapter investigates the *dynamic* relationship between basic oughts, goals and derived oughts.

2.9 Appendix: Proofs

This section is devoted to prove the completeness of the logic Λ_{Enkr} and of its extension $\Lambda_{Enkr,\Box}$.[20]

The completeness proof for Λ_{Enkr} follows the following strategy:

- We want to prove that every consistent set of $\mathcal{L}_2$ formulas has a model
- To do so, we follow the standard method of extending such consistent set to a maximally consistent set (via Lindenbaum Lemma) and then construct an appropriate canonical model.
- The way the canonical model is constructed is as follows:
  - First note that such maximally consistent set of $\mathcal{L}_2$ formulas, call it Λ, contains formulas like atoms, boolean combination thereof, formulas like $O\varphi$, $Goal\varphi$, $D\varphi$ (with φ future tensed) and boolean combinations thereof.
  - We build a canonical tree model such that (i) all literals (i.e., atoms and negated atoms) in Λ are true in the root of the tree, (ii) if $O\varphi$ and $Goal\varphi$ are in Λ, then all histories of the tree make φ true, and (iii) $O\varphi$ and $\neg Goal\varphi$ are in Λ, then no history of the tree make φ true. As a result, if $D\varphi$ is in Λ then all histories make φ true; if $\neg D\varphi$ is in Λ then no history makes φ true.
  - In order to build such canonical tree model, we show that, given Λ, we can construct a linear tree (i.e., a tree given by just one history) canonical model such that (i) it makes true all the literals in Λ and all the formulas φ for which $D\varphi$ is in Λ, and (ii) it makes false some ψ for which $\neg D\psi$ is inΛ.
  - Then we collect all those linear trees (for all ψ for which $\neg D\psi$ is in Λ), and build a branching tree by making all the roots identical.
- We show that the resulting canonical model is an enkratic model which satisfies Λ, and is therefore the model we were looking for.

Before we can prove Theorem 2.17, we need the following auxiliary lemma:

[20] The completeness proofs for Λ_{Enkr} and $\Lambda_{Enkr,\Box}$ in Klein and Marra (2019) are by Dominik Klein. Here I present annotated versions of those proofs. All remaining errors are mine.

Lemma 2.20. *Assume* $\Lambda \subseteq \mathcal{L}_2$ *is maximally* Λ_{Enkr} *consistent. Let* $\Lambda^D = \{\varphi \in \mathcal{L}_1 \mid D\varphi \in \Lambda\}$, *let* $\Lambda^{\neg D} = \{\varphi \in \mathcal{L}_1 \mid \neg D\varphi \in \Lambda\}$ *and let* $\Lambda^{lit} \subseteq \Lambda$ *be the set of all literals, i.e., atoms and negated atoms, occuring in* Λ. *Then for every* $\psi \in \Lambda^{\neg D}$, *there is a linear*[21] *tree* $\mathcal{H}_{\neg\psi} = \langle h_{\neg\psi}, \prec_{\mathcal{H}}\rangle$ *with root* $t_{\neg\psi}$ *such that* $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \neg\psi$ *and* $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \varphi$ *for every* $\varphi \in \Lambda^D \cup \Lambda^{lit}$.

Proof. Let $\chi \in \Lambda^{\neg D}$, we will construct the desired linear tree $\mathcal{H}_{\neg\chi} = \langle h_{\neg\chi}, \prec_{\mathcal{H}}\rangle$. By axioms K_D and D_D, the set $\Lambda^D \cup \{\neg\chi\}$ is Λ_{temp} consistent. Since all $\varphi \in \Lambda^D \cup \{\neg\chi\}$ are future looking, i.e., every atom is in the scope of a modal operator, also $\Lambda^D \cup \{\neg\chi\} \cup \Lambda^{lit}$ is Λ_{temp} consistent. We can hence expand it to a maximally Λ_{temp} consistent subset $\Gamma \subseteq \mathcal{L}_0$. The rest of the proof proceeds by a classic buldozing argument. For the sake of completeness, we hint at the details.

Let $\mathcal{M} = \langle W, R\rangle$ be the Λ_{temp} canonical model over $\mathcal{L}_0$. That is, W is the set of all maximally Λ_{temp} consistent subsets of $\mathcal{L}_0$ and $\Theta R\Sigma$ iff $F\varphi \in \Theta$ for all $\varphi \in \mathcal{L}_0$ with $\varphi \in \Sigma$. Let $H = \{\Sigma \in \mathcal{M} \mid \Gamma R\Sigma\}$. We show that R is transitive and complete on H. Transitivity follows from the 4 axiom. For completeness let $\Theta \neq \Sigma \in H$, wlog $\Theta, \Sigma \neq \Gamma$. We have to show $\Theta R\Sigma$ or $\Sigma R\Theta$.

Pick enumerations $\varphi_0^\Theta, \varphi_1^\Theta \ldots$ of Θ and $\varphi_0^\Sigma, \varphi_1^\Sigma \ldots$ of Σ. For $i \in \omega$ let $\psi_i^* = \bigwedge_{j=1}^i \varphi_j^*$ for $* \in \{\Theta, \Sigma\}$. Note that, since $\Theta \neq \Sigma$, there are $j, k \geq 0$ such that $\varphi_j^\Theta = \neg\varphi_k^\Sigma$. Letting $i_0 = max(j,k)$ we have that $\vdash_{\Lambda_{temp}} \neg(\psi_i^\Theta \wedge \psi_i^\Sigma)$ for all $i > i_0$. As Θ, Σ are maximally consistent we get $\psi_i^\Theta \in \Theta, \psi_i^\Sigma \in \Sigma$ for all $i \in \omega$.

By construction, we have $\Gamma R\Theta$ and $\Gamma R\Sigma$. The Truth Lemma then implies that $F\psi_i^* \in \Gamma$ for $* \in \{\Theta, \Sigma\}$ and all $i \in \omega$. Hence, we have for all i that $F\psi_i^\Theta \wedge F\psi_i^\Sigma \in \Gamma$. By L, this implies that $F(\psi_i^\Theta \wedge \psi_i^\Sigma) \vee F(\psi_i^\Theta \wedge F\psi_i^\Sigma) \vee F(\psi_i^\Sigma \wedge F\psi_i^\Theta) \in \Gamma$. In particular, Γ contains either $F(\psi_i^\Theta \wedge \psi_i^\Sigma)$ for infinitely many i, $F(\psi_i^\Theta \wedge F\psi_i^\Sigma)$ for infinitely many indices i, or $F(\psi_i^\Sigma \wedge \psi_i^\Theta)$ for infinitely many i. Since $\vdash_{\Lambda_{temp}} \neg(\psi_i^\Theta \wedge F\psi_i^\Theta)$ for all but finitely many i, the first case is impossible. We treat the case where Γ contains $F(\psi_i^\Theta \wedge F\psi_i^\Sigma)$ for infinitely many i, the other case being similar. We will show that $F\psi_i^\Sigma \in \Theta$ for all $i \in \omega$. By K_G and the construction of the ψ_i^Σ, this entails that $F\varphi \in \Theta$ for all $\varphi \in \Sigma$, which, together with the definition of R, implies that $\Theta R\Sigma$.

To see that $F\psi_i^\Sigma \in \Theta$ for all $i \in \omega$, assume for a contradiction that this is false, i.e., that there is some i_m with $F\psi_{i_m}^\Sigma \notin \Theta$. By maximal consistency, this implies that $\neg F\psi_{i_m}^\Sigma \in \Theta$. Hence, there is some $\psi_j^\Theta \in \Theta$ with $\vdash_{\Lambda_{temp}} \neg(\psi_j^\Theta \wedge F\psi_{i_m}^\Sigma)$. By

[21] I.e., a tree where $\prec_{\mathcal{H}}$ is a linear order.

construction of the ψ_j, this implies $\vdash_{\Lambda_{temp}} \neg(\psi_{j'}^{\Theta} \wedge F\psi_{j'}^{\Sigma})$ for all $j' > \max(i_m, j)$. In particular, since Γ contains $F(\psi_i^{\Theta} \wedge F\psi_i^{\Sigma})$ for infinitely many i, there is some $j_0 > \max(i_m, j)$ with $F(\psi_{j_0}^{\Theta} \wedge F\psi_{j_0}^{\Sigma}) \in \Gamma$ but $\vdash_{\Lambda_{temp}} \neg(\psi_{j_0}^{\Theta} \wedge F\psi_{j_0}^{\Sigma})$. This, in connection with D_G, contradicts the consistency of Γ. Hence the assumption was false and we obtain that $F\psi_i^{\Sigma} \in \Theta$ for all $i \in \omega$.

To finish the proof, we need to ensure that R is a linear order on H. This is, in general, not true. However, by a classic bulldozing argument (cf. Venema, 2001), we can transform H into a linearly ordered tree $\mathcal{H} = \langle h, \prec_{\mathcal{H}} \rangle$, with $\prec_{\mathcal{H}}$ minimal element Γ such that $\mathcal{H}, \Gamma \vDash \varphi \Leftrightarrow \varphi \in \Gamma$. Renaming Γ to $t_{\neg\chi}$, h to $h_{\neg\chi}$, and $\mathcal{H}$ to $\mathcal{H}_{\neg\chi}$ finishes the proof. $\square$

Now, we can finally show Theorem 2.17. The completeness direction proceeds by constructing a special tree model $\mathcal{T}$, with the property that for each $([\![\varphi]\!]^{t_0} \, \varphi) \in N_O$ either $[\![\varphi]\!]^{t_0} = \mathcal{T}$ or $[\![\varphi]\!]^{t_0} = \emptyset$. Consequently, it will hold that $\mathcal{T}_{Goal} = \mathcal{T}$. The main task of the construction, hence, is to ensure that the histories of $\mathcal{T}$ are such that $\mathcal{T} \subseteq [\![\varphi]\!]^{t_0} \Leftrightarrow D\varphi \in \Lambda$. The corresponding construction bears some resemblance to the completeness proof for ATL (Goranko and van Drimmelen, 2006).

Proof of Theorem 2.17: For completeness, we show that every maximally Λ_{Enkr} consistent subset Λ of $\mathcal{L}_2$ is satisfiable in an enkratic model. Let $\Lambda^D = \{\varphi \in \mathcal{L}_1 \mid D\varphi \in \Lambda\}$, let $\Lambda^{\neg D} = \{\varphi \in \mathcal{L}_1 \mid \neg D\varphi \in \Lambda\}$ and let $\Lambda^{lit} \subseteq \Lambda$ be the set of all literals, i.e., atoms and negated atoms, occuring in Λ. Using Lemma 2.20, we pick linear trees $\mathcal{H}_{\neg\psi} = \langle h_{\neg\psi}, \prec_{\mathcal{H}} \rangle$ with root $t_{\neg\psi}$ for each $\psi \in \Lambda^{\neg D}$ as above. Note that all $t_{\neg\psi}$ share the same atomic valuation, as this is completely determined by Λ^{lit}. Moreover, note that $\neg D\bot \in \Lambda$ by D_D. Hence $\bot \in \Lambda^{\neg D}$ and thus the set of linear trees picked is non-empty.

We have to construct an enkratic model $\mathcal{M} = \langle \mathcal{T}, t_0, v, N_o, \succ_O \rangle$. As tree $\mathcal{T}$ we take the union of the $\mathcal{H}_{\neg\psi}$ where we identify all $t_{\neg\psi}$. Formally, for a linear tree $\mathcal{H} = \langle h, \prec_{\mathcal{H}} \rangle$, let $T_{\mathcal{H}}^{>}$ be the set of all moments but the first of $\mathcal{H}$. Let:

$$T = \{t_0\} \cup \bigcup_{\psi \in B} T_{\mathcal{H}_{\neg\psi}}^{>}$$

and $\prec_{\mathcal{T}}$ the inherited tree-order, making t_0 the root. Finally, the valuation v

is defined by

$$t \in v(p) \quad \text{iff} \quad \begin{cases} t = t_0 \text{ and } p \in \Lambda \\ \text{or } t \in T^{>}_{\mathcal{H}_{\neg\psi}} \text{ and } t \in v_{\mathcal{H}_{\neg\psi}}(p). \end{cases}$$

Finally, we define N_O by $([\![\varphi]\!]^{t_0}, \varphi) \in N_O$ iff $O\varphi \in \Lambda$. Moreover, we pick an arbitrary well-founded ordering $\succ_O$ on $\{\varphi \mid ([\![\varphi]\!]^{t_0}, \varphi) \in N_O\}$.

For the completeness argument, we begin with some observations about the root t_0 of this tree. First we note that, for $([\![\varphi]\!]^{t_0}, \varphi) \in N_O(t_0)$, we have $[\![\varphi]\!]^{t_0} = T$ if $Goal\varphi \in \Lambda$, and $[\![\varphi]\!]^{t_0} = \emptyset$ else. In the former case, we have $D\varphi \in \Lambda$ by GD. By construction, this implies that every $\mathcal{H}_{\neg\psi}$ and hence every branch h/t_0 of $\mathcal{T}$ satisfies $h/t_0 \vDash \varphi$, i.e., $[\![\varphi]\!]^{t_0} = T$, which is what had to be shown. In the other case, $Goal\varphi \notin \Lambda$, we have, by MAX, that $D\neg\varphi \in \Lambda$. Again by construction, every branch h/t_0 of $\mathcal{T}$ satisfies $h/t_0 \vDash \neg\varphi$, i.e., $[\![\varphi]\!]^{t_0} = \emptyset$. The last two observations imply that:

$$N_G = \{[\![\varphi]\!]^{t_0} \mid O\varphi \in \Lambda, Goal\varphi \in \Lambda\}. \tag{2.1}$$

In particular, by GD, $N_G(t_0) = \{\mathcal{T}\}$ if there is some $Goal\psi \in \Lambda$ and $N_G(t_0) = \emptyset$ else. In either case we have $\Cap_{X \in N_G(t_0)} X = \mathcal{T}$.

Now we can show that our model is as desired, i.e., $\mathcal{M}, t_0 \vDash \varphi$ iff $\varphi \in \Lambda$. The argument is an induction over the complexity of φ. As induction base, we show the claim for φ an atom or of the form $O\psi, Goal\psi$ or $D\psi$ for some $\psi \in \mathcal{L}_1$. In the induction step we then show that if the claim holds for $\varphi_1, \varphi_2 \in \mathcal{L}_2$ then also for $\neg\varphi_1$ and $\varphi_1 \wedge \varphi_2$. This induction step is trivial. We only need to show the claim for the induction base.

If φ is atomic, the claim holds by definition of the valuation on t_0. If φ is $O\psi$ for some $\psi \in \mathcal{L}_1$, this follows immediately from the construction of N_O. If φ is $Goal\psi$ for $\psi \in \mathcal{L}_1$ the claim follows immediately from Equation 2.1.

The only non-trivial case is when φ is of the form $D\psi$ for $\psi \in \mathcal{L}_1$. For the left to right direction, assume $\mathcal{M}, t_0 \vDash D\psi$. We have to show $D\psi \in \Lambda$. First, note that $\mathcal{M}, t_0 \vDash D\psi$ implies that $\bigcap_{X \in N_G} X \subseteq [\![\psi]\!]^{t_0}$. Since for each $X \in N_G$ holds that $X = \{\mathcal{T}\}$ or $X = \emptyset$, this implies that $\mathcal{T} \subseteq [\![\psi]\!]^{t_0}$. Assume for a contradiction that $D\psi \notin \Lambda$. By maximality, this implies $\neg D\psi \in \Lambda$. By construction, there is a branch $h_{\neg\psi}$ of $\mathcal{T}$ with $\mathcal{T}, t_0/h_{\neg\psi} \vDash \neg\psi$. In particular, $\mathcal{T} \not\subseteq [\![\psi]\!]^{t_0}$, which is a contradiction. Hence the assumption was false and $D\psi \in \Lambda$. For the right to left direction, assume $D\psi \in \Lambda$. Recall that, by construction, every branch

h of $\mathcal{M}$ satisfies $\mathcal{M}, t_0/h \vDash \psi$. In particular, $\bigcap_{X \in N_G} X \subseteq [\![\psi]\!]^{t_0}$, which implies that $\mathcal{M}, t_0 \vDash D\psi$. □

Proof of Theorem 2.19: The proof borrows heavily from the proof of Theorem 2.17. We only highlight the relevant differences. For weak completeness, we have to show that whenever there is some $\tilde{\varphi} \in \mathcal{L}_\Box$ with $\nvdash_{\neg\Lambda_{Enkr,\Box}} \neg\tilde{\varphi}$, there is some enkratic model $\mathcal{M}, w$ with $\mathcal{M}, w \vDash \tilde{\varphi}$. Let such $\tilde{\varphi}$ be given. Without loss of generality, we can assume that $\tilde{\varphi}$ is in disjunctive normal form, i.e,

$$\tilde{\varphi} = \bigvee_{i=1}^{n} \bigwedge_{j=1}^{k_i} \chi_{i,j}$$

where each χ_i either is an atom, a negated atom, or of the form $X\varphi$ or $\neg X\varphi$ for $X \in \{O, Goal, D, \Box\}$ and $\varphi \in \mathcal{L}_1$. By MAX, we have $\vdash_{\neg\Lambda_{Enkr,\Box}} O\varphi \leftrightarrow (O\varphi \wedge Goal\varphi) \vee (O\varphi \wedge D\neg\varphi)$. We can hence assume, without loss of generality, that every disjunct $\bigwedge_{j=1}^{k_i} \chi_{i,j}$ of φ has the property that, for each $\chi_{i,j}$ of the form $O\psi$, either $Goal\psi$ or $D\neg\psi$ appears as some of the $\chi_{i,j'}$ with $j' \leq k_i$. Likewise, as $\vdash_{\neg\Lambda_{Enkr,\Box}} \neg\Box\varphi \leftrightarrow (\neg\Box\varphi \wedge \neg D\varphi) \vee (\neg\Box\varphi \wedge D\varphi)$, we can assume, that for each $\chi_{i,j}$ of the form $\neg\Box\psi$, either $\neg D\psi$ or $D\psi$ appears as some of the $\chi_{i,j'}$ with $j' \leq k_i$. Moreover, by GO, we can assume that, for each $\chi_{i,j}$ of the form $Goal\psi$, also $O\psi$ appears as some of the $\chi_{i,j'}$ with $j' \leq k_i$ Finally, by D_D, we can assume that, for each $i \leq n$, $\neg D\bot$ appears as some $\chi_{i,j'}$ with $j' \leq k_i$. To show that $\tilde{\varphi} = \bigvee_{i=1}^{n} \bigwedge_{j=j}^{k_i} \chi_i$ is satisfiable in an enkratic model, it suffices to show that one of its disjuncts is satisfiable in an enkratic model. Hence, it suffices to show the claim for $i = 1$, that is for $\tilde{\varphi}$ of the form $\bigwedge_{j=1}^{k} \chi_j$. Let such a φ be given and let $X = \{\chi_1, \dots \chi_k\}$.

In a similar fashion as in Theorem 2.17, let $\Lambda^D = \{\rho \in \mathcal{L}_1 | D\rho \in X\}$, let $\Lambda^\Box = \{\rho \in \mathcal{L}_1 | \Box\rho \in X\}$, let $\Lambda^{\neg D} = \{\rho \in \mathcal{L}_1 | \neg D\rho \in X\}$ and let $\Lambda^{\mathsf{At}} = \{\chi \in X | \chi = p \text{ or } \chi = \neg p\}$. Note that $\Lambda^{\neg D} \neq \emptyset$ as $\bot \in \Lambda^{\neg D}$. As At is infinite, we can pick an atom p_0 that does not occur in any of the formulas in X. Pick a valuation Λ^{lit} extending Λ^{At}, i.e., some maximally consistent $\Lambda^{lit} \subseteq \{p, \neg p \mid p \in \mathsf{At}\}$ with $\Lambda^{lit} \supseteq \Lambda^{\mathsf{At}}$ such that $\neg p_0 \in \Lambda^{lit}$. By a slight adaptation of Lemma 2.20, there are linear trees $\mathcal{H}_{\neg\psi} = \langle h_{\neg\psi}, \prec_{\mathcal{H}} \rangle$, with root $t_{\neg\psi}$ for each $\psi \in \Lambda^{\neg D}$, such that $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \neg\psi$ and $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \chi$ for every $\chi \in \Lambda^D \cup \Lambda^\Box \cup \Lambda^{lit}$. Moreover, let $\Lambda^{\Diamond/D} = \{\neg\rho \in \mathcal{L}_1 | \neg\Box\rho \in X \text{ and } D\rho \in X\}$ be the set of formulas that are possible without being a derived goal. By another slight adaptation of Lemma 2.20, there are linear trees $\mathcal{H}'_\rho = \langle h'_\rho, \prec'_{\mathcal{H}} \rangle$, with root t'_ρ for each $\rho \in \Lambda^{\Diamond/D}$, such that $\mathcal{H}'_\rho, t'_\rho \vDash \rho$ and $\mathcal{H}'_\rho, t'_\rho \vDash \chi$ for every $\chi \in \Lambda^\Box \cup \Lambda^{lit}$.

Before constructing the desired enkratic model, we show the following claim: There is a formula ρ_0 such that $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \rho_0$ for all $\psi \in \Lambda^{\neg D}$, but $\mathcal{H}'_\psi, t'_\psi \vDash \neg\rho_0$ for all $\psi \in \Lambda^{\Diamond/D}$. To see this, note that by definition of sets Λ^D and $\Lambda^{\Diamond/D}$, every member of $\Lambda^{\Diamond/D}$ is of the form $\neg\chi$ with $\chi \in \Lambda^D$. Let $\rho_0 = \bigwedge_{\neg\chi\in\Lambda^{\Diamond/D}} \chi$. We hence have that $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \rho_0$ for all $\psi \in \Lambda^{\neg D}$, but $\mathcal{H}'_\psi, t'_\psi \vDash \neg\rho_0$ for all $\psi \in \Lambda^{\Diamond/D}$, i.e., ρ_0 has the desired property.

Now, we construct a enkratic model $\mathcal{M} = \langle \mathcal{T}, t_0, v, N_O, \succ_O\rangle$ satisfying $\tilde{\varphi}$ as follows. We pick linear trees $\mathcal{H}_{\neg\psi}$ for each $\psi \in \Lambda^D$ and $\mathcal{H}'_\psi$ for all $\psi \in \Lambda^{\Diamond/D}$ as above. As p_0 does not occur in any of the formulas in X and $\neg p_0 \in \Lambda^{lit}$, we can assume wlog that p_0 is globaly false on all $\mathcal{H}_{\neg\psi}$ and $\mathcal{H}'_\psi$. With these, we define a tree $\mathcal{T}$ as in the proof of Theorem 2.17. By the previous claim, there is formula ρ_0 such that $\mathcal{H}_{\neg\psi}, t_{\neg\psi} \vDash \rho_0$ for all $\psi \in \Lambda^{\neg D}$, but $\mathcal{H}'_\psi, t'_\psi \vDash \neg\rho_0$ for all $\psi \in \Lambda^{\Diamond/D}$. Since p_0 is false everywhere, $\rho = \rho_0 \vee Gp_0$ has the same property. Moreover, since p_0 does not occur in X, we also have that $O\rho, \neg O\rho, Goal\rho, \neg Goal\rho$ are all not contained in X. Define the neighborhood N_O by $([\![\varphi]\!]^{t_0}, \varphi) \in N_O$ iff $O\varphi \in X$ or $\varphi = \rho$. Finally, as priority relation $\succ_O$, we pick any well-founded ordering on $\{\varphi \mid O\varphi \in X\} \cup \{\rho\}$ that has ρ as maximal element.

We can now show that $\mathcal{M}, t_0 \vDash \tilde{\varphi}$. Since $\tilde{\varphi} = \bigwedge_{\psi\in X} \psi$, it suffices to show that $\mathcal{M}, t_0 \vDash \psi$ for all $\psi \in X$. If ψ is an atom or negated atom, this follows immediately from the construction. The same holds if ψ is of the form $O\varphi, \neg O\varphi$ or $\Box\varphi$. If ψ is of the form $\neg\Box\varphi$ we have by our assumption that either $D\varphi \in X$ or $\neg D\varphi \in X$. In the first case $\neg\varphi \in \Lambda^{\Diamond/D}$ and $\mathcal{H}'_{\neg\varphi}, t'_{\neg\varphi} \vDash \neg\varphi$, and hence $\mathcal{M}, t_0 \nvDash \Box\varphi$. In the second case, $\varphi \in \Lambda^{\neg D}$ and $\mathcal{H}_{\neg\varphi}, t_{\neg\varphi} \vDash \neg\varphi$ witnessing again that $\mathcal{M}, t_0 \nvDash \Box\varphi$.

For the remaining cases, we define subtrees $\mathcal{S}$ and $\mathcal{S}'$ of $\mathcal{T}$ by $\mathcal{S} = \bigcup\{\mathcal{H}_{\neg\psi} \mid \psi \in \Lambda^{\neg D}\}$ and $\mathcal{S}' = \bigcup\{\mathcal{H}'_\psi \mid \psi \in \Lambda^{\Diamond/D}\}$. Since for each $O\psi \in X$ also $Goal\psi \in X$ or $D\neg\psi \in X$, we can use the same argument as in the previous theorem to show that for $O\psi \in X$, we have that $S \subseteq [\![\psi]\!]^{t_0}$ if $Goal\psi \in X$, and $[\![\psi]\!]^{t_0} \subseteq S'$ if $D\neg\psi \in X$. Since $[\![\rho]\!]^{t_0} = S$, and ρ is the $\succ_O$ maximal element of $\{\varphi \mid O\varphi \in X\} \cup \{\rho\}$, we get that $[\![\varphi]\!]^{t_0}$ with $O\varphi \in X$ can only be in N_G if $[\![\varphi]\!]^{t_0} \nsubseteq S'$, i.e., if $Goal\varphi \in X$. Moreover, since $S \subseteq [\![\varphi]\!]^{t_0}$ whenever $Goal\varphi \in X$, we get that $N_G = \{[\![\varphi]\!]^{t_0} \mid O\varphi, Goal\varphi \in X\}$. By our assumption that $O\varphi \in X$ whenever $Goal\varphi \in X$, this simplifies to $N_G = \{[\![\varphi]\!]^{t_0} \mid Goal\varphi \in X\}$ and hence $\Cap_{[\![\varphi]\!]^{t_0}\in N_G} [\![\varphi]\!]^{t_0} = \mathcal{S}$. From there, the same argument as in the previous theorem implies that $\mathcal{M}, t_0 \vDash \psi$ if $\psi \in X$ is of the form $Goal\varphi$ or $D\varphi$. If ψ is of the form $\neg D\varphi$, note that there is, by construction, a branch $h_{\neg\varphi}$ with $\mathcal{M}, t_0/h_{\neg\varphi} \vDash \neg\varphi$. Since $h_{\neg\varphi} \subseteq \mathcal{S}$ and $\Cap_{[\![\varphi]\!]^{t_0}\in N_G} [\![\varphi]\!]^{t_0} = \mathcal{S}$, this implies

$M, t_0 \models \neg D\varphi$. Finally, if ψ is of the form $\neg Goal\varphi$, we need to distinguish whether $O\varphi \in X$ or not. If not, we have, by construction, that $([\![\varphi]\!]^{t_0}, \varphi) \notin N_O$, which immediately implies that $M, t_0 \models \neg Goal\varphi$. For the case that $O\varphi \in X$, we have by construction that also $D\neg\varphi \in X$. For this case, we have shown above that $[\![\varphi]\!]^{t_0} \subseteq S'$, which implies $[\![\varphi]\!]^{t_0} \notin N_G$. Again, we get $M, t_0 \models \neg Goal\varphi$ as desired. $\square$

Chapter 3

From Oughts to Goals. Part II

This chapter continues the investigation of the Enkratic principle of rationality, by exploring the dynamic relationship between basic oughts, goals and derived oughts. We focus on practical dynamics, i.e., on changes in the possible courses of events the agent believes are open to her. By means of product updates, we represent operations that can add or remove possible courses of events. We investigate how and under which conditions these operations affect the notions at play in Enkrasia, specifically goals and derived oughts.

3.1 Introduction

Our contribution to the logic and the implications of ENKRASIA is grounded on the conceptual distinction between basic oughts, goals and derived oughts. This chapter focuses on the *dynamic* relationship between those notions.

We can distinguish at least three kinds of dynamics relevant to our investigation of ENKRASIA: (i) *temporal dynamics*, concerned with tracking the agent's basis oughts, goals and derived oughts through the progressing of time; (ii) *normative dynamics*, concerned with changes of the basic oughts the agent believes in; and finally (iii) *practical dynamics*, concerned with variations of the possible courses of events the agent believes are open to her. All three kinds of dynamics possibly open interesting formal and conceptual questions, but for this chapter we restrict our attention to *practical dynamics*.

The effects of practical dynamics on oughts and goals are delicate. In this chapter we explore the formal features of practical dynamics, and use resources from the logics of action models to represent the effects that removing or adding possibilities have for the agent's goals and derived oughts. Our motivation is twofold. Firstly, in practical philosophy, a certain stability under practical dynamics is commonly attributed to goals (Bratman, 1987). We aim to contribute to such literature by addressing the conditions for such stability, and by illustrating how the stability of goals can make prominent the conceptual distinction between those basic oughts corresponding to goals and derived oughts. Secondly, while there exists a well-established tradition on logical dynamics, the focus has been so far mainly on the effects on the attitude of belief.[1] This chapter aims at complementing the literature on dynamic logic by drawing attention to oughts and goals.

The chapter has the following structure. We start by introducing a challenge, which illustrates the relevance that practical dynamics has for the notions involved in ENKRASIA (Section 3.2). Then we move to a dynamic extension of our enkratic logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$. Specifically, we define the operation of product update between enkratic models and temporal update models, and show how product updates can be used to represent some illustrative examples (Section 3.3). Finally, in Section 3.3.2 we prove some general results on the effects that removing and adding possibilities have for goals and derived oughts. These general results show that the dynamic relation between basic oughts, goals and derived oughts is less trivial than one could perhaps expect. This concludes our investigation on the logic and the implications of ENKRASIA: in Section 3.4 we compare our approach to other logical frameworks in the literature, and conclude by indicating some possible future directions.

3.2 Challenge III: Dynamic Conditions

There exist certain challenges surrounding ENKRASIA. In the previous chapter, we introduced two of them: they concerned the filtering of inconsistent basic oughts and the restricted validity of the principle of deontic closure. The third challenge, finally, brings *dynamics* into the picture. Our initial reason for focusing on dynamics is conceptual, as it is especially when seen through the lenses of dynamic change that the differences between basic oughts, goals and derived oughts become most prominent. The following two observations illustrate what we have in mind.

[1] Information change in the context of plans (so-called plan management) has been, on the other hand, the focus of works in the field of theoretical artificial intelligence. See, e.g., Bratman et al. (1988), Horty and Pollack (2001), Pollack (1992), Pollack and Horty (1999).

Firstly, it is a well-known feature of goals in plans that they tend to be *stable* under various perturbations (cf. Bratman, 1987, pp. 16,67). Reflecting the fact that goals are ultimately things the agent has committed to, there is a tendency for goals in plans to resist reconsideration, and in particular not to be discarded at every slight change that might occur in the agent's information:

Example 3.1. *Suppose that giving 10 euro back to Ann is a goal in my plan. My plan involves reaching Ann's house either by car or bus, as I believe these are the only options for getting to her place. Now if I learned that my car is broken, I would not simply give up my goal to repay Ann. Normally, I would rather maintain my goal and replan to get to her house by another means, for instance, by bus.*

Let us call *practical dynamics* any changes in the agent's information about what the world looks like. This is, in fact, the kind of dynamics at work in Example 3.1. Then, the basic idea is that goals in plans are not reconsidered whenever practical dynamics occurs. Of course, goals are not irrevocable. They should, for instance, be dropped if they become inconsistent (cf. Bratman, 1987, p.16). Yet, goals are stabler than other notions with respect to practical dynamics.

This leads us to a second observation. Echoing Horty (2015), we can appeal to a sort of *stability test* to illuminate the conceptual distinction between basic oughts, at least those that correspond to goals via ENKRASIA, and derived oughts. The following example shows that derived oughts are generally less stable with respect to practical dynamics:[2]

Example 3.2. *Suppose I ought to repay 10 euro to Ann (basic ought), and I hold the corresponding goal in my plan. Moreover, suppose that I do not have a lot of money, and hence conclude that I ought not go to the movies (derived ought), although I would really like to. Now if I learned that I have additional money at home, sufficient for both repaying Ann and buying a cinema ticket, I would give up that I ought not go to the movies. While I would maintain that I ought to repay Ann, and hence that doing so is a goal in my plan, I would not maintain that I ought to make sure not to go to the movies.*

Hence, while the difference between basic oughts and derived oughts is not explicitly reflected in surface grammar, testing stability with respect to practical dynamics can help to make this conceptual distinction salient.[3]

[2] A version of Example 3.2 is discussed in Horty (2015), p.225.

[3] We thank an anonymous reviewer for drawing attention to the fact that the contrast

The above observations show the relevance of investigating the effects of practical dynamics. Unanticipated obstacles or unexpected opportunities can diversely affect various notions at play in ENKRASIA, specifically goals in plans and derived oughts. The challenge then consists in precisely characterizing how and under which conditions these change dynamically.

3.3 Practical Dynamics

This section provides a formal background to investigate the dynamics of basic oughts, goals and derived oughts. More precisely, we consider how goals and derived oughts change when the agent receives new information that impact her mental picture of the world and the options available to her. In general, this information may trigger significant replanning, as the agent's original options might not be admissible anymore or new, better options have become available. Studying these dynamic variations — we hold — provides additional insights into the relationship between basic oughts, goals and derived oughts.

As mentioned above, our focus here is exclusively on what we have called *practical dynamics*. This means that the dynamics studied here is not fueled by time progression and the according changes to the set of options available. For this exposition, we assume the agent to rest in moment t_0, i.e., she has not yet begun to put her plans into action. Even before beginning to act, the agent might receive information that changes her beliefs of available future courses of events (van Ditmarsch et al., 2008; Baltag et al., 1998). Moreover, we stipulate that this information is purely descriptive: the agent does not receive any information that leads her to adapt or discard any basic oughts. Rather, updates may only concern further available courses of events she had not yet considered, or that certain options she had considered are, in fact, not available. Crucially, leaving the set of basic oughts intact does not entail that the agent's set of goals remains unchanged. Which of the agent's basic oughts translate into goals precisely depends upon whether they are satisfiable in a given situation and, more general, which sets of basic oughts are jointly satisfiable.

In the following, we provide a dynamic account of how the agent's goals and derived oughts change when the available courses of events do. We illustrate this with the following example, which brings together our previous Examples

between basic oughts and derived oughts is not marked in ordinary language. The distinction between the two oughts is rather conceptual.

3.1 and 3.2:

Example 3.3. *Suppose that I ought to repay my friend Ann 10 euro today (basic ought), and that is a goal in my plan. I start to plan my day accordingly. I believe I can get to Ann's house only by bus or car, so it follows that I ought to take the bus or the car (derived ought). Moreover, since my money is scarce, it follows that I ought not go to the movies today (derived ought), although I would really like to. In fact, I have even promised my friend Barbara to go to the movies with her (basic ought), but repaying my debt takes precedence (priority relation). Consider the following epistemic updates.*

> *Update 1: Suppose I learn that the car has a dead battery, and I cannot fix the problem on time to reach Ann's house to give her the 10 euro. So, I conclude I ought to take the bus and replan my day accordingly.*
>
> *Update 2: Suppose that later, I learn that I can simply walk to Ann's (her place is unexpectedly quite close). Hence, now there are again two ways in which I can reach her house: by walking or by bus. I conclude it's no longer true that I ought to take the bus.*
>
> *Update 3: Finally, suppose that I find some extra money at home: It is no longer true that I do not have enough money to repay Ann and go to the movies. I can do both. Hence, it ceases to be true that I ought not go to the movies. In fact, as I had promised my friend Barbara to accompany her, it now follows that I entertain the goal of going to the movies.*

Figure 3.1 illustrates this situation: The top left corner shows the initial tree, the top right corner the tree after learning that the car's battery is dead. The results of the following two updates are depicted in the bottom row.

3.3.1 Product Enkratic Models

We turn now to a formal representation of practical dynamics for the logics Λ_{Enkr} and $\Lambda_{Enkr,\Box}$ introduced in the previous chapter. In line with much work in computer science (e.g. Ciuni and Zanardo, 2010) we assume trees in enkratic models to be discrete. To be precise, we assume that every history is isomorphic to the natural numbers, i.e., it can be written as $h = t_0 \prec_{\mathcal{T}} t_1 \prec_{\mathcal{T}} t_2 \prec_{\mathcal{T}} \ldots$.

For a tree $\mathcal{T} = \langle T, \prec_{\mathcal{T}}\rangle$ let $\prec_{im}$ be the **immediate predecessor relation**, i.e. $x \prec_{im} y$ iff $x \prec_{\mathcal{T}} y$ and there is no $z \in \mathcal{T}$ with $x \prec_{\mathcal{T}} z \prec_{\mathcal{T}} y$. Note that

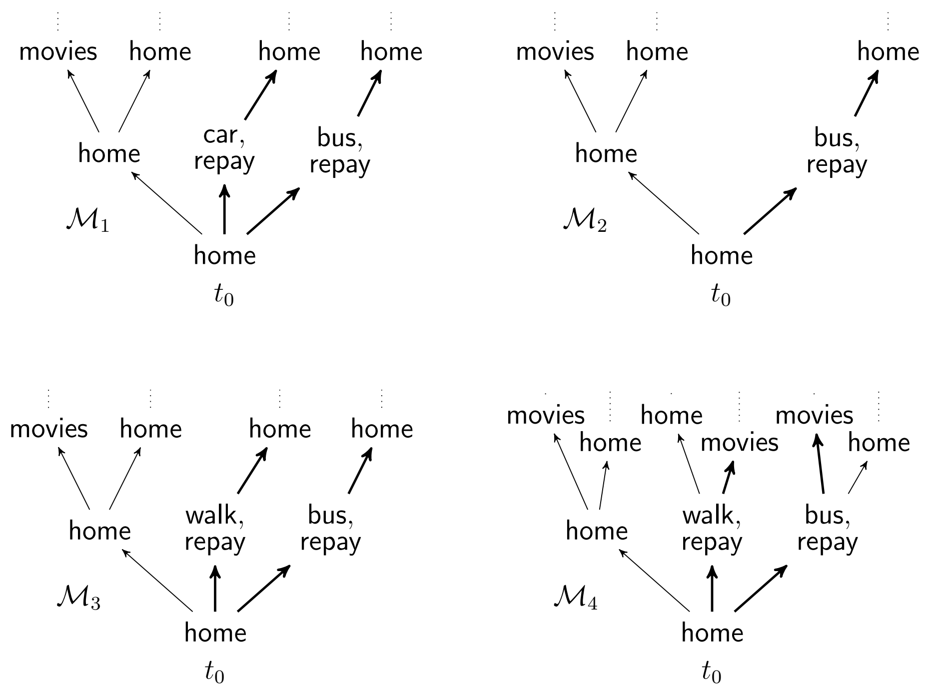


Figure 3.1: Four stages of planning about going to the movies and repaying money, indicated by the atoms repay and movies. In all models, we have $F\mathsf{repay} \succ_O F\mathsf{movies}$ and $N_O = \{(\llbracket F\mathsf{repay}\rrbracket^{t_0}, F\mathsf{repay}), (\llbracket F\mathsf{movies}\rrbracket^{t_0}, F\mathsf{movies})\}$. Bold lines denote each model's admissible subtree.

$\prec_\mathcal{T}$ and $\prec_{im}$ are in a tight relationship. As just shown, $\prec_{im}$ is definable from $\prec_\mathcal{T}$. Under our assumption that every history is isomorphic to the natural numbers, the converse is also true: $\prec_\mathcal{T}$ is the transitive closure of $\prec_{im}$. Hence, providing an enkratic model $\mathcal{M} = \langle \mathbf{T}, t_0, v, N_o, \succ_O \rangle$ is equivalent to providing an **extended enkratic model** $\mathcal{M} = \langle \mathbf{T}, t_0, v, N_o, \succ_O, \prec_{im} \rangle$ that includes the relation $\prec_{im}$. We will make use of this property later.

To technically define the dynamics of models, we refer to the concept of product updates with postconditions (see van Ditmarsch et al. (2008); Baltag et al. (1998) for some technical background). We focus on practical update models, which do not introduce or revoke any basic oughts the agents is exposed to. Rather these models merely change the tree of possible future histories.

Definition 3.4. *A* **(practical) update model** $\mathcal{E} = \langle S, s_0, R_{ims}, pre, post \rangle$ *consists of a set of states S with $s_0 \in S$, a relation $R_{ims} \subseteq S \times S$, and maps $pre : S \to \mathcal{L}_0$ and $post : S \to \mathcal{P}(\mathsf{At} \times \{\top, \bot\})$ such that $(p, \top) \in post(s) \Rightarrow (p, \bot) \notin post(s)$.*

Intuitively, S is a set of possible states with a temporal relation R_{ims} on it, similar to the relation $\prec_{im}$ on T. Each possible state or event s in S can match and modify moments t in T. However, the matching might be subjected to additional conditions to be met by t. These conditions are recorded in $pre(s)$. Finally, a state of the update model might prescribe a change to atomic valuation at moment t. This change is represented by the postcondition $post(s)$, marking when a valuation should be forced true (i.e., $(p, \top) \in post(s)$) or false (i.e., $(p, \bot) \in post(s)$).

For the purposes of this chapter, we make two additional assumptions on the update model. First, we demand the transitive closure R of R_{ims} to be a discrete tree order on S with root s_0 such that R_{ims} is the immediate predecessor relation of R. Second, we require that for any $s \in S$ the set $\{\neg pre(t) \mid sR_{ims}t\}$ is Λ_{temp} inconsistent. In other words: for any formula $\psi \in \mathcal{L}_0$ that is not a Λ_{temp} contradiction, there is some successor t of s such that $pre(t)$ is logically compatible with ψ.

The second of the above assumptions, that for any $s \in S$ the set $\{\neg pre(t) \mid sR_{ims}t\}$ has to be Λ_{temp} inconsistent, is non-standard. In fact, this assumption precludes classic approaches to public announcements or, more generally, the deletion of possible worlds. The condition is needed to ensure that the non-terminality axiom, D_G continues to hold in the updated model. As will become clear in the formal treatment of Example 3.3, this restriction is far less severe than it may seem at first sight. In fact, many cases of deletion can be mimicked by postconditions, adequately transforming superfluous worlds.

Definition 3.5. *Let $\mathcal{M} = \langle \mathbf{T}, t_0, v, N_o, \succ_O, \prec_{im}\rangle$ be an extended enkratic model and $\mathcal{E} = \langle S, s_0, R_{ims}, pre, post\rangle$ be a practical update model. The product* **update of $\mathcal{M}$ with $\mathcal{E}$**, *denoted by $\mathcal{M} \otimes \mathcal{E}$, is the extended enkratic model $\langle\langle T \otimes S, \prec'_T\rangle, (t_0, s_0), v', N'_o, \succ'_O, \prec'_{im}\rangle$ defined as follows:*

- $T \otimes S = \{(t, s) \in T \times S | \mathcal{M}, t \vDash pre(s)\}$
- *Define a relation $\prec'_{im}$ on $T \otimes S$ as $(t, s) \prec'_{im} (t, s')$ iff $t \prec_{im} t'$ and $sR_{ims}s'$. The relation $\prec'_T$ is then the transitive closure of $\prec'_{im}$*
- *The valuation $v' : \mathsf{At} \to \mathcal{P}(T \otimes S)$ is defined by $(t, s) \in v'(p)$ iff either i) $(p, \top) \in post(s)$, or ii) $t \in v(p)$ and $(p, \bot) \notin post(s)$*
- $([\![\varphi]\!]^{(t_0,s_0)}_{\mathcal{M}\otimes\mathcal{E}}, \varphi) \in N_O^{\mathcal{M}\otimes\mathcal{E}}$ *iff* $([\![\varphi]\!]^{t_0}_{\mathcal{M}}, \varphi) \in N_O^{\mathcal{M}}$
- $\succ_O^{\mathcal{M}\otimes\mathcal{E}} = \succ_O^{\mathcal{M}}$.

To begin with, we note that $\mathcal{M} \otimes \mathcal{E}$ is indeed an *enkratic* model. The proof of the following lemma can be found in the appendix, along with all other proofs of this section.

Lemma 3.6. *Let $\mathcal{M}$ be an extended enkratic model and $\mathcal{E}$ a practical update model. Then $\langle\langle T \otimes S, \prec'_T\rangle, (t_0, s_0), v', N'_o, \succ'_O, \prec'_{im}\rangle$ is an extended discrete enkratic model.*

To demonstrate the versatility of this approach, we show how all three updating steps in Example 3.3 can be represented with update models. The first update model $\mathcal{E}_1$ in Figure 3.2 corresponds to learning that the car is not available. Note that by the second additional assumption on update models, the set $\{\neg pre(s') \mid sR_{ims}s'\}$ has to be inconsistent for any $s \in S$. We hence cannot simply delete the car worlds, but need to replace going by car with something else, in this case going by bus. The next update model $\mathcal{E}_2$ corresponds to learning that I could walk to my friend's house. Here, a copy of the going-by-bus world is transformed into a walking-world. Finally, $\mathcal{E}_3$ corresponds to learning that I have sufficient money to see the movies even after repaying my friend.

The three update models displayed in Figure 3.2 generate the sequence of models $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \mathcal{M}_4$ depicted in Figure 3.1. More precisely, we have that $\mathcal{M}_2 \otimes \mathcal{E}_2 = \mathcal{M}_3$ and $\mathcal{M}_3 \otimes \mathcal{E}_3 = \mathcal{M}_4$. For the transition from $\mathcal{M}_1$ to $\mathcal{M}_2$ this is not fully true: $\mathcal{M}_1 \otimes \mathcal{E}_1$ is not the same as $\mathcal{M}_2$, since the former has two duplicate branches of going by bus. However, as $\mathcal{M}_2$ can be gained from $\mathcal{M}_1 \otimes \mathcal{E}_1$ by removing one of these duplicate branch, both models are logically equivalent.

3.3.2 Back to Challenge III: Dynamic Conditions

We show several general results that illustrate the complex relationship between updates, goals and derived oughts. Practical update models, despite not changing the set of basic oughts an agent is exposed to, can have intricate and non-monotonic effects on the agent's goals or derived oughts. In particular we show that the *agent's set of goals need not necessarily grow when her available options grow, and it need not shrink if her options shrink.*

To formulate these results, we fix a piece of notation. Let $Hist(\mathcal{T})$ denote the set of histories of a tree $\mathcal{T}$. For a given situation $\mathcal{M}$, we call practical update model $\mathcal{E}$ a **restriction** if $Hist(\mathcal{M} \otimes \mathcal{E}) \subset Hist(\mathcal{M})$, and an **expansion** if $Hist(\mathcal{M}) \subset Hist(\mathcal{M} \otimes \mathcal{E})$. Note that the first update in Example 3.3 is a

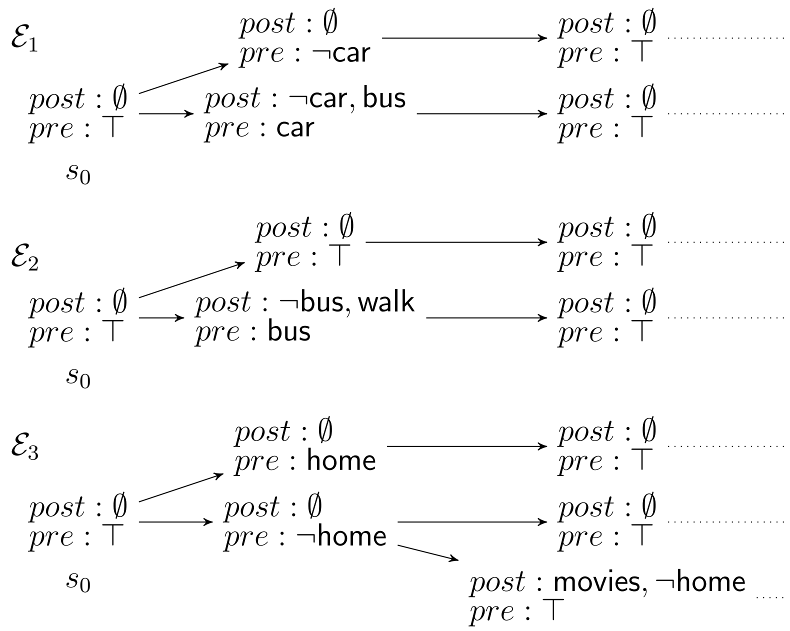


Figure 3.2: Practical update models corresponding to the three updating steps from Example 3.3

restriction, while the second and third update exemplify expansions.

We can turn now to two structural results about restrictions and expansions and how these impact the agent's goals. In the following, we write $n_G^{\mathcal{M}}$ to denote the goal formulas an agent pursues in model $\mathcal{M}$ (more formally, $n_G = \{\varphi \mid \mathcal{M}, t_0 \vDash Goal\varphi\}$).

Lemma 3.7. *Let $\mathcal{M}$ be an enkratic model where for each $([\![\varphi]\!]^{t_0}, \varphi)$ and $([\![\psi]\!]^{t_0}, \psi)$ in N_O with $\psi \succ_O \varphi$ it holds that $\psi \in n_G^{\mathcal{M}}$ whenever $\varphi \in n_G^{\mathcal{M}}$. Let $\mathcal{E}$ be an expansion of $\mathcal{M}$. Then $n_G^{\mathcal{M}} \subseteq n_G^{\mathcal{M}\otimes\mathcal{E}}$.*

Lemma 3.7 identifies a condition under which the agent's set of goals increases if new options become available to her. This additional condition is crucial. In general, an agent might drop some of her goals when new options become available. The following example provides an enkratic model $\mathcal{M}$ and an expansion $\mathcal{E}$ of $\mathcal{M}$ such that $n_G^{\mathcal{M}} \not\subseteq n_G^{\mathcal{M}\otimes\mathcal{E}}$.

Example 3.8. *Consider the enkratic models $\mathcal{M}$, $\mathcal{M}'$ displayed in Figure 3.3. In both models we set $N_O = \{([\![F\mathsf{p}]\!]^{t_0}, F\mathsf{p}), ([\![F\mathsf{q}]\!]^{t_0}, F\mathsf{q}\}), ([\![G\mathsf{r}]\!]^{t_0}, G\mathsf{r}\})$ and $F\mathsf{p} \succ_O F\mathsf{q} \succ_O G\mathsf{r}$. Clearly, $Hist(\mathcal{M}) \subseteq Hist(\mathcal{M}')$ and there is an update model $\mathcal{E}$ such that $\mathcal{M}' = \mathcal{M} \otimes \mathcal{E}$. Hence $\mathcal{M}'$ is an expansion of $\mathcal{M}$. In $\mathcal{M}$ we have $\mathcal{M}, t_0 \vDash GoalF\mathsf{p}$ and $\mathcal{M}, t_0 \vDash GoalG\mathsf{r}$ but $\mathcal{M}, t_0 \not\vDash GoalF\mathsf{q}$. In the*

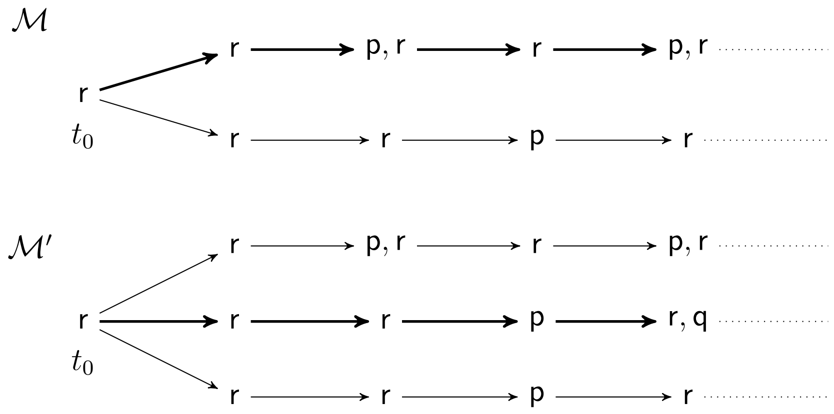


Figure 3.3: An expansion $\mathcal{M}'$ of $\mathcal{M}$ that does not $\subseteq$-increase the set of goals. Bold lines indicate the admissible subtree.

expansion $\mathcal{M}'$, on the other hand, we have $\mathcal{M}', t_0' \vDash GoalF\mathsf{p} \wedge GoalF\mathsf{q}$ but $\mathcal{M}', t_0' \nvDash GoalG\mathsf{r}$. Hence $n_G^{\mathcal{M}} \not\subseteq n_G^{\mathcal{M}\otimes\mathcal{E}}$.

Next, we turn to restrictions. Here, we show that an agent's goal set cannot grow as some of her options are removed.

Lemma 3.9. *Let $\mathcal{M}$ be an enkratic model and let $\mathcal{M} \otimes \mathcal{E}$ be a restriction of $\mathcal{M}$. Then $n_G^{\mathcal{M}} \not\subset n_G^{\mathcal{M}\otimes\mathcal{E}}$.*

This does not imply that a restriction cannot give rise to new goals. When certain goals become unreachable, other basic oughts the agent had discarded before may move into focus. In formal terms: It is, in general, not true that $n_G^{\mathcal{M}\otimes\mathcal{E}} \subseteq n_G^{\mathcal{M}}$. This is the subject of the following example.

Example 3.10. *We use the same construction as in Example 3.8. Again, consider the enkratic models $\mathcal{M}$, $\mathcal{M}'$ displayed in Figure 3.3. Again, we set $F\mathsf{p} \succ_O F\mathsf{q} \succ_O G\mathsf{r}$ and $N_O = \{([\![Fp]\!]^{t_0}, F\mathsf{p}), ([\![F\mathsf{q}]\!]^{t_0}, F\mathsf{q}\}), ([\![G\mathsf{r}]\!]^{t_0}, G\mathsf{r}\})$ in both models. Evidently, $Hist(\mathcal{M}) \subset Hist(\mathcal{M}')$ and there is an practical update model $\mathcal{E}$ such that $\mathcal{M}$ is equivalent to $\mathcal{M}' \otimes \mathcal{E}$. Hence $\mathcal{M}$ is a restriction of $\mathcal{M}'$. Then the same argument as above shows that $n_G^{\mathcal{M}'\otimes\mathcal{E}} \not\subseteq n_G^{\mathcal{M}'}$.*

The two examples above illustrate that purely practical updates can have complex and non-monotonic effects on the agent's goals or derived oughts. In particular, even practical updates merely expanding the set of available histories may trigger the agent to drop certain goals of hers. Partially, such phenomena are due to the exact choice of updating rule. In the present framework, the agent calculates her set of goals from scratch after each update, picking as the new goal set the maximally consistent subset of basic oughts that is maximal in the lexicographic order induced by $\succ_O$.

An alternative updating policy might opt for minimal changes instead, adopting as the updated set of goals some maximally consistent subset that differs minimally from the goal set the agent pursued before the update. While immune to the non-monotonicity described above, such minimal change rules may trigger a different type of non-conservativeness. Take for instance a two step update where an agent is first informed that some history h of a given enkratic tree model $\mathcal{M}$ is not available, followed by a second update indicating that the first information was wrong and h is, in fact available. After executing both updates, the tree of available options is exactly as it was in the starting model $\mathcal{M}$. With the original updating policy described above, the set of goals after both updates is also the same as the initial goal set. However, this would, in general, cease to hold if we followed a minimal change rule instead.[4] We take this to illustrate that the existence of complex interaction patterns between changes of available histories and the set of goals pursued does not hinge on the exact updating policy — rather, it is a general fact of planning when exposed to possibly incompatible sets of basic oughts.

This completes our investigation of the effects of practical dynamics on basic oughts, goals and derived oughts, and concludes the project we embarked on in the previous chapter.

3.4 Conclusion and Open Ends

We pursued two aims: analyzing the logical structure of ENKRASIA, and addressing some of the implications ENKRASIA has, in combination with certain other principles of practical rationality, for deontic logic.

As for the first aim, we have argued that ENKRASIA is a principle of bounded validity. Goals are subjected to two requirements of INTERNAL and STRONG

[4] To see that this holds true for *any* minimal change updating rule, consider a enkratic model $\mathcal{M}$ consisting of three branches f, g and h. Moreover, assume N_O to contain three basic oughts O_1 - O_3 which are satisfied in the subtrees $\{f, g\}, \{f, h\}$ and $\{g, h\}$ respectively. Maximally consistent subsets hence are $\{O_1, O_2\}, \{O_1, O_3\}$ and $\{O_2, O_3\}$. Assume wlog that $\{O_1, O_2\}$ is adopted as set of goals in $\mathcal{M}$. After removing branch f, this set is not consistent anymore, now only $\{O_1, O_3\}$ and $\{O_2, O_3\}$ are maximally consistent. One of these is selected as new set of goals, wlog $\{O_1, O_3\}$. Since $\{O_1, O_3\}$ remains maximally consistent after adding f again, any minimal change rule must retain it as set of goals. In particular, the set of goals after removing and re-adding f is different from before.

CONSISTENCY which basic oughts are not. Both of these conditions set boundaries for the translation of basic oughts into goals. ENKRASIA, then, is only valid as far as it does not conflict with either requirement of consistency.

In relation to the second aim, we have elaborated on the distinction between basic and derived oughts. This distinction allows us to validate crucial deontic inferences (crucially, a reading of the so-called practical inference) without generating an unrestricted validity of deontic closure. In fact, by restricting the conclusions of deontic closure to derived oughts, many of the paradoxical implications usually associated with deontic closure no longer obtain.

The conceptual distinction between basic and derived oughts — ultimately grounded in their interaction with goals — surfaces also in their kinematics. We have provided a dynamical logical framework to account for the notion of practical dynamics, and formally investigated the stability conditions of oughts and goals under such dynamics.

Related approaches. To the best of our knowledge, ENKRASIA has not been previously investigated explicitly from a logical perspective. There are, however, a variety of logical frameworks that deal with notions pertaining to practical rationality. We point to some of these. The analysis presented here is linked to the logical tradition of interpreting intentions according to Bratman's (1987) planning theory. Related works —mainly focusing on normal modal logic— include Cohen and Levesque (1990) and Lorini and Herzig (2008). The latter paper also discusses practical inference, though in the context of forming instrumental intentions rather than derived oughts.

For what concerns the dynamics of intentions and plans, related work includes van der Hoek et al. (2007), with a focus on the operation of deleting plans. A second reference is Icard et al. (2010), who provide an account of intention revision based on AGM theory. Further complementary work includes Craven and Sergot's (2008) account of permitted and obligatory actions within transition systems, and Broersen et al.'s (2001) syntactic, default-based approach on conflicts between beliefs, obligations, intentions and desires.

While various frameworks address the relationship between obligations, plans and intentions, combinability across approaches is sometimes hampered by the fact that these may refer to different things. Consider, for instance, Broersen et al.'s 2001 BOID framework on beliefs, obligations, intentions and desires mentioned above. BOID distinguishes, *inter alia*, between obligations (be they believed by the agent or not) and intentions, denoting actions the agent plans on doing. In the present framework, in contrast, basic oughts are assumed believed by the agent. Goals in a plan, moreover, correspond to basic oughts the agent decided to pursue. These form a category between BOID's obligations and intentions.

Future directions. The present approach fits in with a larger project on the logic of oughts in the context of practical rationality. The analysis and the framework presented here can be extended in different directions. We mention two.

One possible extension of our framework pertains to the relation between ENKRASIA and permissions. We have seen that the relation between oughts and goals is not unidirectional. Oughts translate into goals, but also further, derived oughts can be generated from a given set of goals. However, it seems that not only oughts, but also permissions can be derived from what the agent is committed to bring about. The admissible subtree, denoting the intersection of all goals pursued by the agent, can be thought of as a weakest permission, i.e., as the largest subtree the agent is permitted to arrive in, given her commitments. Naturally, stronger permissions may also hold, permitting the agent to arrive in *any* subtree of the admissible tree. A classic reference on this is Anglberger et al. (2015).

A second possible extension relates to the way in which possible future courses of events are represented. In the current framework we use a tree-structure to represent the agent's choices options among future unfoldings of events. This picture can be refined in various ways. We may, for instance, restrict the agent's ability to fully select the future course of actions. That is, agents may no longer be able to pick a specific branch, but merely to choose some subtree to stay within. Conceptually, this might require a refinement of the principle of STRONG CONSISTENCY, e.g., by demanding that if $Goal\varphi$ then the agent believes she has an available choice option that *guarantees* φ. Formally, this would amount to having equivalence classes of histories (representing choice uncertainty) paired with a quantification over such classes. Similar topics have been investigated by Horty (2001) and Ciuni and Zanardo (2010). The results presented in these works could form a fruitful starting point for expanding the logic of ENKRASIA.

3.5 Appendix: Proofs

Proof of Lemma 3.6: We begin with showing that $\prec'_{\mathcal{T}}$ is a tree-order on $T \otimes S$. Transitivity is immediate, as $\prec'_{\mathcal{T}}$ is transitively closed. For irreflexivity assume the contrary, i.e., assume that there is some (t', s') with $(t', s') \prec'_{\mathcal{T}} (t', s')$. Hence there are $(t', s') \prec'_{im} (t_2, s_2) \prec'_{im} \dots \prec'_{im} (t_n, s_n) = (t', s')$. By definition of $\mathcal{M} \otimes \mathcal{E}$, this entails that $t' = t_1 \prec_{im} t_2 \dots \prec_{im} t_n = t'$. In particular, we get $t' \prec_{\mathcal{T}} t'$, as $\prec_{\mathcal{T}}$ is the transitive closure of $\prec_{im}$. But this contradicts the fact that $\prec_{\mathcal{T}}$ is irreflexive.

Finally for inverse linearity, let $(t, s) \in T \otimes S$. We need to show that:

$$P = \{(t', s') \in T \otimes S | (t', s') \prec'_{\mathcal{T}} (t, s)\}$$

is linearly ordered by $\prec'_{\mathcal{T}}$. Let $(t', s'), (\tilde{t}, \tilde{s}) \in P$ be given. We have to show that $(t', s') \prec'_{\mathcal{T}} (\tilde{t}, \tilde{s})$ or $(\tilde{t}, \tilde{s}) \prec'_{\mathcal{T}} (t', s')$. By assumption, we have that $t', \tilde{t} \prec_{\mathcal{T}} t$. By construction and the assumption that $\prec_{\mathcal{T}}$ has the order type of the natural numbers, there is $n' \in \omega$ such that $(t', s') = (t'_0, s'_0) \prec'_{\mathcal{T}} \dots \prec'_{\mathcal{T}} (t'_{n'}, s'_{n'}) = (t, s)$ with $t'_0 \prec_{im} \dots \prec_{im} t_{n'}$ and $s'_0 R_{ims} \dots R_{ims} s_{n'}$. Likewise, there is $\tilde{n} \in \omega$ such that $(\tilde{t}, \tilde{s}) = (\tilde{t}_0, \tilde{s}_0) \prec'_{\mathcal{T}} \dots \prec'_{\mathcal{T}} (\tilde{t}_{\tilde{n}}, \tilde{s}_{\tilde{n}}) = (t, s)$ with $\tilde{t}_0 \prec_{im} \dots \prec_{im} \tilde{t}_{\tilde{n}}$ and $\tilde{s}_0 R_{ims} \dots R_{ims} \tilde{s}_{\tilde{n}}$. By definition of a practical update model, R_{ims} is a unique predecessor relation, i.e., $x R_{ims} z$ and $y R_{ims} z$ implies $x = y$. We hence have that $s'_{n'-1} = \tilde{s}_{\tilde{n}-1}$, $s'_{n'-2} = \tilde{s}_{\tilde{n}-2} \dots$. The same reasoning yields that $t'_{n'-1} = \tilde{t}_{\tilde{n}-1}$, $t'_{n'-2} = \tilde{t}_{\tilde{n}-2} \dots$. Since $(T, \prec_{\mathcal{T}})$ is a tree, we have that $t' \prec_{\mathcal{T}} \tilde{t}$, $\tilde{t} \prec_{\mathcal{T}} t'$ or $t' = \tilde{t}$. Without loss of generality, we assume the first, the other cases being similar. Since $t' \prec_{\mathcal{T}} \tilde{t}$, we have $n' > \tilde{n}$ and hence $\tilde{t} = t'_{n'-\tilde{n}}$. By the above, this implies that $(\tilde{t}, \tilde{s}) = (t'_{n'-\tilde{n}}, s'_{n'-\tilde{n}})$, and hence $(t', s') \prec'_{\mathcal{T}} (\tilde{t}, \tilde{s})$.

Finally, we have to show that the tree order $\prec'_{\mathcal{T}}$ is serial. To this end, let $(t, s) \in T \otimes S$. We have to show that there is some $(t', s') \in T \otimes S$ with $(t, s) \prec'_{\mathcal{T}} (t', s')$. Note that by seriality of $\mathcal{T}$, there is some t' with $t \prec_{\mathcal{T}} t'$. By our discreteness assumption, we can pick t' such that $t \prec_{im} t'$. Next, we claim that there is some $s' \in S$ with $s R_{ims} s'$ and $\mathcal{M}, t' \vDash pre(s')$. Note that this claim implies that $(t, s) \prec'_{\mathcal{T}} (t', s')$, finishing the proof. So let us show the claim. Assume not. That is, assume that $\mathcal{M}, t' \nvDash pre(s')$ for all $s' \in S$ with $s R_{ims} s'$. Hence, $\mathcal{M}, t' \vDash \neg pre(s')$ for all s' with $s R_{ims} s'$. This shows that $\{\neg pre(t) \mid s R_{ims} t\}$ is consistent, contradicting the assumption that S is a practical update model. □

Proof of Lemma 3.7: Since $\mathcal{M} \otimes \mathcal{E}$ is an expansion of $\mathcal{E}$, the set $S = \{[\![\psi]\!]^{(t_0, s_0)}_{\mathcal{M} \otimes \mathcal{E}} \mid [\![\psi]\!]^{t_0}_{\mathcal{M}} \in N^{\mathcal{M}}_G\}$ in $\mathcal{M} \otimes \mathcal{E}$ is consistent. By assumption, $n^{\mathcal{M}}_G$ is $\succ_O$-upward closed in $\{\varphi \mid ([\![\varphi]\!]^{t_0}_{\mathcal{M}}, \varphi) \in N^{\mathcal{M}}_O\}$ and hence also in $\{\varphi \mid ([\![\varphi]\!]^{t_0}_{\mathcal{M}\otimes\mathcal{E}}, \varphi) \in N^{\mathcal{M}\otimes\mathcal{E}}_O\}$. In particular, the $\prec_{Lex}$ maximally consistent subset of $\{[\![\varphi]\!]^{(t_0,s_0)} \mid ([\![\varphi]\!]^{(t_0,s_0)}, \varphi) \in N^{\mathcal{M}\otimes\mathcal{E}}_O\}$ contains S. This shows that $n^{\mathcal{M}}_G \subseteq n^{\mathcal{M}\otimes\mathcal{E}}_G$. □

Proof of Lemma 3.9: Assume for a contradiction that $n^{\mathcal{M}}_G \subset n^{\mathcal{M}\otimes\mathcal{E}}_G$. This implies that $n^{\mathcal{M}}_G \prec^{\mathcal{M}}_{Lex} n^{\mathcal{M}\otimes\mathcal{E}}_G$. Hence, by construction of the neighborhood $N^{\mathcal{M}}_G$, the set $\{[\![\varphi]\!]^{t_0}_{\mathcal{M}} \mid [\![\varphi]\!]^{t_0}_{\mathcal{M}\otimes\mathcal{E}} \in N^{\mathcal{M}\otimes\mathcal{E}}_G\}$ must be inconsistent, i.e., there is no history h of $\mathcal{T}$ such that $h \in [\![\varphi]\!]^{t_0}_{\mathcal{M}}$ for all φ with $[\![\varphi]\!]^{t_0}_{\mathcal{M}\otimes\mathcal{E}} \in N^{\mathcal{M}\otimes\mathcal{E}}_G$. This, however, is impossible: $N^{\mathcal{M}\otimes\mathcal{E}}_G$ is by definition consistent in $\mathcal{M}\otimes\mathcal{E}$, i.e., there is some history h of $\mathcal{M}\otimes\mathcal{E}$ with $h \subseteq [\![\varphi]\!]^{t_0}$ whenever $[\![\varphi]\!]^{t_0} \in N^{\mathcal{M}\otimes\mathcal{E}}_G$. Since $\mathcal{M}\otimes\mathcal{E}$ is a restriction of $\mathcal{M}$, h is also a history of $\mathcal{M}$, showing that $\{[\![\varphi]\!]^{t}_{\mathcal{M}} \mid [\![\varphi]\!]^{t}_{\mathcal{M}\otimes\mathcal{E}} \in N^{\mathcal{M}\otimes\mathcal{E}}_G\}$ is consistent. □

Chapter 4

Objective Oughts and Reasoning by Cases

In this chapter, we focus on future-dependent oughts, i.e., oughts whose truth (now) is dependent on what will happen in the future. We have two aims: (i) investigating the semantics of future-dependent *objective* oughts and (ii) investigating the validity of Reasoning by Cases for objective oughts. These two aims are related. It is shown that *there is* an interpretation of future-dependent objective oughts for which Reasoning by Cases is not valid.

4.1 Introduction

A familiar view admits that certain oughts in morality and natural language are objective.[1] Consider Oedipus, Jocasta's son. There exists a sense of oughts, the objective one, according to which the sentence "Oedipus ought not marry Jocasta" is true. To a first approximation, objective oughts indicate what is the best action to perform regardless of an agent's epistemic limitations. They are a function of facts of the world, known or unknown, and of sets of values. Hence, in our mythical example, the sentence "Oedipus ought not marry Jocasta" describes what is the right thing to do, the best course of

[1] See, e.g., discussions in Mellor (1983), Oddie and Menzies (1992), Broome (2013), Wedgwood (2003), Wedgwood (2016), Bykvist (2009), Hare (2011), Silk (2014a), Carr (2015). Some of the best-known semantic frameworks for deontic modals are also based on such a notion of objective ought. Those frameworks typically involve a state (or modal base) of possible worlds describing a body of contextually relevant facts, and an ideality function which indicates the deontically best worlds within the state. Kratzer (1991) and Kratzer (2012d)'s semantics is an example: in the Kratzerian framework, deontic modals such as *ought*, *must*, *might* indeed quantify over a *circumstantial* (i.e., not epistemic) modal base.

action given the moral values on incest and given what *is* actually the case — namely, that Jocasta is in fact Oedipus' mother. That Oedipus is ignorant of this plays no role.

Being that objective oughts are a function of the facts of the world, a *futurity problem* emerges: Can objective oughts depend on what will happen in the future? If yes, how should they be accounted for?

We address the above questions, and investigate three solutions to the futurity problem: (i) denying any objective interpretation of future-dependent oughts, (ii) evaluating future-dependent objective oughts with respect to the relevant future facts, (iii) evaluating future-dependent objective oughts with respect to objective probabilities. It is shown that each solution triggers different conceptual and logical implications for oughts. The probabilistic account (iii) is the least standard: nevertheless we argue that it meets some plausible desiderata about the semantics of future-dependent objective oughts, while setting an objectivist standard for decision making under indeterminacy and risk.

The analysis of the futurity problem and of its possible solutions complements the existing literature on objective oughts and, in particular, contributes to the current debate surrounding the so-called Miners' Puzzle. Firstly, it is shown that different senses of objective oughts arise depending on which solution to the futurity problem is adopted. The characterization of objective oughts as oughts from a God's-eye view (Gibbard, 2005) — that is, oughts determined by what will in fact be the case — is, contrary to the common conception, just one of those possible senses.

Relatedly, the analysis presented in this chapter discredits the misconception that objective oughts obey classical rules of inference. Recent literature on the Miners' Puzzle has shown that oughts that are relative to the agent's epistemic standpoint do not validate the inference rule of Reasoning by Cases (Kolodny and MacFarlane, 2010; Willer, 2012; Carr, 2015; Cariani et al., 2013; Bledin, 2015). However, the Miners' Puzzle emerges in a scenario of uncertainty, therefore leaving the logic of objective oughts unaffected. We show that, if the future is possibly indeterminate, then a Miners-like Puzzle emerges for future-dependent objective oughts. Hence, Reasoning by Cases is not generally valid for objective oughts either.

The chapter has the following structure. In Section 4.2, we introduce objective oughts. In Section 4.3, we discuss the Miners' Puzzle and, more generally, Reasoning by Cases under uncertainty. Future-dependent objective oughts and possible solutions to the futurity problem are investigated in Sec.4.4. In Section 4.5, we bring to attention a Miners-like puzzle (called the *Betting Puzzle*) involving (probabilistic) future-dependent objective oughts. It is argued that if the Miners' Puzzle shows that Reasoning by Cases is invalid under uncertainty, the Betting Puzzle shows that Reasoning by Cases is invalid under indeterminacy. Section 4.6 concludes. The appendix 4.7 contains the relevant formal apparatus. It presents a formal semantics for probabilistic future-dependent objective oughts, and shows how such semantics solves the Betting Puzzle. It concludes by comparing the Betting Puzzle to a similar scenario proposed by Horty (2001).

4.2 Objective Oughts

Let us begin with a general schema (GS) to represent objective views on oughts (adapted from Bykvist, 2009):

(GS) At time t an agent α ought objectively to X if and only if X-ing is best in light of F, with F non-epistemic.

In other words, F is a place-holder for whatever makes the action X best, with the constraint that F is not analyzable in epistemic terms. F has, in this minimal sense, an *objective character*.

F's objective character, however, does not necessarily exclude any reference to subjective features of the agent in S, such as her motives, values or desires. For instance, objectivists do not need to deny the role that Greek values on incest play in determining that Oedipus ought not to marry Jocasta. What objectivists need to exclude is that F is defined in terms of the agent's evidence. That Oedipus believes that Jocasta is not his mother is not relevant to determine what he objectively ought to do.

This chapter focuses on a common specification of the above schema, namely the one that analyzes F in terms of facts of the world (for the moment, we set aside the contribution of subjective features such as values). According to such specification, objective oughts are defined as follows (see Gibbard, 2008; Broome, 2013; Kolodny and MacFarlane, 2010):

Definition 4.1. *At time t an agent α* **ought objectively** *to X if and only if X-ing is best in light of the relevant facts.*

Let us call the above definition FACTS. What are those *facts*? To the best of our knowledge, proponents of FACTS do not explicitly commit to any (metaphysical) account of facts. For the purposes of this chapter, we adopt a standard *semantic* account of facts: facts are those propositions that are **settled true**. For instance, it is a fact that Jocasta is the mother of Oedipus in the sense that "Jocasta is the mother of Oedipus" is settled true.[2] Not all settled true propositions will do, however. To qualify as a definition of objective oughts, the relevant facts in FACTS need to be **non-epistemic**: facts about what Oedipus believes, for instance, do not count. In what follows, we will thus talk about facts as *non-epistemic propositions that are settled true*. Moreover, we will treat ought and should as synonyms.

4.3 Reasoning by Cases Under Uncertainty

Objective oughts are not affected by problematic scenarios involving reasoning under uncertainty. A prominent example of these scenarios is the so-called Miners' Puzzle. Here is how Kolodny and MacFarlane (2010) introduce the puzzle:

> Ten miners are trapped either in shaft A or in shaft B, but we do not know which. Flood waters threaten to flood the shafts. We have enough sandbags to block one shaft, but not both. If we block one shaft, all the water will go into the other shaft, killing any miners inside it. If you block neither shaft, both shafts will fill halfway with water, and just one miner, the lowest in the shaft, will be killed.
>
> (Kolodny and MacFarlane, 2010, p.115)

Why is it a puzzle? Given the scenario, the following sentences seem all to be acceptable:

1. You ought to block neither shaft
2. The miners are in shaft A or the miners are in shaft B
3. If the miners are in shaft A, then you ought to block shaft A
4. If the miners are in shaft B, then you ought to block shaft B

[2] There are specific contexts in which a more fine-grained semantic account of facts is to be preferred. For instance Kratzer (2012b) argues that, for what concerns the semantics of the verb *to know* and the semantics of counterfactuals, facts need to be something more specific than proposition— rather, particulars or information units. In this chapter, we adhere to the more standard semantic account of facts as settled true propositions.

However, from the premises (2), (3), (4) via Reasoning by Cases it follows that:[3]

5. You ought to block shaft A or you ought to block shaft B

which, in turn, contradicts (1). The puzzle has been taken to show that Reasoning by Cases is invalid for deontic arguments(Kolodny and MacFarlane, 2010; Willer, 2012; Carr, 2015; Cariani et al., 2013; Bledin, 2015).

For the purposes of the present chapter, two observations are particularly relevant. Firstly, in order to show that Reasoning by Cases fails in the Miners' scenario one does not need a premise as strong as (1). A weaker premise ($1a$) "It is not the case that we ought to block shaft A and it is not the case that we ought to block shaft B" would suffice — as it would correspond to the negation of (5). Premise (1) says something more: we do not simply lack the obligation to block either shaft, we have the obligation to block neither. Such a strong premise may emerge, for instance, from the application of a decision norm such as *MaxiMin*, or *maximization of expected utility.* Table 4.1 depicts the decision problem faced in the Miners' scenario.

	miners are in A	miners are in B
Block A	10	0
Block B	0	10
Block neither shaft	9	9

Table 4.1: Decision problem for the Miners' Puzzle

In Table 4.1, the numerical values are derived from the number of miners saved (thus reflecting the moral values at place in the scenario). The states "miners are in A" and "miners are in B" are equally likely, epistemically speaking (hence, in terms of subjective probabilities). We ought to block neither shaft because that action guarantees the best worst-case outcome (according to MaxiMin) or because that action has the greatest expected utility.[4]

The second, and most important, observation is as follows: the Miners' scenario emerges as a puzzle because it involves a certain epistemic uncertainty, namely,

[3] Reasoning by Cases is a deductive inference rule from the premises $\ulcorner\varphi_1$ or $\varphi_2\urcorner$, $\ulcorner$if φ_1 then $\psi_1\urcorner$ and $\ulcorner$if φ_2 then $\psi_2\urcorner$ to the conclusion $\ulcorner\psi_1$ or $\psi_2\urcorner$.

[4] See Cariani et al. (2013) for an analysis of the Miners' Puzzle in terms of the decision rule MaxiMin, Lassiter (2016) for an analysis in terms of maximization of expected utility, and Carr (2015) for a general overview of the various approaches adopted to solve the Miners' Puzzle.

the uncertainty on where the miners are. Such uncertainty is relevant for oughts that are sensitive to the beliefs/information that an agent possesses.[5] However, it does not affect objective oughts. The position of the miners is already settled. In the sense described above, it is a fact that the miners are in shaft A or it is fact that the miners are in shaft B. In either case, it follows that the objectively best thing to do is blocking the shaft the miners are in. Thus, under the objective reading of ought as in FACTS, (1) is false (as is the weaker (1*a*)), and the conclusion (5) is correctly derived from (2)-(4) by Reasoning by Cases. Objective oughts appear, therefore, to obey such classical inference rule.

4.4 Future-Dependent Objective Oughts

Having introduced the main characteristics of objective oughts, let us now address the *futurity problem* that FACTS generates. The problem centers on the notion of future-dependent oughts, which we clarify below.

With the term *future-dependent oughts* we mean oughts whose truth at t depends on considerations about what will happen at a later time t'. These future-dependent oughts are ubiquitous in morality and natural language.

Consider, for instance, consequentialist ethical theory. According to consequentialism, a sentence like "Agent α ought to do X" is true at t if the action X will bring about the best consequences. Given the temporal relation between an action and its consequences (with the consequences happening *after* the action), consequentialist oughts are thus future-dependent in the sense just outlined: what counts as morally relevant for the rightness of an action at t happens indeed at a later time t'.

Future-dependent oughts are not only technical terms in ethical theory; they are part of natural language as well. We can distinguish at least two kinds of sentences involving future-dependent oughts: (i) sentences that feature oughts depending on some future object or individual (as in "You ought to now congratulate the winner of tomorrow's match", where the referent of the definite description "the winner of tomorrow's match" will only be determined in the future); and (ii) sentences that feature oughts depending on future states of affairs or propositions. This chapter restricts the attention to sentences of the latter kind.

[5]See Carr (2015) for an analysis of subjective oughts, and MacFarlane (2014), Kolodny and MacFarlane (2010) and Silk (2014b) for an account of informational oughts.

Sentences involving future-dependent oughts appear in contexts of both advice and deliberation:

6. ADVISER: You ought to buy fire insurance for your house
7. AGENT: If the coin will land tails, I should bet on tails

The first sentence is an example of advice under risk: considerations about whether and with which probability your house will — in the future — catch fire are, amongst other things, relevant to determine whether you ought to get fire insurance. The second sentence is an example of a conditional of practical deliberation: considerations (and suppositions) on how the coin will land are relevant when deciding how to place a bet.[6]

Now we can introduce the **futurity problem**: Is it possible to interpret future-dependent oughts objectively? That is, are there future-dependent objective oughts? If yes, how should those oughts be evaluated?

FACTS, which interprets objective oughts as functions of facts, does not provide a straightforward answer to these questions. The reason being that, while it seems natural to think of facts about the present and the past (about what is happening or happened), and therefore of objective oughts depending on them, the future is characterized by conceptual and formal challenges. How do considerations about what will happen in the future fit in such a picture of objective oughts? The rest of the chapter is devoted to outline and evaluate possible solutions to this problem. Far from being an orthogonal issue, we show that this has concrete implications for objectivist ethical theory and deontic logic.

To qualify as a solution to the futurity problem, three minimal desiderata are in place:

- *Conservativity.* An adequate solution to the puzzle should (i) maintain an objectivist character, that is, it should fit with the general schema [GS] presented on p.67, and (ii) at most extend FACTS to account for future-dependent oughts.
- *Univocality.* An adequate solution should provide a univocal definition of objective oughts, be they dependent on the past, present or future.

[6] For the sake of simplicity, we consider only *indicative* conditionals of deliberation. For a general defense of conditionals of deliberation as indicatives — and not as counterfactuals — see DeRose (2010).

- *Neutrality about the Future.* An adequate solution should not rely on any structural semantic commitment about future contingency.

The rationale behind the above desiderata is the following. Conservativity assures minimality of change: an adequate solution to the futurity problem should still qualify as objectivist, and provide the same account as FACTS for what concerns oughts that are dependent on the present or the past. Univocality recommends that there be a unique, fixed meaning of objective oughts. Finally, the last desideratum recommends that a solution to the futurity problem does not presuppose any particular answer to the question of future contingency (cf. MacFarlane, 2003).

Let us illustrate the above desiderata via an example. We can imagine an account of objective oughts going roughly as follows: objective oughts that are dependent on the past and the present are defined in light of the relevant facts, while future-dependent oughts are defined in light of a different feature Φ. If the feature Φ were non-epistemic, the account would satisfy conservativity (as it would fit with [GS] and, at most, extend FACTS). However, it would not satisfy univocality: objective oughts would be given a different definition whether they depend on the past, the present or the future. Finally, an account committed to the impossibility of future contingents would violate the desideratum of neutrality.

Equipped with the above desiderata, we turn now to the possible solutions of the futurity problem.

4.4.1 No Future-Dependent Objective Oughts

The first, simplest solution to the futurity problem consists in keeping FACTS as it is, and denying that future-dependent oughts can have an objectivist interpretation. To the extent that FACTS treats objective oughts as functions of facts, and that what will happen in the future is not *yet* a fact, there is no way to determine objectively the truth of sentences involving future-dependent oughts. A natural semantic background for such an account of objective oughts is given by the so-called "growing block" frameworks.[7]

[7] According to "growing block" frameworks, only the past and present objectively exist. Semantically, the history of the actual world ends at the present time, after which there is nothing. Such history, however, grows with the passing of time: new slices are added, and the edge of history moves forward. Since on growing block frameworks there is only the past and the present, FACTS can account only for objective oughts that depend on what happened in the past or is happening in the present. For a critical discussion of the growing

This solution is compatible with having other types of oughts be future-dependent, for instance oughts based on an agent's beliefs or credences about what will happen in the future. It is under the objectivist reading that the truth of future-dependent oughts would not be determined.

Although this solution meets the three desiderata above, it appears too strong. Which implications would such account of objective oughts have? Consider the following scenario: a bookmaker offers you a bet on a coin toss. Assume that, perhaps unknown to the bookmaker herself, both coin's sides are heads, and therefore the coin will certainly land heads. Supposing that you value monetary gain, it seems natural to say that you objectively ought to bet on heads. However, the latter ought is a future-dependent one: after all, what assures monetary gain is the side on which the coin *will* land. Hence, if this solution to the futurity problem were adopted, the truth of "You should bet on heads" could not be determined objectively. This is too cautious an account of objective oughts and objective decision making. In what follows, we show that there are alternative ways that can make sense of (at least some) future-dependent objective oughts.

4.4.2 Future-Dependent Objective Oughts and Facts

A second way of accounting for future-dependent objective oughts is to treat what will happen in the future as a fact. There would be future facts — in the sense of non-epistemic propositions about the future that are settled true — and their status would be on a par with present and past facts. This solution to the futurity problem would not require any modification of FACTS.

This solution would meet the desiderata of conservativity and univocality. There are at least two semantic frameworks in which the desideratum of neutrality can be satisfied: these correspond to the so-called "Peircean" approach to branching time or the so called "thin red line" approach.[8]

Peircean Approach and Branching Time

A widespread semantic approach in temporal logics consists in allowing the temporal evolution of the world to have the shape of a branching tree: the

block views, see Merricks (2006).

[8]We are therefore excluding semantic frameworks that impose linearity of time. Those frameworks would permit one to represent future facts in terms of propositions about the future that are settled true, but they would be committed to denying future contingency — hence not satisfying the desideratum of neutrality.

future is possibly open, and branches represent many possible continuations of the world (Belnap and Green, 1994; Horty and Belnap, 1995; MacFarlane, 2003). For the purposes of this chapter, it is crucial to interpret the branches as genuine *objective possibilities*. The future is possibly open not only in an epistemic sense, but *in re* (see Belnap and Green, 1994, p. 365).

Formally, branching trees can be defined as follows:

Definition 4.2. *Let T be a non-empty set of moments. A* **tree** *is an irreflexive ordered set* $\mathbf{T}=\langle T, \prec\rangle$ *in which the set of the $\prec$-predecessors of any moment t of T is linearly ordered by $\prec$. A* **history** *is a maximal linearly ordered subset of* $\mathbf{T}$.

Such representation in terms of an objective branching tree is neutral with respect to future contingency: the reason being that a linear time is just a special case of branching time.

According to the *Peircean interpretation* of the future (Prior (1967) and Thomason (1984)), at t it is true that $\ulcorner Will\varphi \urcorner$ if and only if φ is true in all future branches of t. Under this interpretation, trees can in fact provide an adequate formal background for future facts. We have said that future facts are propositions about the future that are settled true; under the Peircean interpretation, this amounts to saying that $Will\varphi$ is a future fact if and only if φ is true in all possible futures. Or, to use a toy example: it is already a fact that the coin will land tails if and only if the coin lands tails in all the future possibilities that are open to us now.

Let us consider some further characteristics of the Peircean interpretation of future facts. Firstly, if it is not a fact that the coin will land tails, then it might be the case that in some possible futures the coin lands tails, and in some other possible futures the coin does not land tails.[9] Secondly, even if it is now a fact that the coin will land tails, it does not follow that that has always been so. There might have been a future, possible in the past but not possible anymore, in which it was not a fact that the coin would have landed tails (see Thomason, 1984, p.143).

What are then the implications of having a Peircean account of future-dependent objective oughts? Such a solution to the futurity problem implies that only oughts that depend on already settled propositions about the future can be

[9] On the Peircean interpretation, $\ulcorner \neg Will\varphi \urcorner$ differs from $\ulcorner Will\neg\varphi \urcorner$: the latter implies the former, but not viceversa (see Prior, 1967, p.129).

interpreted objectively. In other words, this solution cannot account for objective oughts that depend on a chancy future. Sentences like "You ought to get fire insurance for your house", in contexts in which there is a certain risk your house will catch fire, or "You should bet on tails", in contexts in which a coin is 0.9 biased, cannot be interpreted objectively — the reason being that, in those contexts, it is not already settled whether your house will catch fire or the coin will land tails. These do not count as facts under the Peircean interpretation, and hence fall outside FACTS.

The Thin Red Line and Branching Time

There is one further approach that could provide a proper semantic background for FACTS: the so-called *thin red line*.[10] We have seen that, according to the standard interpretation of branching time, the world could evolve in the future through different histories. Among all those histories, however, there is only one which corresponds to how the world will actually be. The thin red line precisely marks out the actual future.

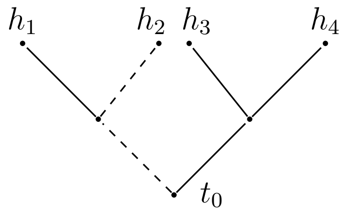


Figure 4.1: Dashed: the *thin red line*.

Consider again our coin example. There is a sense in which at t_0, the moment in which the coin is tossed, the future is open. It is possible that the coin lands tails and it is possible that the coin lands heads. Amongst those two possibilities, however, only one will be realized. That is the one picked up by the thin red line. So, at t_0 it is a fact that the coin will land tails, if that is what will actually happen, i.e., if the coin lands tails at a certain future moment t within the thin-red line. Postulating a thin red line would enable us to have both future facts and different future possibilities accounted for. We can say that it is a fact that the coin will land tails without committing to saying that that is inevitable; after all, the coin could land heads.

What are the implications of the thin red line on future-dependent objective oughts? Consider again the sentences "You ought to get fire insurance for your

[10] For a critical presentation of the thin red line approach, see Belnap and Green (1994); MacFarlane (2003).

house" and "You ought to bet on tails" in the context of a chancy future. Under the thin red line interpretation, these future-dependent oughts can receive an objective reading: their truth (or falsity) depends on whether your house will actually catch fire or the coin will actually land tails, that is, on whether your house catching fire or your coin landing tails lie on the thin red line.

The account of future-dependent objective oughts that emerges here is what Oddie and Menzies (1992) call *actual-outcome objectivism*, and Hare (2011) describes as "what an omniscient creature would advise to do" (see also Gibbard, 2008; Kolodny and MacFarlane, 2010). The sense of objective *ought* that emerges from this solution to the puzzle is the one from a God's-eye perspective: as in the case of an omniscient being who could already see the actual development of the world, the truth value of future-dependent objective oughts is already determined depending on whether the relevant propositions are true or false on the thin red line.

This solution to the futurity problem has, however, some shortcomings. Firstly, Belnap and Green (1994) and MacFarlane (2003) have extensively argued against the commitment of a thin red line. The standard objection is that the stipulation that there exists one, privileged thin red line amongst the possible histories makes unclear in which sense the other possible histories represent objective, ontological possibilities. If now it is a fact that the coin will land tails, then, in a certain sense, that the coin will land tails is already determined. Hence, the alternative histories might at most be epistemically possible (it is a fact that coin will land tails, but we cannot know it yet), but not possible *in re*. If this objection succeeds, then the thin red line solution to the puzzle about futurity does not meet the third desideratum of neutrality. The second shortcoming, or at least limitation, concerns the applicability of an objectivist account of oughts defined from the God's-eye perspective. As in the Peircean interpretation, this account also does not provide any satisfactory objectivist standard for deontic reasoning under indeterminacy: to the contrary, issues related to risk are completely left out from a thin red line objectivist approach.

4.4.3 Future-Dependent Objective Oughts and Objective Probabilities

Let us take a preliminary stock. So far, we have discussed two sorts of solutions to the futurity problem: rejecting an objectivist interpretation of future-dependent oughts or interpreting objective future-dependent oughts as depending on future facts. What those future facts are is closely related to the temporal interpretation adopted: under a Peircean interpretation, future facts are those (non-epistemic) propositions that are true in all future branches; under the thin red line interpretation, future facts are those (non-epistemic) propositions that are true along the one branch representing the actual future.

Both sorts of solutions maintain FACTS as the definition of objective oughts: objective oughts are oughts directly relative to facts. The following solution to the futurity problem takes a different approach, and departs from FACTS by providing a probabilistic account of objective oughts. Such departure is, however, less radical than it seems: probabilities, as it will be shown, play a non-trivial role only for those objective oughts that are future-dependent. Understanding future-dependent objective oughts primarily in terms of probabilities about facts, rather than simply facts, allows us to considerably widen the objectivist perspective: not only does the probabilistic account provide a solution to the futurity problem, but it also sets an objectivist standard for deontic reasoning under indeterminacy and risk.

In what follows, we introduce the probabilistic account of objective oughts and discuss its main characteristics and implications for ethical theory and deontic logic. A formal semantics is presented in the Appendix.

4.4.4 A Probabilistic Proposal

The probabilistic account is defined as follows:

Definition 4.3. *At time t it is true that agent α* **ought objectively** *X if and only if X-ing is best in light of the objective chances of the relevant facts to obtain.*

Where, again, facts are non-epistemic propositions that are settled true. Let us call the above definition PROBABILITIES. It provides a univocal definition of objective oughts that maintains the connection between objective oughts and facts, but such connection is now mediated: objective oughts are functions of the *objective chances* of the relevant facts. Moreover, the probabilistic account qualifies as objectivist according to the general schema [GS] presented on p.67, given the standard distinction between objective probabilities and subjective probabilities (where the latter but not the former express an agent's degrees of credence). What an agent objectively ought to do is *not* defined in terms of an epistemic feature of the agent.

However, the probabilistic account needs some qualifications. There are several possible ways in which PROBABILITIES can be implemented, particularly for what concerns its interaction with temporal frameworks. We describe one of those possible ways, and argue that it provides an adequate solution to the futurity problem.

For continuity of exposition, we propose to understand PROBABILITIES against the sort of branching temporal background that was introduced in Sec.4.4.2 (see Def.4.2). Specifically, we take basic semantic structures to be *future branching trees* as in the one depicted in Fig.4.2 below. The intended

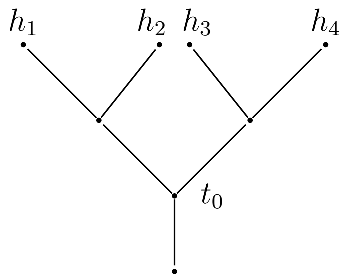


Figure 4.2: Future Branching Tree

interpretation is that a future branching tree represents the temporal evolution of the world *from the perspective of the current moment* t_0: while the past is already determined and — therefore — linear, the future may be open and branch into several objective possibilities.[11]

What we have said in Sec.4.4.2 about objective branching trees and future contingency holds here as well: objective branching trees, and in particular future branching trees, are neutral with respect to future contingency. Linear time is nothing but a future branching tree with just one future branch.

Semantically, we adopt a fine-grained approach and evaluate sentences not simply at moments, but rather at moment/history pairs t/h, indicating what is true at the moment t with respect to the temporal evolution marked out by the history h. Thus, for instance, $\ulcorner Will\varphi \urcorner$ is true at t/h if and only if φ is true at a future moment of t along the history h. This fine-grained approach is

[11] Note that "possible pasts" are excluded from future branching trees. The reason is that with the passing of time, the tree loses some branches: among the different possibilities that were open at a previous time, one is realized while the others are ruled out. From the perspective of the current moment, those possible pasts constitute counterfactual possibilities (i.e., ways the world could have evolved), but do not represent objective possibilities anymore.

commonly referred to as Ockhamist (Prior, 1967; Thomason, 1984; Horty and Belnap, 1995; Horty, 2001).

Taking moment/history pairs as basic points of evaluation allows for a certain expressibility. Firstly, it allows us to recover the Peircean interpretation of facts about the future. If we think of *settleness* at a certain moment t in terms of what holds at t/h for all histories h passing through t, then at t facts about the future are those (non-epistemic) propositions that are true in all future branches of t (see Thomason, 1984).

Secondly, and most importantly, it allows us to discuss not only what the future facts are, but also what the future facts *could* be. In other words, it permits us to consider also those facts about the future that could obtain. Imagine, for instance, that at t there is the risk your house will catch fire. This means that according to some — but not all — histories passing through t your house will catch fire. We say that, conditional on those histories, it is a fact that your house will catch fire. In other words, the fact that your house will catch fire could obtain were the future branching tree constituted of solely those histories in which your house will catch fire.

Why is the talk about future facts that could obtain relevant? The answer is easily seen in relation to PROBABILITIES. By referring to the objective chances of the relevant facts to obtain, the probabilistic account allows one to deal with oughts that depend on a chancy future: the truth of "You ought to get fire insurance for your house" depends precisely on the objective probability of the fact that your house will catch fire —that is, on the objective probability that the world could evolve along those histories in which your house will catch fire. Contrary to the other approaches considered above, PROBABILITIES therefore sets an objectivist standard for reasoning under indeterminacy and risk.

4.4.5 Evaluating the Proposal

Before moving to the implications triggered by the probabilistic account, let us evaluate whether it provides a satisfying solution to the futurity problem. Recall our three desiderata:

- *Conservativity.* Let us start by noticing that PROBABILITIES maintains indeed an objectivist character, that is, it fits the general schema [GS] presented at p.67. Since F is analyzed in terms of objective (i.e., non-epistemic) probabilities, the probabilistic account satisfies condition

(i) of conservativity. Condition (ii) of conservativity recommends that the probabilistic account at most extend FACTS to deal with future-dependent oughts. This second condition is met if we pick a probability function that plays a trivial role (assigning probability 1) for present and past facts. Such a probability function is formally defined in the Appendix. To the extent that the probability function works as expected, PROBABILITIES meets also the second condition of conservativity: PROBABILITIES extends FACTS by providing an account of objective oughts in which the connection between oughts and facts is mediated by objective probabilities, but at the same time by trivializing such mediation for facts about the present and the past.

- *Univocality.* PROBABILITIES provides a univocal definition of objective oughts in terms of objective probabilities, be the oughts dependent on the past, present or future.
- *Neutrality about the Future.* The semantic background of branching trees is not structurally committed on whether the future is contingent or not.

Hence, PROBABILITIES provides a third, alternative solution to the futurity problem.

4.5 Reasoning by Cases Under Indeterminacy

Contrary to the previous solutions to the futurity problem, PROBABILITIES suggests an objective standard for decision making under indeterminacy and risk: the objective probabilities of the relevant facts are what objectivists should consider when determining what ought to be done. This means that for the truth of sentences like "You ought to get fire insurance for your house" and "You should bet on tails", the objective probability that your house will catch fire or the objective probability that the coin will land on tails are relevant features of the world to be taken into account.

It should be noted that this approach does not guarantee future-dependent objective oughts to hold *retrospectively*, so to speak. Suppose that at t_0, given the objective chance of the relevant future circumstances C, an agent α ought objectively to X. However, even a high objective chance of C does not guarantee that such C will actually be the case in the future. Therefore it could be the case that, given the objective chance that C will happen tomorrow, the agent objectively ought to do X now, but that tomorrow it is false that the agent objectively ought to have done X. There is no contradiction, however, between "An agent α objectively ought to do X" true at t_0, and "The agent α objectively ought to have done X" false at t_1. The context has changed.

4.5.1 The Betting Puzzle

We have seen that the Miners' Puzzle does not affect objective oughts. Does that mean that objective oughts always validate Reasoning by Cases? No, at least if PROBABILITIES is adopted. To show this point, we bring to attention the following **betting puzzle**. A similar scenario can already be found in (Horty, 2001, p.53); the difference is that future-dependent objective oughts are the focus of the betting puzzle presented here.

Consider the following betting scenario:

> I offer you a bet. I will toss a fair coin (hence 0.5 objective chance that the coin will land tails and 0.5 objective chance that the coin will land heads) and ask you to guess on which side the coin will land. If your guess is correct, you win 150$. However, entering in the bet costs you 90$.

Table 4.2 represents the decision problem at t_0:

	Will Tails	Will Heads
Bet Tails	60	-90
Bet Heads	-90	60
Bet nothing	0	0

Table 4.2: Decision Problem for the Betting Scenario

The decision problem for the Betting Puzzle is structurally identical to the one in the Miners' Puzzle. Crucially, in terms of maximization of (objective) expected utility, the following sentences all appear to be acceptable:[12]

8. The coin will land tails or the coin will land heads
9. If the coin will land tails, you should bet on tails

[12] Contrary to what happens in the Miners' Puzzle, the rule MaxiMin is not directly applicable to the Betting Puzzle, at least not to the extent that PROBABILITIES is adopted. In fact, probabilities (be they objective or subjective) do not play any role in MaxiMin. However, MaxiMin would be compatible with the following modification of PROBABILITIES: At time t it is true that agent α ought objectively to X if and only if X-ing is best in light of the *objective possibilities, or objective chances* of the relevant facts to obtain.

10. If the coin will land heads, you should bet on heads

11. You should not bet

While, by Reasoning by Cases, it would follow:

12. You should bet on tails or you should bet on heads

which contradicts (11).

If the Miners' Puzzle is taken to show that Reasoning by Cases is invalid for belief/information dependent oughts, the same goes for the Betting Puzzle with respect to the objective oughts of PROBABILITIES.The crucial difference between the two, the Miners' Puzzle and the Betting Puzzle, is that the former concerns uncertainty, while the latter involves indeterminacy.

4.6 Conclusion

The standard definition of objective oughts as oughts dependent on relevant facts leaves future-dependent oughts unexplained. We have investigated three main strategies to account for future-dependent objective oughts. In particular, we have shown that there is a probabilistic interpretation of future-dependent objective oughts that meets some natural desiderata and provides an objectivist standard for reasoning under indeterminacy. Finally, we have argued that Reasoning by Cases is not a valid inference rule for objective oughts in scenarios of indeterminacy.

4.7 Appendix: A Formal Semantics

Let us begin by defining probability trees and future branching frames.[13]

Definition 4.4. *A* **probability tree** *is a tuple* $\mathcal{E}=\langle \mathbf{T}, S, Pr\rangle$*, where* $\mathbf{T}=\langle T, \prec \rangle$ *is a finite tree,* S *is an algebra over the set of histories in* $\mathbf{T}$*, and* Pr *is a function assigning to each element of* S *a number in [0,1] satisfying the following:*

- $Pr(\mathbf{T})=1$
- $Pr(\mathbf{T}' \cup \mathbf{T}'')= Pr(\mathbf{T}')+Pr(\mathbf{T}'')$ *if* $\mathbf{T}'$ *and* $\mathbf{T}''$ *are disjoint elements of* S.

[13] For simplicity, we are limiting the exposition to the case of $\mathbf{T}$ finite. Def.4.4 above can be easily extended for $\mathbf{T}$ infinite by considering σ-algebras, closed under *countable* union, instead of algebras. (See Halpern, 2003, p. 15). Moreover, for simplicity, we will identify a tree $\mathbf{T}$ with the set of its histories $Hist(\mathbf{T})$ whenever disambiguation is clear by context.

In other words, the probability function Pr assigns to those sets of histories in $\mathbf{T}$ that are in S their objective chance to be realized. Note that this does *not* imply that all maximal subtrees of $\mathbf{T}$ are assigned objective chances. In fact, there might be combinations of courses of events for which no objective chance can be determined.

Definition 4.5. *A* **forward branching frame** *is a tuple* $\mathcal{F}=\langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0\rangle$, *where* $\mathcal{E}^{t_0}=\langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}\rangle$ *is a probability tree, and* t_0 *is the moment "now" such that* $t_0 \in h'$*, for all histories* h' *in* $\mathbf{T}^{t_0}$*.*

Definition 4.6. *A* **forward branching model** *is a tuple* $\mathcal{M} = \langle \mathcal{F}, v\rangle$ *with* $\mathcal{F}=\langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0\rangle$ *a forward branching frame. The valuation function* v *maps each sentence letter from the background language into a set of* t/h *pairs (with* $t \in h$*) from* $\mathbf{T}^{t_0}$*.* **Truth** *at* t/h *is defined as follows:*

- $\mathcal{M}, t/h \models p$ *iff* $t/h \in v(p)$ *for* p *atomic*
- $\mathcal{M}, t/h \models \neg\varphi$ *iff* $\mathcal{M}, t/h \not\models \varphi$
- $\mathcal{M}, t/h \models \varphi \wedge \psi$ *iff* $\mathcal{M}, t/h \models \varphi$ *and* $\mathcal{M}, t/h \models \psi$
- $\mathcal{M}, t/h \models Will\varphi$ *iff there is a* $t' \in h$ *such that* $t \prec t'$ *and* $\mathcal{M}, t'/h \models \varphi$
- $\mathcal{M}, t/h \models Past\varphi$ *iff there is a* $t' \in h$ *such that* $t' \prec t$ *and* $\mathcal{M}, t'/h \models \varphi$

A formula is **settled** *true at* t *if it is true at* t/h *for all histories* h *passing through* t*.*

Definition 4.7. *Consider a forward branching model* $\mathcal{M} = \langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0, v\rangle$*. Let* φ *be a sentence of the background language. The* **proposition** *expressed by* φ *in* $\mathcal{M}$ *at* t*, written* $\mathbf{T}_{\varphi}^{\mathcal{M},t}$*, is defined as follows:*

- $\mathbf{T}_{\varphi}^{\mathcal{M},t}=\bigcup\{h|\mathcal{M}, t/h \models \varphi\}$

For simplicity, we write $\mathbf{T}_{\varphi}^{t}$ instead of $\mathbf{T}_{\varphi}^{\mathcal{M},t}$ — hence leaving the reference to $\mathcal{M}$ implicit.

From the above definitions it follows that the probability of those propositions that are settled true at t_0 is trivially 1 (since the proposition expressed by formulas that are settled true at t_0 is $\mathbf{T}^{t_0}$ and $Pr^{t_0}(\mathbf{T}^{t_0}) = 1$).

We turn to the semantics of objective oughts. We use the operator $\mathcal{O}$ to indicate objective oughts.

Definition 4.8. *A* **deontic model** $\mathcal{M}=\langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0, v, d\rangle$ *is a forward branching model plus a deontic selection function* d *that maps each probability* *tree* $\mathcal{E}'=\langle \mathbf{T}', S', Pr' \rangle$ *to a set of histories of* $\mathbf{T}'$ *that are deontically best — where* $\mathbf{T}' \subseteq \mathbf{T}^{t_0}$ *is a subtree given by some histories of* $\mathbf{T}^{t_0}$, $S' = \{x \cap Hist(\mathbf{T}') \mid x \in S^{t_0}\}$, *and* $Pr' = Pr^{t_0}(\cdot|\mathbf{T}')$.

Definition 4.9. *Let* $\mathcal{M}=\langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0, v, d \rangle$ *be a deontic model. Moreover, let* $\mathbf{T}^t$ *be the subtree given by all histories of* $\mathbf{T}^{t_0}$ *passing through* t, $S^t = \{x \cap Hist(\mathbf{T}^t) \mid x \in S^{t_0}\}$, $Pr^t = Pr^{t_0}(\cdot|\mathbf{T}^t)$ *and finally* $\mathcal{E}^t=\langle \mathbf{T}^t, S^t, Pr^t \rangle$. *The* **truth** *of an* **objective ought** *at* t/h *is defined as follows:*

- $\mathcal{M}, t/h \models \mathcal{O}\varphi$ *iff* $\mathcal{M}, t/h' \models \varphi$ *for every* $h' \in d(\mathcal{E}^t)$

Conditional objective oughts can be be derived from Def.4.9 by requiring Pr^{t_0} to be normalized with respect to the subtree satisfying the antecedent of the conditional and shifting d to that subtree (Kolodny and MacFarlane, 2010; Willer, 2012; Yalcin, 2010; Carr, 2015).

Definition 4.10. *Let* $\mathcal{M} = \langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0}, t_0, v, d \rangle$ *be a deontic model. Moreover, let* $\mathbf{T}^t_\psi$ *be the proposition expressed by* ψ *in* $\mathcal{M}$ *at* t, $S^t_\psi = \{x \cap Hist(\mathbf{T}^t_\psi) \mid x \in S^{t_0}\}$, $Pr^t_\psi = Pr^{t_0}(\cdot|\mathbf{T}^t_\psi)$, *and* $\mathcal{E}^t_\psi=\langle \mathbf{T}^t_\psi, S^t_\psi, Pr^t_\psi \rangle$. *The truth of a* **conditional objective ought** *at* t/h *is defined as follows:*

- $\mathcal{M}, t/h \models \psi \Rightarrow \mathcal{O}\varphi$ *iff* $\mathcal{M}, t/h' \models \varphi$ *for every* $h' \in d(\mathcal{E}^t_\psi)$

In other words: $\ulcorner \psi \Rightarrow \mathcal{O}\varphi \urcorner$ is true iff $\ulcorner \mathcal{O}\varphi \urcorner$ is true with respect to the subtree generated by assuming ψ and the objective chances conditionalized to such subtree.

Betting Puzzle. We construct a tree-like deontic model $\mathcal{M}$ of the betting scenario. Suppose you are currently considering what to bet: the relevant decision tree features six different histories (as six are the combinations between possible ways of betting and the outcome of the toss). The scenario is represented in Fig.4.3, where T and H stand respectively for "The coin lands tails" and "The coin lands heads", while $\mathsf{Bt}/\mathsf{Bh}/\mathsf{Bn}$ stand for "You bet on tails/heads/nothing".

We show now that all acceptable sentences of the Betting Puzzle are true at t_0 in each history h of the above model. Firstly we note that for each history h it holds that $\mathcal{M}, t_0/h \models Will\mathsf{T} \vee Will\mathsf{H}$.

Moreover, we define the deontic selection function d as follows (implicitly reflecting the rule of maximization of (objectively) expected utility):

For $\mathcal{E}^{t_0}_{\text{will T}} = \langle \mathbf{T}^{t_0}_{\text{will T}}, S^{t_0}_{\text{will T}}, Pr^{t_0}_{\text{will T}} \rangle$, $d(\mathcal{E}^{t_0}_{\text{will T}}) = \{h' \in \mathbf{T}^{t_0}_{\text{will T}} \mid \mathcal{M}, t_0/h' \models \mathsf{Bt}\}$. That is, under the supposition that the coin will land tails, the deontic

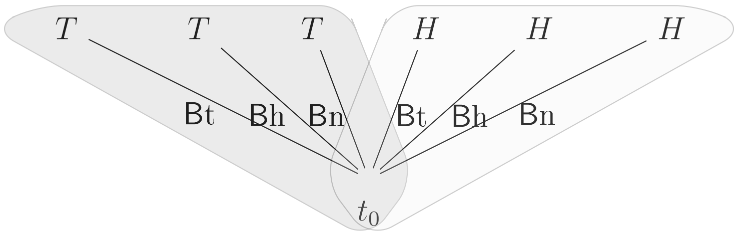


Figure 4.3: Tree-like deontic model $\mathcal{M}$ for the Betting Scenario. Dark grey indicates the subtree $\mathbf{T}^{t_0}_{\text{will T}}$ in which $Will\mathsf{T}$ is settled true; Light grey indicates the subtree $\mathbf{T}^{t_0}_{\text{will H}}$ in which $Will\mathsf{H}$ is settled true. It is assumed that the objective chances for $\mathbf{T}^{t_0}_{\text{will T}}$ and $\mathbf{T}^{t_0}_{\text{will H}}$ to obtain are 0.5 respectively.

function would recommend to bet on tails.

Similarly, for $\mathcal{E}^{t_0}_{\text{will H}} = \langle \mathbf{T}^{t_0}_{\text{will H}}, S^{t_0}_{\text{will H}}, Pr^{t_0}_{\text{will H}} \rangle$, $d(\mathcal{E}^{t_0}_{\text{will H}}) = \{h' \in \mathbf{T}^{t_0}_{\text{will H}} \mid \mathcal{M}, t_0/h' \models \mathsf{Bh}\}$. Similarly to the previous case: under the supposition that the coin will land heads, the deontic function would recommend to bet on heads.

On the other hand, for $\mathcal{E}^{t_0} = \langle \mathbf{T}^{t_0}, S^{t_0}, Pr^{t_0} \rangle$, $d(\mathcal{E}^{t_0}) = \{h' \in \mathbf{T}^{t_0} \mid \mathcal{M}, t_0/h' \models \mathsf{Bn}\}$. Indeed, given the global context described by Table 4.2, betting nothing is the action that maximizes expected utility.

By semantic definition, it follows that (for all histories h in the model):
$\mathcal{M}, t_0/h \models Will\mathsf{T} \vee Will\mathsf{H}$
$\mathcal{M}, t_0/h \models Will\mathsf{T} \Rightarrow \mathcal{O}\mathsf{Bt}$
$\mathcal{M}, t_0/h \models Will\mathsf{H} \Rightarrow \mathcal{O}\mathsf{Bh}$
$\mathcal{M}, t_0/h \models \mathcal{O}\mathsf{Bn}$.

Thus, the above model validates all acceptable sentences of the Betting Puzzle. The unacceptable conclusion, however, does not hold:
$\mathcal{M}, t_0/h \not\models \mathcal{O}\mathsf{Bt} \vee \mathcal{O}\mathsf{Bh}$

Hence the Betting Puzzle is solved.

Comparison with the Gambling Problem. We would like to conclude by comparing the Betting Puzzle discussed in this chapter with a similar one proposed by Horty (2001), and known as *the gambling problem*. There are two relevant differences. The first difference between the two puzzles is conceptual. The notion of future-dependent oughts is central to the Betting Puzzle, while in the gambling problem such a notion remains implicit. The second difference is formal, and concerns how these puzzles are solved. We have seen that the semantics provided here solves the Betting Puzzle by both (i) blocking the unacceptable conclusion that would follow via Reasoning by Cases (i.e., that you should bet on tails or you should bet on heads) and (ii) validating the acceptable judgment that you should not bet. It is the latter aspect of the solution that is not present in Horty (2001). In fact, the deontic framework adopted by Horty (2001) would *always* falsify sentences like "You should not bet", or "You ought to block neither shalf" — no matter what the stakes (in terms of monetary gains or lives saved) are in the respective scenarios. More generally, Horty's (2001) deontic framework appears to have certain built-in (decision-theoretic) commitments that make it impossible to derive an ought that contradicts the ought that would follow via Reasoning by Cases. This observation is the content of the following lemma.

Let $\odot$ and $[i\ cstit]$ denote Horty's (2001) deontic operator and stit operator. For the relevant definitions, we refer the reader to (Horty, 2001, p.77).[14]

Lemma 4.11. *In utilitarian stit models, the following formula is valid:*

$$(A \vee B) \wedge \odot([i\ cstit\varphi_1]/A) \wedge \odot([i\ cstit\varphi_2]/B) \supset \neg \odot [i\ cstit(\neg\varphi_1 \wedge \neg\varphi_2)]$$

Proof. We sketch the proof of a more general result, namely that:

$$(A \supset C) \wedge \odot([i\ cstit\varphi]/A) \supset \neg \odot ([i\ cstit\neg\varphi]/C)$$

The original lemma follows from the above (with C := $A \vee B$). The proof proceeds by contradiction.

Assume, by absurdum, that there is a utilitarian stit model M and index m/h in which it is true that $(A \supset C) \wedge \odot([i\ cstit\varphi]/A) \wedge \odot([i\ cstit\neg\varphi]/C)$.

1. From $\odot([i\ cstit\varphi]/A)$ it follows that, when restricting to A-histories (i.e, all histories h' such that A is true at m/h'), all non-strongly dominated actions contain only φ-histories. Moreover there is at least one such action K.

2. From $([i\ cstit\neg\varphi]/C)$ it follows that, when restricting to C-histories, all non-strongly dominated actions contain only $\neg\varphi$-histories.

3. Since $(A \supset C)$, the set of all C-histories is a superset of the set of all A-histories.

[14] Lemma 4.11 resulted from a collaboration between me, Soroush Rafiee Rad and Dominik Klein.

4. From 1 and 2, it follows that K is strongly dominated in the set of C-histories, i.e., there is an action K' strongly dominating K on the set of all C-histories and such that it contains only $\neg\varphi$-histories.
5. By 3 and 4, K' weakly dominates K on the set of A-histories.
6. In particular, there is a non-strongly dominated action on the set of A-histories which (only) contains $\neg\varphi$-histories. This contradicts 1.

□

Notes

- In the previous version of the manuscript, the forward branching frame in Def.4.5 was defined as "$\mathcal{F}=\langle \mathbf{T}, S, Pr^{t_0}, t_0\rangle$, where Pr^{t_0} is Pr normalized to that unique subtree that has t_0 as now" . Such definition has been corrected, and uniformly changed.
- In the previous version of the manuscript, Def.4.8 did not specify $\mathbf{T}', S'$ and Pr'. This has been uniformly changed. I thank Allard Tamminga for urging me to clarify this.
- The model $\mathcal{M}$ at page 85 was mistakenly referred to with the letter D. This has been uniformly corrected.

Chapter 5

When Is Reasoning by Cases Valid?

This chapter brings together the counterexamples to Reasoning by Cases that have been recently proposed, amongst others, by Kolodny and MacFarlane (2010), Bledin (2014) and Carr (2015). The counterexamples all involve indicative conditionals whose consequents are epistemically, deontically or probabilistically modalized. We provide a unified explanation of why these counterexamples emerge, and put forward a sufficient criterion for the validity of Reasoning by Cases. Specifically, we first show that the failure of Reasoning by Cases is not due to the presence of modals in the consequents of the indicative conditionals. Finally, we argue that Reasoning by Cases is valid when it involves consequents that have a certain property, namely upward local persistence.

5.1 Introduction

This chapter is about the deductive inference rule of *Reasoning by Cases* and the scope of its validity.We say the scope of its validity because the rule, which goes from the premises ⌜φ_1 or φ_2⌝, ⌜if φ_1, then ψ_1⌝ and ⌜if φ_2, then ψ_2⌝ to the conclusion ⌜ψ_1 or ψ_2⌝, has recently been the subject of much dissent. Perhaps the most famous attack comes from the so-called *Miners' Puzzle*, brought to attention by Kolodny and MacFarlane (2010), which shows that Reasoning by Cases is invalid when it involves indicative conditionals whose consequents contain deontic modals. The Miners' Puzzle is not an isolate case, however. Another counterexample, this time featuring indicative conditionals with epistemic modals in the consequents, has been proposed by Bledin (2014). Moreover, a third counterexample with indicative conditionals involving probabilistic modals in the consequents can be found, for instance, in Carr (2015).

The primary aim of this chapter is to provide a general account that could not only illustrate *that* the interaction between indicative conditionals and deontic, epistemic and probabilistic modals gives rise to the above counterexamples to Reasoning by Cases, but also determine *when* Reasoning by Cases is a valid inference rule. The account we defend here contributes to the literature in at least two ways. First, it is general. Albeit the above counterexamples have generated much recent debate, the literature so far has mainly focused on each of those counterexamples taken in isolation. The starting point of this chapter is instead the three counterexamples mentioned above, taken together. This general approach can highlight the commonalities that the counterexamples share and that would be overlooked had we considered only one single counterexample. Moreover, considering the three counterexamples together permits us to have a broader view on the distinctive behavior of indicative conditionals and modals. A second contribution to the literature consists in providing a sufficient criterion for the validity of Reasoning by Cases that improves on the alternative accounts that can be found in Cantwell (2008) and Bledin (2014). In particular, the criterion defended here extends and generalizes the one presented by Bledin (2014). If the criterion we propose is on the right track, then it is good news for the supporters of classical reasoning, as it vindicates the use of Reasoning by Cases in a wide variety of contexts.

The chapter proceeds as follows. The next section introduces the three counterexamples mentioned above. The goal of the section is not to provide a full-fletched defense of those counterexamples, but rather to discuss what those counterexamples have in common. Specifically, the counterexamples all feature deontic, epistemic or probability modals that take narrow scope under indicative conditionals. Section 5.3 shows that, contrary to what might be expected, the invalidity of Reasoning by Cases is not strictly speaking due to the interaction between modals and indicative conditionals. To prove this point, we defend a novel counterexample which does not involve any modal but only indicative conditionals. Section 5.4 constitutes the positive contribution of this chapter: it brings together the observations of the previous sections, and puts forward a criterion sufficient to determine the scope of the validity of Reasoning by Cases. The criterion is based on the formal notion of local persistence throughout contexts, and is expressed using the formal machinery of the so-called *informational logic*. It is shown that the counterexamples considered so far all violate the criterion, while at the same time the criterion allows for a wide range of instances of Reasoning by Cases (e.g., those involving atomic sentences and boolean combination thereof, and those involving the epistemic modal *might*). The chapter concludes by comparing the sufficient criterion defended here to the alternative proposals in the literature.

Let us now begin with a closer look at the counterexamples to Reasoning by Cases.

5.2 Three Modal Counterexamples

5.2.1 The One with Deontic Modals

We begin with the counterexample to Reasoning by Cases from Kolodny and MacFarlane (2010). The counterexample involves deontic modals, and it is mostly known as the *Miners' Puzzle*. Recall the background scenario (Kolodny and MacFarlane, 2010, p.115): Ten miners are either in shaft A or in shaft B, but we do not know which. Their lives are at stake, and we can block at most one of the shafts. If we block the shaft the miners are in, all ten miners live. If we block the shaft the miners are not in, all ten miners will be killed. If we block neither shaft, one miner will be killed. What should you do? Given the scenario and the information you possess, the intuitive answer is: you ought to block neither of the shafts, as this guarantees that nine miners are saved.[1]

[1] Two remarks are in place here. First remark: there are two available readings of the sentence "You ought to block neither shaft". The strong reading corresponds to the deontic modal *ought* scoping over negation, and the weaker one to the negation scoping over *ought*. Roughly speaking, the strong reading of "You ought to block neither shaft" indicates that you have the obligation of refraining from blocking either shaft, while the weak reading indicates that you do not have the obligation of blocking shaft A and you do not have the obligation of blocking shaft B. This chapter follows Willer (2012); Carr (2015); Cariani et al. (2013); Bledin (2015), and adopts the stronger reading. However whichever reading is adopted does not make any difference to the issues under discussion here. Both readings generate a counterexample as, in both cases, the sentence "You ought to block neither shaft" turns out to be incompatible with the conclusion derived by Reasoning by Cases. Second remark: as discussed by Carr (2015), the acceptability of the sentence "You ought to block neither shaft" is sensitive to the changes in the scenario's figures. Understanding which decision rules guide our moral judgments in miners-like scenarios is a project worth pursuing, but remains an intrinsically distinct problem with respect to the one we are concerned about here (cf. Cariani et al., 2013). The problem this chapter addresses is strictly *logical* (i.e., the scope of validity of Reasoning by Cases), and not *decision-theoretic*. What matters for the logical problem is that there exists a scenario in which what would logically follow via Reasoning by Cases is not acceptable.

Simultaneously, here is one way you can reason about the situation:

1	The miners are in A or they are in B
2	If they are in A, I ought to block A
3	If they are in B, I ought to block B
4	I ought to block A or I ought to block B

The premises of the above argument are all acceptable: in the scenario it holds that the miners are in shaft A or they are in shaft B; moreover, under the supposition that the miners are in a specific shaft (in turn, shaft A or shaft B), the right thing to do is blocking that shaft, as this allows all ten miners to be saved. From those premises, however, via Reasoning by Cases it follows that you ought to block A or you ought to block B. And here is the paradoxical nature of the scenario: this conclusion contradicts what we have agreed on, namely that you ought to block neither of the shafts.

We can now see why the Miners' Puzzle constitutes a counterexample to Reasoning by Cases: it shows that there are instances of such an inference rule that move from acceptable premises to an unacceptable conclusion. This particular instance has the following form:

1	$p_1 \vee p_2$
2	$p_1 \Rightarrow \text{OUGHT}\ q_1$
3	$p_2 \Rightarrow \text{OUGHT}\ q_2$
4	$\text{OUGHT}\ q_1 \vee \text{OUGHT}\ q_2$

We follow the standard practice of using $\vee$ and $\Rightarrow$ to indicate disjunction and indicative conditional, respectively, and the sentential operator OUGHT for the natural language deontic modal *ought*.

Let us have a closer look at the logical form of the counterexample, and in particular at its premises. Premise 1 consists of a disjunction of atomic sentences, while premises 2 and 3 are conditional sentences whose consequents are deontically modalized. There is one important assumption on which the counterexample relies, and that we would like to stress at this point: the assumption that the logical form of the argument matches its surface form, and in particular that the the deontic modal *ought* takes *narrow-scope* under the indicative conditional. This assumption can be questioned, and perhaps the most common way of doing this consists in endorsing a wide-scope analysis of *ought*. According to such analysis, in premises 2 and 3 the operator OUGHT scopes over the indicative conditional, giving rise to the following, non-trivial logical form:

1	$p_1 \vee p_2$
2	OUGHT $(p_1 \Rightarrow q_1)$
3	OUGHT $(p_2 \Rightarrow q_2)$
4	OUGHT $q_1 \vee$ OUGHT q_2

Formalized in this way, the argument would not constitute a counterexample to Reasoning by Cases because, clearly, the argument would not be a genuine instance of the deductive inference rule. Criticisms against the general plausibility of the wide-scope analysis of *ought* have been raised by Silk (2014b), and further arguments against a sort of "wide-scoping solution" to the Miners' Paradox can be found in Kolodny and MacFarlane (2010) and Charlow (2013). We will come back to the issue of wide-scoping in Sec.5.3. For the moment, it is sufficient to point out that the counterexample to Reasoning by Cases generated by the Miners' Puzzle relies on the availability of a narrow-scope reading for the deontic modal *ought*.

5.2.2 The One with Epistemic Modals

A second counterexample to Reasoning by Cases can be found in Bledin (2014). This time the counterexample involves epistemic modals — specifically the epistemic modal *must* — in the consequents of indicative conditionals. Before introducing the counterexample, it is better to spend a few words on the epistemic reading of the modal *must* and its dual *might*. Under the epistemic reading, the modals *must* and *might* relate to some piece of knowledge or information: *must p* indicates that the available knowledge/information warrant or makes necessary that p, while *might p* indicates that given the available knowledge/information it is possible that p (Kratzer, 2012d; Veltman, 1985, 1996; Yalcin, 2007; Willer, 2013, 2015). For instance, upon noticing that the lights in Roger's study are on and knowing that Roger always turns all lights off whenever he leaves his house, you can conclude that, in view of your information, Roger must be at home.[2]

Keeping fixed the interpretation discussed above, we can now turn to the counterexample involving epistemic modals. What follows is an adaptation

[2]Example from Kratzer (2012a).

from (Bledin, 2014, p.300). The scenario proceeds as follows: on the 8th of May, a seminar will take place at your university — a nice occasion to also talk to your supervisor about your new project or to talk about an upcoming concert with your colleagues. You know that the 8th of May is either Monday or Tuesday, but do not know which. Also, you know that if it is Monday then your supervisor Ann must be in her office, and that if it is Tuesday then your colleague Barbara must be in her office. What should you then expect on the 8th of May? Unfortunately, given the information available to you, it is possible that Ann is not in her office on the 8th of May (and so you will not have the chance to talk about your new project) and it is possible that Barbara is not in her office on the 8th of May (and so you will not have the chance to talk to her about the concert). In other words, Ann might not be in and Barbara might not be in. However, you can reason as follows:

1	It is Monday or it is Tuesday
2	If it is Monday, then Ann must be in
3	If it Tuesday, then Barbara must be in
4	Ann must be in or Barbara must be in

That is, via Reasoning by Cases you conclude that, in light of your information, Ann must be in or Barbara must be in. But this conclusion contradicts what we established: namely that, given your information, Ann might not be in her office and Barbara might not be in her office. After all, your information does not warrant that you will have the chance to talk about your project nor does it warrant that you will have the chance to talk about the concert.[3]

Hence, the scenario discussed here constitutes a second counterexample to Reasoning by Cases. It shows that there are instances of the following form that move from acceptable premises to an unacceptable conclusion:

1	$p_1 \vee p_2$
2	$p_1 \Rightarrow \text{MUST } q_1$
3	$p_2 \Rightarrow \text{MUST } q_2$
4	$\text{MUST } q_1 \vee \text{MUST } q_2$

[3]It warrants a weaker conclusion: namely that you will have the chance to talk about your project or the concert. Indeed, given your information, it must be the case that Ann is in or Barbara is in. As Bledin (2014) rightly points out, there is a difference of scope: in the weaker conclusion, *must* scopes over a disjunction, while in the conclusion of the main text *must* scopes under a disjunction.

This counterexample shares the same assumptions as the previous one: namely that the logical form of the above argument matches its surface form, and in particular that the sentential operator MUST takes narrow scope under the indicative conditional.[4]

5.2.3 The One with Probability Modals

Let us present the third counterexample that can be found in the literature. This counterexample features probability modals, such as *probably* or *likely*, in the consequents of indicative conditionals.

Just a few introductory words before delving into the counterexample. Probability modals such as *probably* or *likely* indicate what is probable relative to a certain body of information — the standard assumption being that, for φ to be probable, the likelihood of φ must be higher than 0.5 (Hamblin, 1959; Kratzer, 1991; Yalcin, 2007, 2010).[5] For instance, upon receiving the information that a coin is 0.7 biased towards heads up, you can conclude that the coin will probably land heads. Following Yalcin (2012b), we treat *probably* and *likely* as synonyms.

The counterexample introduced here is due to Carr (2015).[6] Here is the scenario:

> Suppose that it is either Monday or Tuesday, but you do not know which. And suppose the miners could be in shaft A, shaft

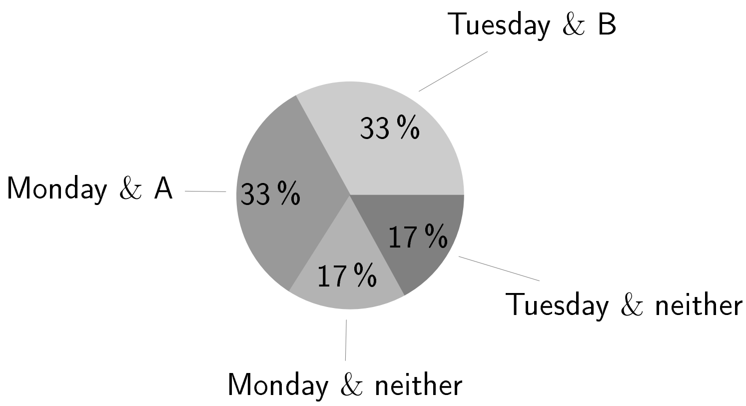
[4] It is perhaps also worth indicating what the counterexample does *not* assume: it does not take a stance on whether the epistemic modal *must* has to be treated as an indirect evidential or not (cf. Kratzer, 2012a; Veltman, 1985, 1996; Yalcin, 2007; von Fintel and Gillies, 2010; Willer, 2015). Generally, the indirect evidentiality camp does not support the inference from MUSTφ to φ, as the former is taken to express something weaker than the latter. To the contrary, the inference from MUSTφ to φ is taken to be unproblematic in Veltman (1996), Yalcin (2007) and Bledin (2014). The framework adopted in Sec.5.4, deriving from Bledin (2014), actually validates the inference from MUSTφ to φ, but nothing in our analysis relies on such inference.

[5] Fixing the threshold at 0.5 is motivated by the idea that φ is probably the case if and only if it is more likely than not. Although such an account of *probably* and *likely* is widely accepted, some concerns have been raised about having an absolute threshold fixed at 0.5. For instance, it is questionable whether a sentence like "The coin will probably land heads" is true when the coin is only 0.51 biased towards heads up. Others have proposed a relative, non-absolute threshold to account for the fact that sometimes φ is taken to be probable when it is more likely than each of its alternatives. For a critical discussion, see Yalcin (2010). However, since none of the above concerns directly applies to the counterexample discussed in this section, we stick with the standard 0.5 threshold.

[6] For another counterexample involving probability modals in the consequents of indicative conditionals, see Moss (2015).

> B, or neither shaft: the likelihood of each is $\frac{1}{3}$. The miners are in shaft A only if it is Monday, and in shaft B only if it is Tuesday. If it is Monday, it is $\frac{2}{3}$ likely that the miners are in shaft A. And if it is Tuesday, it is $\frac{2}{3}$ likely that they are in shaft B.
>
> (Carr, 2015, p.683)

This is all the information given about the situation. What follows is a visual representation of the breakdown.



You reason as follows:

1	It is Monday or it is Tuesday
2	If it is Monday, then the miners are likely in A
3	If it Tuesday, then the miners are likely in B
4	The miners are likely in A or the miners are likely in B

The premises of the above argument are all acceptable: in the scenario it holds that it is Monday or Tuesday; under the supposition that it is Monday, then it is likely that the miners are in A (as the likelihood of the miners being in A, provided it is Monday, is $\frac{2}{3}$) and, similarly, under the supposition that it is Tuesday, then it is likely that the miners are in B. However, the conclusion of the argument is not acceptable: given the available information, it does not hold that the miners are likely in A nor it holds that the miners are likely in B. It is equally probable that the miners are in A, that they are in B, or that they are in neither shaft.

Hence, reasoning by cases from the above premises leads to an unacceptable conclusion. Once again, if we assume that the logical form of the above argument corresponds to its surface form, and in particular that the sentential operator PROBABLY takes narrow scope under the indicative conditional, we are left with a counterexample of the following form:

$$
\begin{array}{r|l}
1 & p_1 \vee p_2 \\
2 & p_1 \Rightarrow \text{PROBABLY } q_1 \\
3 & p_2 \Rightarrow \text{PROBABLY } q_2 \\
\hline
4 & \text{PROBABLY } q_1 \vee \text{PROBABLY } q_2
\end{array}
$$

5.2.4 What in Common?

Let us take a stock. So far we have presented the three main counterexamples to Reasoning by Cases that can be found in the literature. These are cases in which an unacceptable conclusion follows from a set of acceptable premises. Which lesson can be drawn from those counterexamples? First of all, all the counterexamples reduce to the following form:

$$
\begin{array}{r|l}
1 & p_1 \vee p_2 \\
2 & p_1 \Rightarrow \text{MODAL } q_1 \\
3 & p_2 \Rightarrow \text{MODAL } q_2 \\
\hline
4 & \text{MODAL } q_1 \vee \text{MODAL } q_2
\end{array}
$$

Where MODAL stands for a deontic/epistemic/probabilistic modal, which takes narrow-scope under the indicative conditional. This commonality between the counterexamples naturally suggests that the invalidity of reasoning by cases is due — in a certain sense — to the interaction between deontic/epistemic/probabilistic modals and indicative conditionals. Indeed, Bledin (2015) recommends to "exercise caution" when using Reasoning by Cases in languages that contain those modals and indicative conditionals.[7] Applying the general form of Reasoning by Cases in such languages, he continues, can lead to absurd results.

Let us begin from the suggestion that Reasoning by Cases is invalid because it involves deontic/epistemic/probabilistic modals in the consequents of indicative conditionals. Putting it positively, this suggestion provides an indication

[7] To be precise, Bledin (2015) talks about "informational modals". Informational modals are the modals we refer to in the main text, namely the deontic/epistemic/probabilistic modals that are relative to a certain body of information. Hence, our change of wording should not make any substantial difference. More about the term "informational modals" will be said in Sec.5.4.

of the scope of validity of Reasoning by Cases, and leads us to the following hypothesis.

Take the general form of Reasoning by Cases:

1	$\varphi_1 \vee \varphi_2$
2	$\varphi_1 \Rightarrow \psi_1$
3	$\varphi_2 \Rightarrow \psi_2$
4	$\psi_1 \vee \psi_2$

If we consider exclusively the conditions to impose to the consequents ψ_1 and ψ_2 of the indicative conditionals, it holds that:

> **First Hypothesis**: If ψ_1 and ψ_2 are not deontically, epistemically nor probabilistically modalized, then Reasoning by Cases is valid.

The above hypothesis is persuasive: evidence in its favor is given by the fact that the only counterexamples we have encountered so far involved deontic/epistemic/probabilistic modals. However, such a hypothesis is not the final word on the matter. In particular, we suggest that it does not properly track the scope of validity of Reasoning by Cases: Reasoning by Cases may fail even in the absence of deontic/epistemic/probabilistic modals.

5.3 A Novel Counterexample: Just Conditionals

The goal of this section is to improve on the *First Hypothesis*. Specifically, we first show that the *First Hypothesis* is not correct by defending a counterexample to Reasoning by Cases that does not involve any deontic/ epistemic/ probabilistic modal. In light of this novel counterexample, we put forward a *Second Hypothesis*.

Here is the scenario: Imagine I offer you the chance to bet on a coin toss. Suppose you have accepted the bet, but have not yet announced which outcome you will bet upon. Hence I know that you will bet on tails or you will bet on heads, but do not know which. Before tossing the coin, I reason as follows:

> P1: You bet on tails or You bet on heads
>
> P2: If you bet on tails, then if the coin lands tails you win

P3: If you bet on heads, then if the coin lands heads you win

The above premises are all acceptable: P1 holds as you have accepted the bet, and premises P2 and P3 hold in virtue of the betting scenario. However, from these premises, the following conclusion does not intuitively follow:

C: If the coin lands tails you win, or, if the coin lands heads you win

After all, you could lose your bet. But the above conclusion would follow, were Reasoning by Cases applied.

To see that the scenario above constitutes a counterexample to Reasoning by Cases, let us schematically represent it as follows:[8]

1 | You bet on tails $\vee$ You bet on heads
2 | You bet on tails $\Rightarrow$ (The coin lands tails $\Rightarrow$ you win)
3 | You bet on heads $\Rightarrow$ (The coin lands heads $\Rightarrow$ you win)
4 | (The coin lands tails $\Rightarrow$ you win) $\vee$ (The coin lands heads $\Rightarrow$ you win)

With the above logical form in mind, let us now spend a few words more on why the conclusion is unacceptable. Recall that you have not yet announced which option you will bet upon. Given the available information, there are four live possibilities:

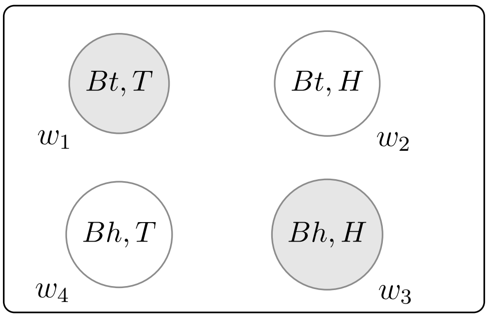


Figure 5.1: Open Possibilities in the Betting Scenario

Where Bt and Bh stand for *You bet on tails* and *You bet on heads*; while T and H stand for *The coin lands tails* and *The coin lands heads* respectively. The possibilities which lead you to win the bet are those colored in grey.

[8]There is an implicit futurity in the above scenario: you *will* bet, the coin *will* land and (possibly) you *will* win. For simplicity, we ignore this issue here.

The sentence “If the coin lands tails you win, or, if the coin lands heads you win” does not hold because, even if it is settled that the coin will land tails or heads, there is the possibility that you lose the bet, depending on whether you will bet on tails or heads. The given scenario does not rule out the possibility of your betting on heads and the coin landing tails, nor the possibility of your betting on tails and the coin landing heads. In those cases (i.e., w_2 and w_4), you would definitely lose your bet! In other words: it holds that if the coin lands tails you win *provided* you bet on tails, and that if the coin lands heads you win *provided* you bet on heads. But since it is left open on which one of the two options (heads or tails) you will bet, then neither “If the coin lands tails you win” nor “If the coin lands heads you win” hold unconditionally. Hence the argument moves from acceptable premises to an unacceptable conclusion.

Two observations are in order: first, the counterexample relies on the same assumptions of the counterexamples discussed in Sec.5.2, namely that the logical form of the natural language argument matches its surface form, and that the operator in the consequent of the indicative conditional has a narrow-scope reading. In the present case, that operator is itself an indicative conditional — so we are left with an invalid instance of Reasoning by Cases involving right-nested indicative conditionals. This brings us to the second observation: importantly, the present counterexample does *not* involve any deontic/epistemic/probabilistic modal. The *First Hypothesis* is therefore falsified. Reasoning by cases does not fail only when the consequents are epistemically/deontically/probabilistically modalized. The issue has a more general nature.

5.3.1 Two Objections and Replies

We consider two objections to the analysis so far presented.

1) *Reasoning by Cases is invalid under the narrow-scope reading. This shows we should reject the narrow-scope reading.*

This objection suggests that, given that the counterexamples we have considered so far all assumed a narrow-scope reading of the operators in the consequents, it is the narrow-scope reading which should be dismissed — not the principle of Reasoning by Cases. Consider the counterexample presented in this section. Under a wide-scope analysis, its non-trivial logical form is the following:

1	$p_1 \vee p_2$
2	$(p_1 \wedge q_1) \Rightarrow q_2$
3	$(p_2 \wedge q_3) \Rightarrow q_2$
4	$(q_1 \Rightarrow q_2) \vee (q_3 \Rightarrow q_2)$

Clearly, the above argument is not an instance of Reasoning by Cases. Hence even if the argument is invalid, Reasoning by Cases is not to blame.

There are two possible ways of articulating the above objection. One way, which we call "always wide-scope", recommends to uniformly reject narrow-scope. Another way, which we call "sometimes wide-scoping", argues that sometimes narrow-scope should be rejected. We consider these positions in turn.

An "always wide-scope" objection is under-motivated in the present context. Adopting "always wide-scope" would indeed save Reasoning by Cases from the counterexamples considered above, but at the price of blocking other intuitively good arguments. Consider the following argument:

1	You bet on heads
2	If you bet on heads, then if the coin lands heads you win
3	If the coin lands heads you win

The argument appears unobjectionable: it holds that, provided you have bet on heads, if the coin lands heads you win. If the "always wide-scope" were adopted, the above argument — which under the narrow-reading is an instance of Modus Ponens — would be rejected. Hence the motivation for "always wide-scope" is thin: the counterexamples to Reasoning by Cases are blocked, but so are intuitively good instances of Modus Ponens.[9]

The "sometimes wide-scoping" claim is weaker, as it does not amount to denying that the narrow-scope reading is ever available. Rather, it amounts to supporting the availability of wide-scope readings. As such, this position is compatible with the analysis presented here. It is compatible with granting, for instance, that the logical form of an argument should normally match its surface form. In order to constitute an objection against the narrow-scope

[9] The same line of response to wide-scope analyses is adopted by Yalcin (2012b) in defense of his counterexample to Modus Tollens.

reading of the counterexamples above, the "sometimes wide-scoping" theorist would need to argue that all those cases count as abnormal — so that a deviation from the surface form is justified. Following Yalcin (2012b) and Cantwell (2008), we are inclined to say that the burden of such proof is carried by the "sometimes wide-scoping" theorist. In the meantime, we hold that the counterexamples fall within the normal cases in which logical form and surface form match.

2) *Indicative conditionals have a modal flavor. So, the First Hypothesis was not too far off: It can be straightforwardly weakened by requiring modal-free consequents.*

The way indicative conditionals have so far been informally interpreted has indeed a modal component: conditionals like $\ulcorner\varphi \Rightarrow \psi\urcorner$ involve a *shift* to an hypothetical context in which φ holds — a shift that we have indicated using expressions like *under the supposition that* φ or *provided that* φ. This shifting interpretation traces back to (a semantic version of) the so-called "Ramsey Test", and is standardly adopted (Kratzer, 2012a; Veltman, 1996; Gillies, 2010; Yalcin, 2007; Willer, 2014). Granted that indicative conditionals hold a modal flavor as just described, the objection leads to the following hypothesis:

> **Second Hypothesis**: If ψ_1 and ψ_2 are not modalized, then Reasoning by Cases is valid.

The *Second Hypothesis* is akin to the analysis of Cantwell (2008), who notices that indicative conditionals which appear in the consequents of other indicative conditionals give rise to the same kind of logical phenomena (in particular, violations of classical logic principles) as other modalities.

In what follows, we show why and how the *Second Hypothesis* should be replaced. Although the *Second Hypothesis* is true, it is stronger than required.

5.4 Information-Sensitivity and Persistence

The discussion has been largely "pre-theoretical" so far: we have introduced three counterexamples to Reasoning by Cases that can be found in the literature, and defended a novel counterexample involving right-nested indicative conditionals — all this without committing to any specific formal framework. In this section, we present a notion — *upward local persistence* — that is relevant to improve on the *Second Hypothesis* above. Such a notion is better expressed in formal terms.

There are different frameworks that could serve as a formal background for the notion of upward local persistence.[10] In what follows, we focus on some recent developments of informational logic (Yalcin, 2007; Bledin, 2014). Informational logic quite naturally embodies the accounts of modals and conditionals that are adopted by the proponents of the three counterexamples considered in Sec.5.2, and therefore permits us to fruitfully engage in the debate about the scope of validity of Reasoning by Cases.

5.4.1 Informational Logic

Some linguistic expressions are relative to the body of information that is assumed in discourse or reasoning. At this point, this does not come as a surprise: in the previous sections, we have seen that modals such as *ought*, *must*, *may*, *probably* can be dependent on a certain body of information, and that the truth of sentences involving deontic, epistemic or probabilistic modals may vary with changes in the available information. In informational logic, this basic picture is formalized by evaluating sentences at a double-index (w, i), consisting of a possible world w and an information state i. It is convenient to begin by modeling i simply as a finite set of worlds.

Before delving into the details on how those information-sensitive expressions are evaluated in informational logic, it is worth highlighting the general characteristics of such a framework. What properly distinguishes informational logic from the other frameworks mentioned above is the notion of *support*:[11]

Definition 5.1. *Let φ be a sentence of the background language and i an information state. The information state i* **supports** *φ just in case φ is true at (w, i) for all $w \in i$.*

Hence to say that φ is supported in i is to say that φ expresses a property that holds for i irrespectively of any "vantage point" $w \in i$. The notion of *support* serves as a basis for the definition of logical consequence, which, in turn, can

[10] E.g., data semantics (Veltman, 1985), the Kratzerian framework for modality and conditionals (Kratzer, 2012a,c,d), Kolodny and MacFarlane's semantics (Kolodny and MacFarlane, 2010).

[11] The formal notion of *support* traces back at least to Veltman (1996). In the main text, we refer to the definition of *support* adopted in Bledin (2014). Bledin's definition differs from Veltman's in being static and, ultimately, appealing to the notion of *truth at* (w, i). For the purposes of the present chapter, Bledin's formulation has some advantages with respect to Veltman's. We will come back to this when discussing the notion of upward local persistence.

be interpreted as preservation of properties of information states (Yalcin, 2007; Bledin, 2014):

Definition 5.2. *A formula* ψ *is a* **logical consequence** *of a set of formulas* $\varphi_1, ..., \varphi_n$*, written* $\varphi_1, ..., \varphi_n \models \psi$*, just in case there is no* i *which supports the premises* $\varphi_1, ..., \varphi_n$ *but does not support the conclusion* ψ*.*

The above definitions provide us with a general overview of informational logic: sentences are evaluated at indeces; support at i is defined as truth at indeces (w, i) for all $w \in i$; and logical consequence is defined in terms of preservation of support.

Let us turn to the semantics of the background language. We adopt a step by step approach: first the semantic clauses of atoms and boolean connectives, then of deontic and epistemic modals, and finally of the indicative conditional. The semantic clause for the probabilistic modal requires some modification of the underlying notion of an information state, and so it is presented last.

Let V be an interpretation function that maps each atom in the background language and world w to a truth value $\{T, F\}$. Truth at (w, i) for atoms and boolean connectives is defined as follows:[12]

- p is true at (w, i) iff $V(p, w) = T$
- $\neg\varphi$ is true at (w, i) iff φ is not true at (w, i)
- $\varphi \vee \psi$ is true at (w, i) iff φ is true at (w, i) or ψ is true at (w, i)
- $\varphi \wedge \psi$ is true at (w, i) iff φ is true at (w, i) and ψ is true at (w, i)
- $\varphi \supset \psi$ is true at (w, i) iff φ is not true at (w, i) or ψ is true at (w, i)

Note that atoms and boolean connectives receive the standard interpretation. In particular, it is worth stressing that the parameter i plays no role in the evaluation of an atomic sentence: atoms denote world properties which are not sensitive to variations of information.

Deontic, epistemic and probabilistic modals are, on the other hand, sensitive to those variations. Under such reading, they are called *information modals*. Let us consider them in turn, starting from *ought*.

Let d be a function which maps each information state i to a set of deontically

[12] Unless specified otherwise, the semantic clauses are taken from Bledin (2014) who, in turn, credits Yalcin (2007) and Kolodny and MacFarlane (2010) (the latter especially for the semantic clause of *ought*).

ideal worlds $d(i) \subseteq i$. The deontic *ought* O is a universal quantifier over the set of deontically ideal worlds:

- $O\varphi$ is true at (w, i) iff $\forall v \in d(i)$, φ is true at (v, i)

In other words, a sentence $O\varphi$ is true at (w, i) if and only if φ is true in all the worlds which are deontically ideal relative to the information state i.

Epistemic modals *must* ($\Box$) and *might* ($\Diamond$) express, respectively, what is necessary and possible given a certain body of information. That is, they are universal and existential quantifiers over i:

- $\Box\varphi$ is true at (w, i) iff $\forall v \in i$, φ is true at (v, i)
- $\Diamond\varphi$ is true at (w, i) iff $\exists v \in i$, φ is true at (v, i)

Now we turn our attention to the indicative conditional. In Sec.5.3.1 we argued that indicative conditionals involve a shift to a hypothetical context in which the antecedent holds. This idea is made precise as follows.

First, let us write $[i + \varphi]$ to indicate a maximal φ-subset of an information state i. More precisely, $[i + \varphi]$ is such that:

i) It is a φ-subset of i, i.e., $[i + \varphi] \subseteq i$ such that $\forall w \in [i + \varphi]$: φ is true at $(w, [i + \varphi])$

ii) It is maximal, i.e., there is no i' such that $[i + \varphi] \subset i' \subseteq i$ and i' is a φ-subset of i.

Intuitively, $[i + \varphi]$ is a maximal subset of i where φ is true everywhere.

Equipped with this definition, we can finally turn to the semantic clause of the indicative conditional (Kolodny and MacFarlane (2010)):

- $\varphi \Rightarrow \psi$ is true at (w, i) iff $\forall [i + \varphi]$, $\forall v \in [i + \varphi]$: ψ is true at $(v, [i + \varphi])$

where, as just explained, $[i + \varphi]$ is a maximal φ-subset of i. In case there is just one maximal φ-subset of i (for instance, when φ does not contain any modality), the semantic clause of the conditional reduces to the following, simpler one:

- $\varphi \Rightarrow \psi$ is true at (w, i) iff $\forall v \in [i + \varphi]$: ψ is true at $(v, [i + \varphi])$

Indicative conditionals involve a *change of context*: antecedents restrict the information state in which consequents are evaluated. True indicative conditionals are those whose consequents are true for every world in that restricted information state. Or, to phrase this using the notion of support as defined in Def.5.1: an indicative conditional is true if and only if the information state which supports the antecedent also supports the consequent.

The case of the informational modal *probably* requires some further sophistication, as the simple definition of information states as sets of worlds is not sufficient. Following Yalcin (2010) and Yalcin (2012b), we take information states to be probability spaces, i.e., pairs $i = (s_i, Pr_i)$ where s is a finite set of worlds and Pr is a probability function obeying the standard laws of probability calculus.[13] The semantic clause for *probably* ($\triangle$) is the following:

- $\triangle\varphi$ is true at (w, i) iff $Pr_i(\{v : \varphi \text{ is true at } (v, i)\}) > 0.5$

Roughly speaking, $\triangle\varphi$ is true at (w, i) if and only if the likelihood of φ relative to i is higher than 0.5. This concludes our presentation of the semantics of informational logic.

What principles does the semantics of modals and indicative conditionals reflect? From what we have observed so far, it follows that it is the parameter i that plays a crucial role in the semantics of information modals and indicative conditionals. Contrary to what is the case for atoms and boolean connectives, if a sentence like $O\varphi$, $\Box\varphi$, $\Diamond\varphi$, $\triangle\varphi$ or $\varphi \Rightarrow \psi$ is true at (w, i) then it is true at (v, i) for every world $v \in i$. One way to see this result is to interpret the semantics of informational modals and indicative conditionals through the lenses of a kind of *expressivism*: informational modals and indicative conditionals do not describe properties of worlds, but rather express irreducible features of informational states. This kind of expressivism has been defended, for instance, by Yalcin (2012b) and Starr (2016), and seems particularly suited to some dynamic versions of the informational logic framework. Specifically, it is suited to those dynamic versions which dismiss the notion of *truth at* (w, i) altogether, and establish the semantics on different grounds.[14] However, the

[13] In particular, Pr satisfies:

- Normalization: $Pr(s)$=1
- Non-negativity: $Pr(p) \geq 0$, where p is a subset of s
- Additivity: $Pr(p \cup q) = Pr(p) + Pr(q)$, where p,q are disjoint subsets of s.

See Yalcin (2010, 2012b) and, for a general introduction to probability calculus, Halpern (2003).

[14] For this type of dynamic semantics, the seminal reference is Veltman (1996). In Veltman (1996), meaning is not associated to truth conditions but rather with a so-called *context change potential*, that is, with the effect that a certain sentence triggers in the information state. The context change potential of some linguistic expressions (like atoms) consists in

present semantics of informational modals and indicative conditionals structurally relies on the notion of *truth at* (w, i), which serves as basis for the notions of *support* and *logical consequence.* An alternative interpretation is hence available: rather than irreducible features of information states, informational modals and indicative conditionals can be taken to describe global properties of worlds *given* an information state. In the section Sec.5.4.2, we show that there is at least one, technical, reason to do so.

Before moving on, let us consider some of the (in)validities informational logic predicts.

Remark 5.3. *Modus Ponens is valid, as it is Reasoning by Cases with the material conditional. However, Reasoning by Cases with indicative conditionals is invalid. In particular, none of the problematic arguments described in Sec.5.2–5.3 is valid.*

1. $\varphi, \varphi \Rightarrow \psi \models \psi$
2. $\varphi_1 \vee \varphi_2, \varphi_1 \supset \psi_1, \varphi_2 \supset \psi_2 \models \psi_1 \vee \psi_2$
3. $p_1 \vee p_2, p_1 \Rightarrow Oq_1, p_2 \Rightarrow Oq_2 \not\models Oq_1 \vee Oq_2$
4. $p_1 \vee p_2, p_1 \Rightarrow \Box q_1, p_2 \Rightarrow \Box q_2 \not\models \Box q_1 \vee \Box q_2$
5. $p_1 \vee p_2, p_1 \Rightarrow \triangle q_1, p_2 \Rightarrow \triangle q_2 \not\models \triangle q_1 \vee \triangle q_2$
6. $p_1 \vee p_2, p_1 \Rightarrow (q_1 \Rightarrow q_2), p_2 \Rightarrow (q_3 \Rightarrow q_2) \not\models (q_1 \Rightarrow q_2) \vee (q_3 \Rightarrow q_2)$

Proof of the validity of Modus Ponens can be found in Bledin (2014), while the validity of Reasoning by Cases with material conditionals follows from the fact that boolean connectives receive a standard interpretation in informational logic. This is a good result, as what this chapter is after is *not* Reasoning by Cases with the material conditional. More importantly, informational logic predicts Reasoning by Cases with indicative conditionals not to be generally valid. In particular, all the counter-intuitive arguments in Sec.5.2–5.3 are blocked.[15] [16]

ruling out worlds which do not have the properties described by those linguistic expressions. Informational modals and indicative conditionals, on the other hand, express properties of information states, and their context change potential simply consists in testing whether the relevant information state has those properties. Their meaning is therefore irreducible to any property that holds at the world-level. See also (Willer, 2013, 2014; Gillies, 2010; Yalcin, 2012a).

[15] Counter-models to prove *3-6* can be straightforwardly derived from the scenarios described in Sec.5.2–5.3.

[16] For the sake of illustration, we sketch the counter-model of the problematic argument

5.4.2 Reasoning by Cases and Context Change

Equipped with a more precise account of informational modals and indicative conditionals, we can now see that the inference rule of Reasoning by Cases involves a *double context change.* Consider again the general form of the rule:

1	$\varphi_1 \vee \varphi_2$
2	$\varphi_1 \Rightarrow \psi_1$
3	$\varphi_2 \Rightarrow \psi_2$
4	$\psi_1 \vee \psi_2$

The first context change happens in premises 2) and 3): from a global state i which supports $\ulcorner \varphi_1 \vee \varphi_2 \urcorner$ we move to its partial states i' and i'' which support $\ulcorner \varphi_1 \urcorner$ and $\ulcorner \varphi_2 \urcorner$ respectively. This is the shifting effect of indicative conditionals' antecedents. The second context change happens at the level of 4): from the partial states i' and i'' that support the consequents $\ulcorner \psi_1 \urcorner$ and $\ulcorner \psi_2 \urcorner$ respectively, we move back to the global state i and conclude that i supports $\ulcorner \psi_1 \vee \psi_2 \urcorner$.

For the second context change to be legitimate, however, the information that is incorporated into each of the partial states need to survive in i. This survival is not to be captured in terms of support, but rather in terms of *truth at a world given an information state.* To illustrate this point, an example might

described in Sec.5.3. The relevant information state i is the one of Figure 5.1 (where Win is true at (w_1, i) and (w_3, i)). Premises (P1) $Bt \vee Bh$, (P2) $Bt \Rightarrow (T \Rightarrow Win)$ and (P3) $Bh \Rightarrow (H \Rightarrow Win)$ are all supported in i. Let us show that all premises are true at any (w_j, i), for $w_j \in i$. (P1) is trivial, let us turn to (P2). To show that $\ulcorner Bt \Rightarrow (T \Rightarrow Win) \urcorner$ is true at (w_j, i), let us consider $[i + Bt] = \{w_1, w_2\}$. We need to show that $\ulcorner T \Rightarrow Win \urcorner$ is true at $(w_1, [i + Bt])$ and at $(w_2, [i + Bt])$. We write $[[i + Bt] + T]$ for the T-maximal subset of $[i + Bt]$. Since $[[i + Bt] + T] = \{w_1\}$ and Win is true at $(w_1, [[i + Bt] + T])$, it holds that $\ulcorner T \Rightarrow Win \urcorner$ is true at $(w_1, [i + Bt])$ and at $(w_2, [i + Bt])$, which is what we needed to show.The case for (P3) is similar to (P2). On the other hand, the conclusion (C) $(T \Rightarrow Win) \vee (H \Rightarrow Win)$ is not supported in i. In particular, $\ulcorner T \Rightarrow Win \urcorner$ is not true at any (w_k, i), for $w_k \in i$, because $[i + T] = \{w_1, w_4\}$ and Win is not true at $(w_4, [i + T])$. Similarly, $\ulcorner H \Rightarrow Win \urcorner$ is not true at (w_k, i) because $[i + H] = \{w_2, w_3\}$ and Win is not true at $(w_2, [i + H])$. Since none of the disjuncts is true at (w_k, i), neither is their disjunction (C).

be helpful. Consider the following valid instance of Reasoning by Cases:

1	The coin lands tails $\vee$ The coin lands heads
2	The coin lands tails$\Rightarrow$ you win 10$
3	The coin lands heads$\Rightarrow$ you win nothing
4	You win 10$ $\vee$ You win nothing

Let i' be a partial state which supports "The coin lands tails". The information incorporated into i', namely that you win 10$, does not need to be supported in the global state i. That would be too strong, as it would amount to requiring that the body of information i warrants your 10$ win. A better way to go is then to understand the survival of the information that you win 10$ in terms of truth at the world-level: what needs to be ensured is that every world that carries the information that you win 10$ with respect to the state i' carries the same information with respect to the global state i. Only in this way can we legitimately conclude that the body on information i is such that you win 10 or you win nothing.

World-level information survival from partial states to the global state can be captured by means of the following notion (Veltman, 1985; Gillies, 2010):

Definition 5.4. *A sentence* φ *is* **upward locally persistent** *iff* $\forall w, i', i$ *with* $i' \subseteq i$*: if* φ *true at* (w, i') *then* φ *true at* (w, i)*.*

In other words, a sentence is upward locally persistent when its truth at a world given an information state i' persists at the same world given a larger information state i. Examples of upward local persistent sentences are atoms and boolean combinations thereof. This is not surprising, as the truth of an atomic sentence does not depend on the information state parameter. More remarkable is that a sentence like $\ulcorner \Diamond \varphi \urcorner$, containing the information modal *might*, is upward locally persistent (that is because *might* is interpreted as existential quantification).

5.4.3 Final Hypothesis

Let us take stock. We have shown that Reasoning by Cases involves a double context change, and in particular it involves a context change from partial states back to the global state. To ensure that the latter change is legitimated, we have argued that the information that a world carries given a partial state should survive in the global state. Upward local persistence guarantees such survival. This leads to the third, and final hypothesis concerning the scope of validity of Reasoning by Cases:

> **Final Hypothesis**: If ψ_1 and ψ_2 are upward locally persistent, then Reasoning by Cases is valid.

5.4.4 Assessing the Final Hypothesis

The analysis of Reasoning by Cases as involving a double context change and the formulation of the *Final Hypothesis* that follows from such analysis provide an explanation of *why* Reasoning by Cases fails in the four counterexamples considered in Sec.5.3–5.4. In all those cases, the consequents ψ_1 and ψ_2 are not upward locally persistent.

Remark 5.5. *None of the following sentences are upward locally persistent:*

- Oq
- $\Box q$
- $\triangle q$
- $q_1 \Rightarrow q_2$

The cases of $\ulcorner \Box q \urcorner$ and $\ulcorner q_1 \Rightarrow q_2 \urcorner$ are easy to see: the epistemic modal *must* and the indicative conditional are interpreted as universal quantification (hence not persistent in an enlarged domain of quantification). For $\ulcorner Oq \urcorner$, it is sufficient to point out that the function d can assign different sets of deontically ideal worlds to different information states. It is even possible that $d(i') \not\subset d(i)$ for $i' \subset i$. Thus, a sentence like $\ulcorner Oq \urcorner$ is neither upward locally persistent nor, we might say, downward locally persistent.[17] Concerning the modal *probably*, it should be noted that the probability of an outcome can be higher that 0.5 on a certain set of possible outcomes, but its probability can decrease once the set of possible outcomes is enlarged.

The notion of upward local persistence, on which the *Final Hypothesis* is based, marks a clear distinction within informational modals. It also indicates in which sense indicative conditionals resemble certain informational modals but not others. In this sense, the *Final Hypothesis* makes a finer distinction than the *Second Hypothesis*, which only refers to modalized sentences. Moreover, the *Final Hypothesis* predicts that Reasoning by Cases is valid for the epistemic

[17] To the extent that permissions are interpreted as the dual of oughts, then also sentences like Pq are neither upward nor downward locally persistent.

modal *might*, which we have shown being upward locally persistent. This prediction is independently supported by the observation that the epistemic *might* appears to validate other logical principles, such as Modus Tollens, that are known to be problematic for the other informational modals and the indicative conditional (cf Yalcin, 2012b)).

The *Final Hypothesis*, therefore, predicts a wide array of valid instances of Reasoning by Cases: for example, all those in which ψ_1 and ψ_2 are atomic, boolean combinations of atoms, or of the form $\ulcorner \Diamond \psi \urcorner$. In setting the boudaries of Reasoning by Cases, Bledin (2014) discusses only simple valid instances of Reasoning by Cases in which ψ_1 and ψ_2 are atomic. The *Final Hypothesis* indicates how Bledin's (2014) analysis can be extended and generalized. Horty (2001) and Kooi and Tamminga (2007) present two criteria (respectively: sufficient, and necessary and sufficient) for the validity of specific instances of Reasoning by Cases: those involving *dominance oughts* and *weak action oughts.* While our *Final Hypothesis* concerns the general scope of validity of Reasoning by Cases in different domains, Horty (2001) and Kooi and Tamminga (2007) set finer boundaries within the deontic domain.

Let us conclude by mentioning that the *Final Hypothesis* is, in a certain sense, too cautious. Upward local persistence constitutes a sufficient criterion to identify valid instances of Reasoning by Cases, but not a necessary one. There are some valid instances of Reasoning by Cases in which ψ_1 and ψ_2 are not upward locally persistent (notice, indeed, that the *Final Hypothesis* is not a biconditional). We discuss here two of those valid instances.

First, as shown by Bledin (2014), from the premises $\ulcorner \Box\varphi_1 \vee \Box\varphi_2 \urcorner$, $\ulcorner \Box\varphi_1 \Rightarrow \psi_1 \urcorner$ and $\ulcorner \Box\varphi_2 \Rightarrow \psi_2 \urcorner$, one can correctly derive $\ulcorner \psi_1 \vee \psi_2 \urcorner$ even for ψ_1 and ψ_2 not upward locally persistent. This observation is consistent with our analysis of Reasoning by Cases for which the principle generally involves a double context change. In the special case in which the first premise is of the form $\ulcorner \Box\varphi_1 \vee \Box\varphi_2 \urcorner$, there is no proper shift to any partial state: the global state already supports φ_1 or it already supports φ_2. Hence the requirement of upward local persistence, meant to ensure a legitimate change from partial states back to the global state, is vacuous.

A second example of a valid instance of Reasoning by Cases that does not make use of the requirement of upward local persistence is the following (where φ_1

and φ_2 partition the global state, and $\triangle$ indicates *probably*):

1	$\varphi_1 \vee \varphi_2$
2	$\varphi_1 \Rightarrow \triangle\psi_1$
3	$\varphi_2 \Rightarrow \triangle\psi_1$
4	$\triangle\psi_1$

Similarly, a valid instance is obtained if *probably* is substituted in the above example by the epistemic *must*. Is the above pattern generalizable? Unfortunately not. Consider again the Miners' scenario described in Sec.5.2. It is possible to slightly modify the Miners' Puzzle as follows:

1	The miners are in A $\vee$ they are in B
2	They are in A $\Rightarrow$ you ought to block one shaft
3	They are in B $\Rightarrow$ you ought to block one shaft
4	You ought to block one shaft

The argument has the same pattern as the one just considered for *probably*. The premises are all acceptable, given the Miners' scenario: the miners are in A or they are in B, if they are in A then you ought to block one shaft (shaft A), if they are in B then you ought to block one shaft (shaft B). However, the conclusion is not acceptable: after all, you still ought to block neither shaft.

5.5 Conclusion

When is Reasoning by Cases valid? We argued that it is valid if the consequents of indicative conditionals are upward locally persistent. Upward local persistence is a sufficient criterion, general enough to account for a wide variety of intuitively valid instances of Reasoning by Cases but effective enough to rule out problematic instances of Reasoning by Cases involving oughts, the epistemic *must*, *probably* and right-nested indicative conditionals. Beyond the particular framework in which the criterion was formulated, the analysis presented in this chapter aimed at showing how indicative conditionals and informational modals interact in Reasoning by Cases.

Notes

- In the previous version of the manuscript, the first part of Def.5.1 indicated (w, i) as an index. To ease readibility, this has been omitted.
- Footnote 16 has been added to the previous version of the manuscript.

Chapter 6

Conclusion

This thesis has not advanced a unique deontic logic, but rather sought to shed some light on the various facets of the logical structure of oughts. We see the novelty of our contribution as mainly resulting from broadening the perspective on oughts. The logic of Enkrasia presented in the first part of this work considered the structure of oughts in their rational interaction with goals in plans. The second part focused on the inference rule of Reasoning by Cases and investigated the scope of its validity — not only within the deontic domain, but also in comparison with other modal domains.

We conclude by providing a general overview of the chapters' contributions.

Chapter 2 began by considering an interpretation of the rationality principle of Enkrasia, by which if an agent believes strongly and with conviction she ought to X then X-ing is a goal in her plan. We argued that such relation between believed oughts and goals is not trivial. In fact, it is neither a serial relation nor a unidirectional one. As for the first aspect, not all basic oughts correspond to goals. The logical framework developed in Chapter 2 captures the idea that Enkrasia is indeed a principle of bounded logical validity: it is valid within consistency bounds. As for the second aspect, we argued that further oughts can be derived from goals. These derived oughts register the necessary conditions for the fulfillment of all the agent's goals, but do not translate back into further goals to be planned for in themselves. The neighborhood logic we proposed captured some fine-grained semantic distinctions between basic oughts (which function as the input of the agent's deliberation), goals and derived oughts. Finally we have showed that drawing the distinction between basic and derived oughts can help discriminate between valid and invalid instances of deontic closure. *Chapter 3* continued the investigation of Enkrasia by considering how believed oughts and goals vary under dynamics. We limited the attention to practical dynamics, which only concern changes in the possible courses of events open to the agent. While this sort of dynamics leaves basic oughts unaltered, it significantly affects goals and derived oughts. By using resources from dynamic epistemic logic, we represented both the growth and the decrease of the agent's options. These dynamics made prominent the distinction between basic oughts, goals and derived oughts. Goals, in particular, displayed a relative stability under practical dynamics — in line with what is generally theorized in planning theory.

In *Chapter 4* and *Chapter 5* we turned to our second topic of inquiry: the inference rule of Reasoning by Cases. Our investigation was carried out under the assumption that the so-called Miners' Puzzle provides a genuine counterexample to the validity of Reasoning by Cases. Our aim was not to defend this assumption, although we are sympathetic towards it. Rather, our general concern was about which sort of lessons can be drawn from it. Firstly, one could speculate that although the Miners' Puzzle shows that Reasoning by Cases is not generally valid in the deontic domain, there are some oughts — the objective ones — for which Reasoning by Cases is still a valid inference rule. In fact, the Miners' Puzzle concerns situations of epistemic uncertainty, and objective oughts are not tied to the information agents possess. Chapter 4, however, showed that this conjecture does not hold: there exists an interpretation of objective oughts for which Reasoning by Cases is invalid. We illustrated this by constructing a Miners-like Puzzle that was not ultimately founded on uncertainty, but on indeterminacy. Thus, even if Reasoning by Cases might still be valid for objective oughts under certain interpretations, it is not so for all of them. Chapter 5 broadened the perspective by considering further counterexamples to Reasoning by Cases. These occur across different domains: not only deontic, but also epistemic and probabilistic. As modality is the common denominator of those counterexamples, one could hypothesize that the invalidity of Reasoning by Cases is indeed due to its application to modals. We argued that this hypothesis both undergenerates and overgenerates. It undergenerates because a further counterexample to Reasoning by Cases can be constructed employing only indicative conditionals. It overgenerates as it misses finer distinctions between modals. A more promising approach lies in the observation that a double context-change is involved in Reasoning by Cases. We theorized that the common denominator of all the above-mentioned counterexamples is that they involve sentences whose truth values do not persist when contexts change. Persistence, thus, enabled us to elaborate a sufficient criterion for the validity of Reasoning by Cases. This result, as well as the contribution presented in the first part of the thesis, emerged by attempting to broaden the perspective on oughts. Doing so has allowed certain new questions to be asked, and — hopefully — a few answers to be drawn.

Bibliography

Anglberger, A. J., N. Gratzl, and O. Roy (2015). Obligation, free choice, and the logic of weakest permissions. *The Review of Symbolic Logic 8*(4), 807–827.

Baltag, A., L. S. Moss, and S. Solecki (1998). The Logic of Public Announcements, Common Knowledge, and Private Suspicions (extended abstract). In *Proc. of the Intl. Conf. TARK 1998*, pp. 43–56. Morgan Kaufmann Publishers.

Bellos, A. (March 2016). Can you solve it? The logic question almost everyone gets wrong. *The Guardian*. Retrievied from https://www.theguardian.com/science/2016/mar/28/can-you-solve-it-the-logic-question-almost-everyone-gets-wrong.

Belnap, N. and M. Green (1994). Indeterminism and the thin red line. *Philosophical Perspectives 8*, 365–388.

Blackburn, P., M. de Rijke, and Y. Venema (2001). *Modal Logic*. Cambridge University Press.

Bledin, J. (2014). Logic informed. *Mind 123*(490), 277–316.

Bledin, J. (2015). Modus ponens defended. *The Journal of Philosophy 112*(2), 57–83.

Brachman, R. J., H. J. Levesque, and R. Reiter (1992). *Knowledge representation*. MIT press.

Bratman, M. (1987). *Intention, plans, and practical reason*. Harvard University Press.

Bratman, M. E., D. J. Israel, and M. E. Pollack (1988). Plans and resource-bounded practical reasoning. *Computational intelligence 4*(3), 349–355.

Broersen, J., M. Dastani, and L. Van Der Torre (2001). Resolving conflicts between beliefs, obligations, intentions, and desires. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pp. 568–579. Springer.

Broome, J. (2013). *Rationality Through Reasoning.* Wiley-Blackwell.

Brown, M. A. (2004). Rich deontic logic: a preliminary study. *Journal of Applied Logic 2*(1), 19–37.

Bykvist, K. (2009). Objective versus subjective moral oughts. In R. S. Lars-Göran Johansson, Jan Österberg (Ed.), *Logic, Ethics, and All That Jazz. Essays in Honour of Jordan Howard Sobel*, pp. 39–65. Uppsala Philosophical Studies.

Cantwell, J. (2008). Changing the modal context. *Theoria 74*(4), 331–351.

Cariani, F. (2016). Consequence and contrast in deontic semantics. *The Journal of Philosophy 113*(8), 396–416.

Cariani, F., M. Kaufmann, and S. Kaufmann (2013). Deliberative modality under epistemic uncertainty. *Linguistics and Philosophy 36*(3), 225–259.

Carr, J. (2015). Subjective ought. *Ergo, an Open Access Journal of Philosophy 2*(27).

Charlow, N. (2013). What we know and what to do. *Synthese 190*(12), 2291–2323.

Ciuni, R. and A. Zanardo (2010). Completeness of a branching-time logic with possible choices. *Studia Logica 96*(3), 393–420.

Cohen, P. R. and H. J. Levesque (1990). Intention is choice with commitment. *Artificial Intelligence 42*(2), 213– 261.

Craven, R. and M. Sergot (2008). Agent strands in the action language nc+. *Journal of Applied Logic 6*(2), 172 – 191.

DeRose, K. (2010). The conditionals of deliberation. *Mind 119*(473), 1–42.

van Ditmarsch, H., W. van der Hoek, and B. Kooi (2008). *Dynamic Epistemic Logic.* Springer.

Dowell, J. (2012). Contextualist solutions to three puzzles about practical conditionals. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics*, Volume 7, pp. 271–303. Oxford Scholarship Online.

von Fintel, K. (2012). The best we can (expect to) get? Challenges to the classic semantics for deontic modals. Unpublished manuscript.

von Fintel, K. and A. S. Gillies (2010). *Must*...stay...strong! *Natural Language Semantics 18*(4), 351–383.

Gibbard, A. (2005). Truth and correct belief. *Philosophical Issues 15*, 338–350.

Gibbard, A. (2008). *Reconciling Our Aims: In Search of Bases for Ethics.* Oxford University Press.

Gillies, A. S. (2010). Iffiness. *Semantics and Pragmatics 3*(4), 1–42.

Goranko, V. and S. van Drimmelen (2006). Complete axiomatization and decidability of alternating-time temporal logic. *Theoretical Computer Science 353*(1-3), 93–117.

Halpern, J. (2003). *Reasoning about Uncertainty.* MIT Press.

Hamblin, C. L. (1959). The modal "probably". *Mind 68*(270), 234–240.

Hare, C. (2011). Obligation and regret when there is no fact of the matter about what would have happened if you had not done what you did. *Noûs 45*(1), 190–206.

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

Hilpinen, R. and P. McNamara (2013). Deontic logic: A historical survey and introduction. In *Handbook of deontic logic and normative systems*, Volume 1, pp. 3–136. College Publications.

van der Hoek, W., W. Jamroga, and M. Wooldridge (2007). Towards a theory of intention revision. *Synthese 155*(2), 265–290.

Horty, J. F. (2001). *Agency and Deontic Logic.* Oxford University Press.

Horty, J. F. (2012). *Reasons as Defaults.* Oxford University Press.

Horty, J. F. (2015). Requirements, oughts, intentions. *Philosophy and Phenomenological Research 91*(1), 220–229.

Horty, J. F. and N. Belnap (1995). The deliberative stit: A study of action, omission, ability and obligation. *Journal of Philosophical Logic 24*(6), 583–644.

Horty, J. F. and M. E. Pollack (2001). Evaluating new options in the context of existing plans. *Artificial Intelligence 127*(2), 199–220.

Icard, T., E. Pacuit, and Y. Shoham (2010). Joint revision of belief and intention. In *Proceedings of the 12th International Conference on Knowledge Representation*, pp. 572–574.

Klein, D. and A. Marra (2019). From oughts to goals. A logic for enkrasia. *Studia Logica (to appear).* The final authenticated version is available online at: https://doi.org/10.1007/s11225-019-09854-5.

Kolodny, N. (2005). Why be rational? *Mind 114*(455), 509–563.

Kolodny, N. and J. MacFarlane (2010). Ifs and oughts. *The Journal of philosophy 107*(3), 115–143.

Kooi, B. and A. Tamminga (2007). Moral reasoning by cases. RUC-ILLC Workshop on Deontic Logic, Roskilde University. Extended abstract (unpublished).

Kratzer, A. (1991). Modality. In A. v. Stechow and D. Wunderlich (Eds.), *Handbuch Semantik/Handbook Semantics*, pp. 639–650. Berlin and New York (de Gruyter).

Kratzer, A. (2012a). Conditionals. In *Modals and Conditionals*, pp. 86–108. Oxford University Press.

Kratzer, A. (2012b). Facts: Particulars or information units? In *Modals and Conditionals*, pp. 161–183. Oxford University Press.

Kratzer, A. (2012c). The notional category of modality. In *Modals and Conditionals*, pp. 27–69. Oxford University Press.

Kratzer, A. (2012d). What *must* and *can* must and can mean. In *Modals and Conditionals*, pp. 4–20. Oxford University Press.

Lassiter, D. (2016). Linguistic and philosophical considerations on bayesian semantics. In N. Charlow and M. Chrisman (Eds.), *Deontic Modality*, pp. 82–116. Oxford University Press.

Lorini, E. and A. Herzig (2008). A logic of intention and attempt. *Synthese 163*(1), 45–77.

MacFarlane, J. (2003). Future contingents and relative truth. *The Philosophical Quarterly 53*(212), 321–336.

MacFarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications.* Clarendon Press.

Marra, A. (2016). Objective oughts and a puzzle about futurity. In O. Roy, A. Tamminga, and M. Willer (Eds.), *Deontic Logic and Normative Systems. Proceedings of the 13th International Conference DEON*, pp. 171–186. College Publications.

Marra, A. (2018). When is reasoninig by cases valid? Some remarks on persistence and information sensitivity. To be submitted.

Marra, A. and D. Klein (2015). Logic and ethics: An integrated model for norms, intentions and actions. In W. van der Hoek, W. H. Holliday, and W. Wang (Eds.), *Logic, Rationality, and Interaction. Proceedings of the 5th International Workshop LORI*, pp. 268–281. Springer.

Mason, E. (2013). Objectivism and prospectivism about rightness. *Journal of Ethics and Social Philosophy 7*(2), 1–21.

Mellor, D. H. (1983). Objective decision making. *Social Theory and Practice 9*(2/3), 289–309.

Merricks, T. (2006). Good-bye growing block. In D. Zimmerman (Ed.), *Oxford Studies in Metaphysics.* Oxford University Press.

Moss, S. (2015). On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics 8*(5), 1–81.

Nair, S. (2014). Consequences of reasoning with conflicting obligations. *Mind 123*(491), 753–790.

Nair, S. (2016). Conflicting reasons, unconflicting oughts. *Philosophical Studies 173*(3), 629–663.

Oddie, G. and P. Menzies (1992). An objectivist's guide to subjective value. *Ethics 102*(3), 512–533.

Pacuit, E. (2017). *Neighborhood semantics for modal logic*. Short Textbooks in Logic. Springer.

Parfit, D. (1988). What we together do. Unpublished manuscript.

Pollack, M. E. (1992). The uses of plans. *Artificial Intelligence 57*(1), 43–68.

Pollack, M. E. and J. F. Horty (1999). There's more to life than making plans: plan management in dynamic, multiagent environments. *AI Magazine 20*(4), 71.

Prior, A. (1967). *Past, Present and Future*. Oxford University Press.

Ross, A. (1941). Imperatives and logic. *Theoria 7*, 53–71.

Shpall, S. (2013). Wide and narrow scope. *Philosophical Studies 163*(3), 717–736.

Silk, A. (2014a). Evidence sensitivity in weak necessity deontic modals. *Journal of Philosophical Logic 43*(4), 691–723.

Silk, A. (2014b). Why 'ought' detaches: Or, why you ought to get with my friends (if you want to be my lover). *Philosophers' Imprint 14*(7).

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Starr, W. (2016). Dynamic expressivism about deontic modality. In N. Charlow and M. Chrisman (Eds.), *Deontic Modality*, pp. 355–396. Oxford University Press.

Thomason, R. (1984). *Combinations of Tense and Modality*. Reidel Publishing Company.

Thomason, R. H. (2000). Desires and defaults: A framework for planning with inferred goals. In *KR*, pp. 702–713. Citeseer.

Veltman, F. (1985). *Logics for Conditionals*. Ph. D. thesis, University of Amsterdam.

Veltman, F. (1996). Defaults in update semantics. *Journal of Philosophical Logic 25*(3), 221–261.

Veltman, F. (2011). Or else, what? imperatives on the borderline of semantics and pragmatics. Unpublished notes.

Venema, Y. (2001). Temporal logic. In L. Goble (Ed.), *The Blackwell guide to philosophical logic*, pp. 203–223. Blackwell.

Wedgwood, R. (2003). Choosing rationally and choosing correctly. In *Weakness of Will and Practical Irrationality*, pp. 201–230. Oxford University Press.

Wedgwood, R. (2016). Objective and subjective 'ought'. In *Deontic Modality*. Oxford University Press.

Willer, M. (2012). A remark on iffy oughts. *The Journal of Philosophy 109*(7), 449–461.

Willer, M. (2013). Dynamics of epistemic modality. *Philosophical Review 122*(1), 45–92.

Willer, M. (2014). Dynamic thoughts on ifs and oughts. *Philosophers' Imprint 14*(28), 1–30.

Willer, M. (2015). An update on epistemic modals. *Journal of Philosophical Logic 44*(6), 835–849.

Von Wright, G. H. (1963). Practical inference. *The philosophical review 72*(2), 159–179.

Yalcin, S. (2007). Epistemic modals. *Mind 116*(464), 983–1026.

Yalcin, S. (2010). Probability operators. *Philosophy Compass 5*(11), 916–937.

Yalcin, S. (2012a). Context probabilism. In M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz, and M. Westera (Eds.), *Logic, Language and Meaning. Proceedings of the 18th Amsterdam Colloquium*, pp. 12–21. Springer.

Yalcin, S. (2012b). A counterexample to modus tollens. *Journal of Philosophical Logic 41*(6), 1001–1024.

Yalcin, S. (2016). Modalities of normality. In N. Charlow and M. Chrisman (Eds.), *Deontic Modality*, pp. 230–255. Oxford University Press.