Deontic Logic and Preference Change

van Benthem, J.; Liu, F.

Publication date
2014

Document Version
Final published version

Published in
The IfCoLog Journal of Logics and their Applications

Deontic Logic and Preference Change

Johan van Benthem
University of Amsterdam and Stanford University

Fenrong Liu
Tsinghua University, Beijing, China

Abstract

The normative realm involves deontic notions such as obligation or permission, as well as information about relevant actions and states of the world. This mixture is not static, given once and for all. Both information and normative evaluation available to agents are subject to changes with various triggers, such as learning new facts or accepting new laws. This paper explores models for this setting in terms of dynamic logics for information-driven agency. Our paradigm will be dynamic-epistemic logics for knowledge and belief, and their current extensions to the statics and dynamics of agents' preferences. Here the link with deontics is that moral reasoning may be viewed as involving preferences of the acting agent as well as moral authorities such as lawgivers, one's conscience, or yet others. In doing so we discuss a large number of themes: primitive 'betterness' order versus reason-based preferences (employing a model of 'priority graphs'), the entanglement of preference and informational attitudes such as belief, interactive social agents, and scenarios with long-term patterns emerging over time. Specific deontic issues considered include paradoxes of deontic reasoning, acts of changing obligations, and changing norm systems. We conclude with some further directions, as well as a series of pointers to related work, including different paradigms for looking at these same phenomena.

Keywords: Deontic Logic, Preference Change, Epistemic Logic, Public Announcement Logic.

This paper is an updated and revised version of a draft chapter for the *Handbook of Deontic Logic*, that was written originally in 2009. Given the recent increasing interest in our central themes of reason-based preference and preference dynamics, we are publishing the present version in the IF-COLOG Journal at the kind suggestion of Dov Gabbay. We are grateful to Guillaume Aucher and Davide Grossi for many useful comments and pointers, the majority of which will feed into the final chapter version when the Handbook appears.

1 Agency, information, and preference

Agents pursue goals in this world, acting within constraints in terms of their information about what is true, as well as norms about what is right. The former dimension typically involves acts of inference, observation, as well as communication and other forms of social interaction. The latter dimension involves evaluation of situations and actions, 'coloring' the agents' view of the world, and driving their desires, decisions, and actions in it. A purely informational agent may be rational in the sense of clever reasoning, but a *reasonable* agent is one whose actions are in harmony with what she wants. The two dimensions are intimately related. For instance, what we want is influenced by what we believe to be true as well as what we prefer, and normally also, we only seek information to further goals that we desire.

This balance of information and evaluation is not achieved once and for all. Agents must constantly cope with new information, either because they learn more about the current situation, or because the world has changed. But equally well, agents constantly undergo changes in evaluation, sometimes by intrinsic changes of heart, but most often through events with normative impact, such as accepting a command from an authority. These two forms of dynamics, too, are often entangled: for instance, learning more about the facts can change my evaluation of a situation.

A third major aspect of agency is its social interactive character. Even pure information flow is often driven by an epistemic gradient: the fact that different agents know different things leads us to communicate, whether in cooperative inquiry or adversarial argumentation, perhaps until a state of equilibrium is reached such as common knowledge or common belief. But also more complex forms of interaction occur, such as merging beliefs, where differences in informational authority may play a crucial role. Again, very similar phenomena play on the normative side. Norms, commitments and duties usually involve other agents, both as their source and as their target, and whole institutions and societies are constructed in terms of social choice, shared norms and rules of behavior.

In this chapter, we will discuss how current dynamic-epistemic logics can model the above phenomena, both informational and preferential, and we will show what results when this perspective is taken to normative reasoning and deontic logic. Our treatment will be brief, and for a much more elaborate sample of this style of thinking about the normative realm, we refer to [17]. In pursuing this specific line, we are not denying the existence of other valid approaches to deontic dynamics, and we will provide a number of references to other relevant literature.

2 Dynamic logics of knowledge and belief change

Before analyzing preference or related deontic notions, we first develop the basic methodology of this paper for the purely informational case, where the first 'dynamic-epistemic logics' arose in the study of information change.

2.1 Epistemic logic and semantic information

Dynamic logics of agency need an account of underlying of static states that can be modified by suitable triggers: actions or events. Such states usually come from existing systems in philosophical or computational logic whose models can serve as static snapshots of the dynamic process. In this paper, we start with a traditional modal base system of epistemic logic, referring to the standard literature for details (cf. [45] and [30]).

Definition 1. *Let a set of propositional variables Φ be given, as well as a set of agents A. The epistemic language is defined by the syntax rule*

$$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \qquad \textit{where } p \in \Phi,\ a \in A.$$

Remark: Single agents, interacting agents, and groups. For convenience, we will focus on single agents in this paper, although this still allows us to describe interacting individual agents where needed through iterations of modalities. Epistemically important notions with groups themselves as agents, such as 'common knowledge' or 'distributed knowledge', are deferred to our discussion at the end.

Semantic models for the epistemic language encode agents' 'information ranges' in the form of equivalence classes of binary uncertainty relations for each agent.[1] These support a standard compositional truth definition.

Definition 2. *An* epistemic model *is a tuple $\mathfrak{M} = (W, \{\sim_a\}_{a\in A}, V)$ with W a set of epistemically possible states (or 'worlds'), $\sim_a$ an equivalence relation on W, and V a valuation function from Φ to subsets of W.*

Definition 3. *For an epistemic model $\mathfrak{M} = (W, \{\sim_a \mid a \in A\}, V)$ and any world $s \in S$, we define $\mathfrak{M}, s \models \varphi$ (epistemic formula φ* is true in $\mathfrak{M}$ at *s) by induction on the structure of the formula φ:*

1. $\mathfrak{M}, s \models \top$ *always.*

[1] The approach of this paper will also work on more general relations such as pre-orders, but we start with this easily visualizable epistemic case for expository purposes.

2. $\mathfrak{M}, s \models p$ iff $s \in V(p)$.

3. $\mathfrak{M}, s \models \neg\varphi$ iff *not* $\mathfrak{M}, s \models \varphi$.

4. $\mathfrak{M}, s \models \varphi \wedge \psi$ iff $\mathfrak{M}, s \models \varphi$ *and* $\mathfrak{M}, s \models \psi$.

5. $\mathfrak{M}, s \models K_a\varphi$ iff *for all* t *with* $s \sim_a t : \mathfrak{M}, t \models \varphi$.

Using equivalence relations in our models yields the well-known modal system **S5** for each individual knowledge modality, without interaction laws for different agents. Just for concreteness, we record this basic fact here:

Theorem 4. *Basic epistemic logic is axiomatized completely by the axioms and inference rules of the modal system **S5** for each separate agent.*

Few researchers see our basic modalities and the simple axioms of modal **S5** as expressing genuine properties of 'knowledge' – thus making earlier polemical discussions of epistemic 'omniscience' or 'introspection' expressed by these axioms obsolete. Our interpretation of the above notions is as describing the *semantic information* that agents have available (cf. [13]), being a modest but useful building block in analyzing more complex epistemic and deontic notions. We will allow ourselves the use of the word 'know' occasionally, however: old habits die hard.[2]

Static epistemic logic describes what agents know on the basis of their current semantic information. But information flows, and a richer story must also include dynamics of actions that produce and modify information. We now turn to the simplest case of this dynamics: reliable public announcements or public observations, that shrink the current information range.

2.2 Dynamic logic of public announcement

The pilot for the methodology of this paper is 'public announcement logic' (PAL), a toy system describing a combination of epistemic logic and one dynamic event, namely, *announcement* of new 'hard information' expressed in some proposition φ that is true at the actual world. The corresponding 'update action' $!\varphi$ transforms a current epistemic model $\mathfrak{M}, s$ into its definable submodel $\mathfrak{M}|\varphi, s$ where all worlds

[2] There is a fast-growing literature on more sophisticated logical analyses of genuine knowledge (cf. [76], [27], [123]), which also seems relevant to modeling and reasoning in the deontic realm. However, the main points to be made in this paper are orthogonal to these additional refinements of the logical framework.

that did not satisfy φ have been eliminated. This model update is the basic scenario of obtaining information in the realm of science but also of common sense, by shrinking one's current epistemic range of uncertainty.[3]

To describe this phenomenon, the *language* of PAL has two levels, using both formulas for propositions and action expressions for announcements:

$$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid [A]\varphi$$
$$A := !\varphi$$

The new dynamic formula $[\varphi]\psi$ says that "after updating with the true proposition φ, formula ψ holds":

$$\mathfrak{M}, s \models [!\varphi]\psi \;\text{ iff }\; \text{if } \mathfrak{M}, s \models \varphi, \text{ then } \mathfrak{M}|\varphi, s \models \psi.$$

This language can make characteristic assertions about knowledge change such as $[!\varphi]K_a\psi$, which states what agent a will know after having received the hard information that φ. In particular, the knowledge change before and after an update can be captured by so-called *recursion axioms*, a sort of recursion equations for the 'dynamical system' of PAL, relating new knowledge to knowledge that agents had before. Here is the complete logical system for information flow under public announcement (two original sources are [53], [111]):

Theorem 5. *PAL is axiomatized completely by the usual laws of the static epistemic base logic plus the following recursion axioms:*

1. $[!\varphi]q \leftrightarrow (\varphi \to q)$ *for atomic facts q*
2. $[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$
3. $[!\varphi](\psi \wedge \chi) \leftrightarrow ([!\varphi]\psi \wedge [!\varphi]\chi)$
4. $[!\varphi]K_a\psi \leftrightarrow (\varphi \to K_a[!\varphi]\psi)$

These elegant principles analyze reasoning about epistemic effects of receiving hard information, through observation, communication, or other reliable means. In particular, the knowledge law reduces knowledge after new information to 'conditional knowledge' that the agent had before, but in a subtle recursive manner. This prudence of design for PAL is necessary since the process of information update can change truth values of epistemic assertions. Perhaps, initially, I did not know that p, but after the event $!p$, I do.

[3]The name 'public announcement logic' may be unfortunate, since the logic describes updates with hard information from whatever source, but no consensus has emerged yet on a rebaptism.

There are several noteworthy features to this approach. We already stressed the recursive nature of reducing new knowledge to pre-existing knowledge, a feature that is typical of dynamical systems. Also, the precise way in which this happens involves breaking down the 'postconditions' behind the dynamic modalities $[!\varphi]$ compositionally on the basis of their shape.

Next, as things stand here, repeating these steps, the stated features drive a 'reduction process' taking every formula of our dynamic-epistemic language eventually to an equivalent formula inside the static epistemic language. In terms of semantics and expressive power, this means that a current static model 'pre-encodes' all information about what might happen when agents communicate what they know. In terms of the logic, the reduction procedure means that PAL is *axiomatizable* and *decidable*, since it inherits these features from the epistemic base logic.

However, it is also important to note that the latter sweeping dynamics-to-statics reduction is not an inevitable feature of dynamic-epistemic analysis. In recent versions of the semantics for PAL, the available sequences of information updates may be constrained by global *protocols* that regulate available events in the current process of inquiry. In that case, no reduction is possible to the base logic, and the dynamic logic, though still employing recursion equations, while also remaining axiomatizable and decidable, comes to encode a genuine new kind of 'procedural information' (cf. [14]). Protocols also make sense for deontic purposes, because of the procedural character of much normative behavior, and we will briefly return to this perspective at the end of this chapter.

In what follows, PAL will serve as a pilot example for many other complex cases, for example, changes in beliefs, preferences, and obligations. In each case, the 'triggering events' can be different: for instance, beliefs can change by signals of different force: hard or more 'soft', and obligations can change through actions of commanding by a normative authority. In many cases, the domain of the model does not change, but rather its *ordering pattern.*[4] However, the general recursive methodology of *PAL* will remain in force, though in each case, with new twists.

2.3 From knowledge to belief and soft information

Knowledge rests on hard information, but most of the information that we have and act on is soft, giving rise to *beliefs*, that are not always true, and that can be revised when shown inadequate. One can think of learning from error as the more creative ability, beyond mere recording of reliable information in the agent's environment.

[4] One example of this approach, even in the epistemic realm, are 'link cutting' versions of updating after announcement: cf. [90], [126], [26], that will be used later on in scenarios where we may want to return to worlds considered earlier in the process.

Again we need to start with a convenient static base for our investigation. One powerful model for soft information and belief reflects the intuition that we believe those things that hold in the *most plausible* worlds in our epistemic range. I believe that this train will take me home on time, even though I do not know that it will not suddenly fly away from the tracks. But the worlds where it stays on track are more plausible than those where it flies off, and among the latter, those where it arrives on time are more plausible than those where it does not.

The long history for this way of modeling belief includes non-monotonic logic in artificial intelligence ([124], [32], [86], [50], [51]), the semantics of natural language (cf. [139]), as well as the philosophical literature on epistemology, and logics of games (cf. [129], [10]).

The common intuition of relative plausibility leads to the following semantics:

Definition 6. *An* epistemic-doxastic model $\mathfrak{M} = (W, \{\sim_a\}_{a\in A}, \{\leq_a\}_{a\in A}, V)$ *consists of an epistemic model* $(W, \{\sim_a\}_{a\in A}, V)$ *as before, while the* $\leq_a$ *are binary comparative plausibility pre-orders for agents between worlds.*

Intuitively, these comparison orders might well be *ternary* $\leq_{a,s} xy$ saying that, in world s, agent a considers world x at least as plausible as y.[5] For convenience in this chapter, however, our semantics assumes that plausibility orderings are the same for epistemically indistinguishable worlds: that is, agents know their plausibility judgements. Assuming that plausibility is a pre-order, i.e., reflexive and transitive, but not necessarily connected, leaves room for the existence of genuinely incomparable worlds – but much of what we say in this chapter also holds for the special case of *connected* pre-orders where any two worlds are comparable.[6] As with epistemic models, our logical analysis works largely independently from specific design decisions about the ordering, important though they may be in specific applications.

One can interpret many logical languages in these comparative order structures. In what follows, we work with modal formalisms for the usual reasons of perspicuous formulation and low complexity (cf. [29]).

First of all, there is *absolute belief* as truth in all most plausible worlds:

> $\mathfrak{M}, s \models B_a\varphi$ iff $\mathfrak{M}, t \models \varphi$ *for all those worlds* $t \sim_a s$ *that are maximal in the order* $\leq_a xy$ *in the* $\sim_a$*-equivalence class of* s.

[5] In particular, ternary world-dependent plausibility relations are found in the semantics of conditional logic: cf. [89], [127], models for games: cf. [130], [13], as well as in recent logical analyses of major paradigms in epistemology: [76].

[6] Connected orders are equivalent to the 'sphere models' of conditional logic or belief revision theory (cf. [62], [120]) – but in these areas, too, a generalization to pre-orders has been proposed: cf. [35], [124], [138].

But the more general notion in our models is that of a *conditional belief*:

$\mathfrak{M}, s \models B_a^{\psi}\varphi$ iff $\mathfrak{M}, t \models \varphi$ *for all those worlds* $t \sim_a s$ *that are maximal for* $\leq_a xy$ *in the set* $\{u \mid s \sim_a u$ *and* $\mathfrak{M}, u \models \psi\}$.[7]

Conditional beliefs generalize absolute beliefs, which are now definable as $B_a^{\top}\varphi$. They *pre-encode* absolute beliefs that we will have *if* we learn certain things. Indeed, the above semantics for $B_a^{\psi}\varphi$ is formally similar to that for conditional assertions $\psi \Rightarrow \varphi$. This allows us to use known results from [35], [138]:

Theorem 7. *The logic of* $B_a^{\psi}\varphi$ *is axiomatized by standard propositional logic plus the laws of conditional logic over pre-orders.*

Deductively stronger modal logics also exist in this area, such as the popular system **KD45** for absolute belief. The structural content of their additional axioms can be determined through standard modal frame correspondence techniques (see [29], [23]).

Digression: Further relevant attitudes. Modeling agency with just the notions of knowledge and belief is mainly a tradition inherited from the literature. In a serious study of agency the question needs to be raised afresh what is our natural repertoire of attitudes triggered by information. As one interesting example, the following operator has emerged recently, in between knowledge and belief qua strength. Intuitively, 'safe belief' is belief that agents have which cannot be falsified by receiving true new information.[8] Over epistemic plausibility models $\mathfrak{M}$, its force is as follows:

Definition 8. *The modality of* safe belief $B_a^{+}\varphi$ *is interpreted as follows:*

$\mathfrak{M}, s \models B_a^{+}\varphi$ iff *for all worlds* $t \sim_a s$*: if* $s \leq_a t$*, then* $\mathfrak{M}, t \models \varphi$.

Thus, the formula φ is to be true in all accessible worlds that are at least as plausible as the current one. This includes the most plausible worlds, but it need not include all epistemically accessible worlds, since the latter may include some less plausible worlds than the current one. The logic for safe belief is just **S4**, since it is in fact the simplest modality over the plausibility order.

[7] These intuitive maximality formulations must be modified in models allowing infinite sequences in the plausibility ordering. Trivialization can then be avoided as follows (cf. the exposition of plausibility semantics in [54]): $\mathfrak{M}, s \models O^{\psi}\varphi$ iff $\forall t \sim s : \exists u : (t \preceq u$ *and* $\mathfrak{M}, u \models \psi$ *and* $\forall v \sim s$: (*if* $u \preceq v$ *and* $\mathfrak{M}, v \models \psi$, *then* $\mathfrak{M}, v \models \varphi$)).

[8] This notion has been proposed independently in AI [125], philosophy [131], learning theory, and game theory [8], [11].

A notion like this has the conceptual advantage of making us see that agents can have more responses to information than just knowledge and belief.[9] But there is also the technical advantage that the simple modality of safe belief can define more complex notions such as conditional belief (see [85], [33], [13]) which can lead to simplifications of logics for agency.

2.4 Dynamic logics of belief change

Having set up the basic attitudes, we now want to deal with explicit acts or events that update not just knowledge, but also agents' beliefs.[10]

Hard information The first obvious triggering event are the earlier public announcements of new hard information. Their complete logic of belief change can be developed in analogy with the earlier dynamic epistemic logic PAL, again via world elimination. Its key recursion axiom for new beliefs uses conditional beliefs:

Fact 9. *The following formula is valid in our semantics:*

$$[!\varphi]B_a\psi \leftrightarrow (\varphi \to B_a^{\varphi}[!\varphi]\psi)$$

To keep the complete dynamic language in harmony, we then also need a recursion axiom for the conditional beliefs that are essential here:

Theorem 10. *The dynamic logic of conditional belief under public announcements is axiomatized completely by*

(a) any complete static logic for the model class chosen,

(b) the PAL recursion axioms for atomic facts and Boolean operations,

(c) the following recursion axiom for conditional beliefs:

$$[!\varphi]B_a^{\chi}\psi \leftrightarrow (\varphi \to B_a^{\varphi \wedge [!\varphi]\chi}[!\varphi]\psi)$$

This analysis also extends to safe belief, with this recursion law:

Fact 11. *The following PAL-style axiom holds for safe belief:*

$$[!\varphi]B_a^{+}\psi \leftrightarrow (\varphi \to B_a^{+}(\varphi \to [!\varphi]\psi)).$$

[9] Other relevant notions include the 'strong belief' of [131], [10].

[10] For a much more extensive up-to-date treatment of logic-based belief revision, cf. the chapter [28] in the forthcoming *Handbook of Logics of Knowledge and Belief.*

Using this equivalence, which behaves more like the original central PAL axiom, one can show that safe belief has its intuitively intended feature. Safe belief in factual propositions (i.e., those not containing epistemic or doxastic operators) remains safe belief after updates with hard factual information.[11]

Soft information But belief change also involves more interesting triggers, depending on the quality of the incoming information, or the trust agents place in it. 'Soft information upgrade' does not eliminate worlds as what hard information does, but rather *changes the plausibility order*, promoting or demoting worlds according to their properties. Here is one widely used way in which this can happen: an act of 'radical', or 'lexicographic' upgrade.[12]

Definition 12. *A* radical upgrade $\Uparrow\varphi$ *changes the current plausibility order* $\leq$ *between worlds in* $\mathfrak{M}, s$ *to create a new model* $\mathfrak{M}\Uparrow\varphi, s$ *where all* φ*-worlds in* $\mathfrak{M}, s$ *become better than all* $\neg\varphi$*-worlds, while, within those two zones, the old plausibility order* $\leq$ *remains as it was.*

No worlds are eliminated here, it is the ordering pattern that adapts. There is a matching upgrade modality for this in our dynamic language:

$\mathfrak{M}, s \models [\Uparrow\varphi]\psi$ iff $\mathfrak{M}\Uparrow\varphi, s \models \psi$.

This supports one more dynamic completeness theorem (cf.[22]).

Theorem 13. *The logic of radical upgrade is axiomatized completely by*

(a) a complete axiom system for conditional belief on the static models,

(b) the following recursion axioms:

$$[\Uparrow\varphi]q \leftrightarrow q, \quad \textit{for all atomic proposition letters } q$$
$$[\Uparrow\varphi]\neg\psi \leftrightarrow \neg[\Uparrow\varphi]\psi$$
$$[\Uparrow\varphi](\psi \wedge \chi) \leftrightarrow ([\Uparrow\varphi]\psi \wedge [\Uparrow\varphi]\chi)$$
$$[\Uparrow\varphi]B^{\chi}\psi \leftrightarrow (E(\varphi \wedge [\Uparrow\varphi]\chi) \wedge B^{\varphi\wedge[\Uparrow\varphi]\chi}[\Uparrow\varphi]\psi)$$
$$\vee (\neg E(\varphi \wedge [\Uparrow\varphi]\chi) \wedge B^{[\Uparrow\varphi]\chi}[\Uparrow\varphi]\psi$$

[11] Unlike with plain belief, the latter recursion does not involve a move to an irreducible new notion of 'conditional safe belief'. Indeed, given a definition of conditional belief in terms of safe belief, the more complex recursion law in Theorem 10 can be derived.

[12] In this section, we drop epistemic accessibility, and focus on plausibility order only.

Here the operator 'E' is the existential epistemic modality, and we need to add a simple recursion axiom for knowledge, that we forego here.[13]

There are many further policies for changing plausibility order whose dynamic logic can be axiomatized in a similar manner. For instance, 'conservative upgrade' $\uparrow\varphi$ only puts the *most plausible* φ-worlds on top in the new model, leaving the rest in their old positions. For general results on complete logics, see [22], [10] and [12]. In particular, [117] is an excellent source for variety of policies in belief revision theory that is not tied to the specific dynamic logic methodology employed in this paper.

2.5 General dynamic methodology and its applications

We have spent quite some time on the above matters because they represent a general methodology of *model transformation* that works for many further phenomena, including changes in preference, and the even richer deontic scenarios that we will be interested in eventually.

Model transformations of relevance to agency can be much more drastic than what we have seen here, extending the domains of available worlds and modifying their relational structure accordingly. In the dynamic-epistemic logic of general observation DEL, different agents can have different access to the current informational event, as happens in card games, communication with security restrictions, or other social scenarios. This requires generalizing PAL as well as the above logics of belief change, using a mechanism of 'product update' to create more complex new models (cf. [9], [136], [12]).

Appropriately extended update mechanisms have been applied to many further aspects of agency: changes in intentions ([118], [80]), trust ([75]), inference ([137]), questions and inquiry ([19]), as well as complex scenarios in games ([105], [13]) and social information phenomena generally ([121], [7], [66]). Yet, in this paper, we will stick mainly with the much simpler pilot systems presented in the preceding sections.

3 Deontic logic as preference logic

Having set up the machinery for changing informational attitudes, we now turn to our next interest, the realm of normative evaluation for worlds or actions and the matching dynamic deontic logics. Here we will follow a perhaps not uncontroversial track: our treatment of deontic notions and scenarios will be based on *preference* structure and its changes. We believe that this is a conceptually good way of looking at deontic notions, and at the same time, it lends itself very well to treatment by our

[13] As before, it is easy to extend this analysis of soft upgrade to safe belief.

earlier methods, since at an abstract level, doxastic plausibility order and deontic betterness order are very similar. The results that follow in the coming sections are largely from [91], [54], and [94].[14]

Let us say a few more words about the connection between deontic logic and preference, to justify our approach in this paper. Deontic logic is the logical study of normative concepts such as obligation, prohibition, permission and commitment. This area was initiated by von Wright in [140] who introduced the logic of absolute obligation. As a reaction to paradoxes with this notion, conditional obligation was then proposed in [141], [143] and [47]. Good reviews systematizing the area are found in [3], [4].

One often thinks of deontic logic as the study of some accessibility relation from the actual world to the set of 'ideal worlds', but the more sophisticated view ([67], [48] and [81]) has models with a binary comparison relation that we may call 'betterness'.[15] Such more general comparisons make sense, for instance, when talking and reasoning about 'the lesser of two evils', or about 'improvement' of some given situation.

Naturally, this is precisely the ordering semantics that we have already seen for belief, and it would be tedious to indulge in formal definitions at this stage that the reader can easily construct for herself. Our base view would be that of binary *pre-orders* as before, for which we will now use the notation R to signal a change from the earlier plausibility interpretation. As usual, imposing further constraints on the ordering will generate deductively stronger deontic logics.

The binary relation R now interprets $O\varphi$ (absolute obligation) as φ *being true in all best worlds*, much like belief with respect to plausibility. Likewise, we interpret conditional obligation $O^{\psi}\varphi$ like conditional belief: φ holds *in the best* ψ*-worlds*.[16]

For further information on deontic logic, we refer to [4] and various chapters in the forthcoming Handbook [52]. Our emphasis in this paper will be mainly on interfacing with this field.

As we already noted at the start of this paper, deontic ordering shows intuitive analogies with the notion of *preference*. One can think of betterness as reflecting the preferences of a moral authority or law-giver, and in the happy Kantian case

[14]To unclutter notation, here and henceforth, we will mostly suppress agent indices for modal operators and their corresponding relations.

[15]Hansson argued that von Wright-type deontic logic can be naturally interpreted in terms of a preference relation 'is at least as ideal as' among possible worlds – an ordering that we will call 'betterness' in what follows.

[16]There are also more abstract neighborhood versions of this semantics, where the current proposition plays a larger role in terms of binary deontic betterness relations R^{ψ}, where one can set $\mathfrak{M}, s \models O^{\psi}\varphi$ iff for all t in W with $sR^{\psi}t, \mathfrak{M}, t \models \varphi$.

where agents' duties coincide with their inclinations, deontic betterness *is* in fact the agent's own preference. We claim no novelty for this line of thought, which was advocated forcefully as early as [67]. With this twist, we can then avail ourselves of existing studies of preference structure and evaluation dynamics, a line of thinking initiated in [134] and [135], though we now take the dynamic-epistemic road.

By way of background to what follows, we note that preference logic is a vigorous subject with its own history. For many new ideas and results in the area, we refer to [71] and [63]. What we will do next in this paper is survey some recent developments in the study of preference statics and dynamics, emphasizing those that are of relevance to deontic logic, an area where we will return eventually toward the end of this paper.[17]

4 Static preference logic

In the coming sections, we will discuss basic developments in modal preference logic, starting with its statics, and then continuing with the dynamics of preference change. Our treatment follows ideas from [33] and [65], and for the dynamics, we rely on [20] and [26].

4.1 General modal preference logic

Our basic models are like in decision theory or game theory: there is a set of alternatives (worlds, outcomes, objects) ordered by a primitive ordering that we dub 'betterness' to distinguish it from richer notions of preference.[18]

Definition 14. *A* modal betterness model *is a tuple* $\mathfrak{M} = (W, \preceq, V)$ *with* W *a set of worlds or objects,* $\preceq$ *a reflexive and transitive relation over these, and* V *is a valuation assigning truth values to proposition letters at worlds.*[19]

The order relation in these models also induces a strict variant $s \prec t$:

If $s \preceq t$ but not $t \preceq s$, then t is *strictly better* than s.

[17]Preference logic tends to focus on describing the agents' own preferences, rather than those of others, but what we have to say applies equally well to multi-agent settings such as moral scenarios, or games, where different preference orders interact in crucial ways.

[18]To repeat an earlier point, while each agent has her own betterness order, in what follows, merely for technical convenience, we suppress indices wherever we can.

[19]As we said before, we use pre-orders since we want the generality of possibly non-total preferences. Still, total orders, the norm in areas like game theory, provide an interesting specialization for the results in this chapter – but we will only mention it in passing.

Here is a simple modal language that can say a lot about these structures:

Definition 15. *Take any set of propositional variables Φ, with p ranging over Φ. The* modal betterness language *has this inductive syntax rule:*

$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle\leq\rangle\varphi \mid \langle<\rangle\varphi \mid E\varphi.$

The intended reading of $\langle\leq\rangle\varphi$ is "φ is true in a world that is at least as good as the current world", while $\langle<\rangle\varphi$ says that "φ is true in a world that is strictly better than the current world. ". In addition, the auxiliary *existential modality $E\varphi$* says that "there is a world where φ is true". As usual, we write $[\leq]\varphi$ for the defined universal modality $\neg\langle\leq\rangle\neg\varphi$, and we use $[<]$ and U for the duals of $\langle<\rangle\varphi$ and E, respectively. Combinations of these modalities can capture a wide variety of binary preference statements comparing propositions, witness the cited literature.

The interpretation of this modal language over our models is as follows:

Definition 16. *Truth conditions for the atomic propositions and Boolean combinations are standard. Modalities are interpreted like this:*

- $\mathfrak{M}, s \models \langle\leq\rangle\varphi$ *iff* *for some t wih $s \preceq t$, $\mathfrak{M}, t \models \varphi$.*
- $\mathfrak{M}, s \models \langle<\rangle\varphi$ *iff* *for some t with $s \prec t$, $\mathfrak{M}, t \models \varphi$.*
- $\mathfrak{M}, s \models E\varphi$ *iff* *for some world t in W, $\mathfrak{M}, t \models \varphi$.*

The defined modalities use the obvious universal versions of these clauses. For concreteness, we state the standard calculus to come out of this.

Theorem 17. *Modal betterness logic is completely axiomatized by*

1. *the system **S4** for the preference modality,*
2. *the system **S5** for the universal modality,*
3. *the connecting law $U\varphi \rightarrow [\preceq]\varphi$,*
4. *three axioms for the strict betterness modality: cf. [16].*

4.2 Special features of preference

Next we briefly survey three special logical features of preference structure that go beyond standard modal logic of pre-orders, and that will eventually turn out to be of interest to deontics as well.

Lifting to generic preferences. While betterness relates specific objects or worlds, preference is often used generically for comparing different *kinds* of things. Ever since [142], logicians have also studied preferences $P(\varphi, \psi)$ between propositions, viewed as properties of worlds, or of objects.

There is not one such notion, but many, that can be defined by a *lift* of the betterness order among worlds to sets of worlds, cf. [65], [16], [94]. For instance, compare your next moves in a game, identified with the set of outcomes that they lead to. Which move is 'better' depends on the criterion chosen: maybe we want the one with the highest possible outcome, or the one with the highest minimally guaranteed outcome, etcetera.

Such options are reflected in various quantifier combinations for the lifting. In particular, von Wright had a $\forall\forall$-type preference between sets P, Q:

$$\forall x \in P \; \forall y \in Q\colon \; x \preceq y.$$

A simpler, but also useful example is the modal $\forall\exists$-type

$$\forall x \in P \; \exists y \in Q\colon \; x \preceq y.$$

This says that for any P-world, there is a Q-world which is at least as good as that ψ-world. In the earlier game setting, this stipulation would say that the most preferred moves have the highest maximal outcomes. This ubiquitous $\forall\exists$ generic preference can be defined in the above modal preference language, using the universal modality ranging over all worlds:

$$P^{\forall\exists}(\varphi, \psi) := U(\psi \rightarrow \langle\leq\rangle\varphi).$$

This generic preference $P\varphi\psi$ satisfies the usual properties for preference, reflexivity and transitivity: for instance, $P\varphi\psi$ and $P\psi\chi$ imply $P\varphi\chi$.[20]

Ceteris paribus clauses. Unlike plausibility, preference ordering seldom comes in pure form: the comparison between alternatives is often entangled with other considerations. Again, games provide an example. Usually, players do not compare moves via the sets of all their possible outcomes, but rather, they compare the *most plausible* outcomes of their moves. This is the so-called *normality sense* of ceteris paribus preference: we do not compare all the φ and ψ-worlds, but only the 'normal ones' in some relevant sense. This belief restriction, observed by many authors, will return in our discussion of doxastic entanglement of preference in Section 8.

But there are also other natural senses of taking a ceteris paribus clause. It was noticed already in [142] that there is also an 'equality sense' of preference, involving

[20] Other stipulationss lead to other generic preferences. This proliferation may be a problem (e.g., 'doing what is best' depends on one's stipulation as to 'best'), but there is no consensus in the literature. A logical approach at least helps make the options clear.

a hidden assumption of *independence*. In that case, one only make comparisons between worlds where some things or issues are held constant, in terms of giving the same truth values to some specified set of atomic propositions, or complex formulas. The logic of equality-based preference is axiomatized and analyzed in detail in [16].

Richer preference languages. Modal languages are just one step on a ladder of formalisms for analyzing reasoning practices. It has been claimed that richer languages are needed to faithfully render basic preference notions, cf. [38] on first-order preferences among objects, [60] on first-order languages of social choice, [20] on hybrid modal preference languages for defining backward induction solutions in games, the hybrid modal language of 'desire' and 'freedom' for decision making in [64], or the modal fixed-point languages for games used in [13]. Though we will mainly use modal formalisms to make the essential points to follow, we will mention the relevance of such richer preference formalisms occasionally.

5 World based dynamics of preference change

Now let us look at how given preferences can change. Intuitively, there are many acts and events that can have such an effect. Perhaps the purest form is a radical *command* by some moral authority to do something. This makes the worlds where we act better than those where we do not, cf. [146]: at least, if we 'take' the order as a legitimate instruction, and change our evaluation accordingly, overriding any preferences that we ourselves might have had. Technically, this dynamics will change a current betterness relation in a model. This can be studied entirely along the lines already developed here for information dynamics.[21]

5.1 Betterness change

[26] is a first systematic study of betterness change using methods from dynamic-epistemic logic. The running example in their approach is a weak 'suggestion' $\sharp\varphi$ that a proposition φ be the case. This relatively modest ordering change leaves the set of worlds the same, but it removes any preferences that the agent might have had for $\neg\varphi$-worlds over φ-worlds among these.[22]

The main general point to note here is that events with evaluative import can act as triggers that change some current betterness relation on worlds. In particular, a suggestion $\sharp\varphi$ leads to the following model change:

[21] Of earlier treatments, we mention [135], based on [139].

[22] Similar operations have come up recently in logical treatments of relevant alternatives theories in epistemology, when modeling changes in what is considered relevant to making or evaluating a knowledge claim. Cf. [77], [24].

Definition 18. *Given any modal preference model* $(\mathfrak{M}, s)$*, the* suggestion upgrade $(\mathfrak{M}\sharp\varphi, s)$ *has the same domain, valuation, and actual world as* $(\mathfrak{M}, s)$*, but the new preference relations are now*

$$\preceq_i^* = \preceq_i - \{(s,t) \mid \mathfrak{M}, s \models \varphi \text{ and } \mathfrak{M}, t \models \neg\varphi\}$$

In preference models $\mathfrak{M}$, a matching dynamic modality is interpreted as:

$$(\mathfrak{M}, s) \models [\sharp\varphi]\psi \quad \text{iff} \quad \mathfrak{M}_{\sharp\varphi}, s \models \psi$$

Again, complete dynamic logics exist (cf. [26]). The reader will find it useful to scrutinize the key recursion law for preferences after suggestion.[23]

Theorem 19. *The dynamic preference logic of suggestion is completely axiomatized by the following principles:*

1. $\langle\sharp\varphi\rangle p \;\leftrightarrow\; p$
2. $\langle\sharp\varphi\rangle\neg\psi \;\leftrightarrow\; \neg\langle\sharp\varphi\rangle\psi$
3. $\langle\sharp\varphi\rangle(\psi \wedge \chi) \;\leftrightarrow\; (\langle\sharp\varphi\rangle\psi \wedge \langle\sharp\varphi\rangle\chi)$
4. $\langle\sharp\varphi\rangle\langle\leq\rangle\psi \;\leftrightarrow\; (\neg\varphi \wedge \langle\leq\rangle\langle\sharp\varphi\rangle\psi) \vee (\varphi \wedge \langle\leq\rangle(\varphi \wedge \langle\sharp\varphi\rangle\psi))$
5. $\langle\sharp\varphi\rangle E\psi \;\leftrightarrow\; E\langle\sharp\varphi\rangle\psi$

Similar completeness results are presented in [94] for dynamic logics that govern many other kinds of normative action, such as the 'strong commands' corresponding to our earlier radical plausibility upgrade. Following this instruction, deontically, the agent incorporates the wish of some over-riding authority.

5.2 Deriving changes in defined preferences

This is an analysis of betterness change and modal statements about it local to specific worlds. But it also applies to the earlier lifted *generic preferences*. As an illustration, consider the $\forall\exists$-lift defined earlier:

Fact 20. *The following equivalence holds for generic* $\forall\exists$ *preference:*

$$\langle\sharp A\rangle P^{\forall\exists}(\varphi, \psi) \quad \textit{iff} \quad P^{\forall\exists}(\langle\sharp A\rangle\varphi, \langle\sharp A\rangle\psi) \wedge P^{\forall\exists}((\langle\sharp A\rangle\varphi \wedge A), (\langle\sharp A\rangle\psi \wedge A)).$$

We omit the simple calculation for this outcome. Similar results may be obtained for other set liftings such as Von Wright's $\forall\forall$-version.

Finally, the recursive style of dynamic analysis presented here also applies to various forms of ceteris paribus preference.

[23] Technically, the simplicity of this law reflects the clear analogy between our universal preference modality and the earlier doxastic notion of safe belief.

5.3 General formats for betterness change

Behind our specific examples of betterness change, there lies a much more general theory that works for a wide class of triggering events that change betterness or evaluation order. One widely applicable way of achieving greater generality uses programs from *propositional dynamic logic PDL*.

For instance, suggesting that φ is defined by the program:

$$\sharp\varphi(R) := (?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi) \cup (?\neg\varphi; R; ?\varphi).$$

where R is the given input relation, while the operations $?\varphi$ test whether the relevant proposition φ, or related ones, hold. In particular, the disjunct $(?\varphi; R; ?\varphi)$ means that we keep all old betterness links that run from φ-worlds to φ-worlds.

This definition is equivalent in PDL to the more compact program expression

$$\sharp\varphi(R) := (?\neg\varphi; R) \cup (R; ?\varphi).$$

Again we keep all old R-links, except for those that ran from φ-worlds to $\neg\varphi$-worlds.

Likewise, our plausibility changers for belief revision can be defined in this format. For instance, the earlier 'radical upgrade' is defined by

$$\Uparrow\varphi(R) := (?\varphi; R; ?\varphi) \cup (?\neg\varphi; R; ?\neg\varphi) \cup (?\neg\varphi; \top; ?\varphi)$$

Here the constant symbol $\top$ denotes the universal relation that holds between any two worlds. This reflects the original meaning of this transformation: all φ-worlds become better than all $\neg\varphi$-worlds, whether or not they were better before, and within these two zones, the old ordering remains.[24]

Given any PDL program definition of the above sort, one can automatically write recursion laws for the complete dynamic logic of its induced model change, cf. [26] for the precise algorithm. As an illustration, here is the straightforward computation for suggestions:

$$\begin{aligned} \langle\sharp\varphi\rangle\langle R\rangle\psi &\leftrightarrow \langle(?\neg\varphi; R) \cup (R; ?\varphi)\rangle\langle\sharp\varphi\rangle\psi \\ &\leftrightarrow \langle ?\neg\varphi; R\rangle\langle\sharp\varphi\rangle\psi \vee \langle R; ?\varphi\rangle\langle\sharp\varphi\rangle\psi \\ &\leftrightarrow (\neg\varphi \wedge \langle R\rangle\langle\sharp\varphi\rangle\psi) \vee \langle R\rangle(\varphi \wedge \langle\sharp\varphi\rangle\psi). \end{aligned}$$

[24] Conservative upgrades can be dealt with in a similar way. As commands, these leave the agent more of her original preferences: so, differences with radical commands will show up in judgments of 'conditional betterness', as discussed in the literature on conditional obligation: see [67].

For alternative general formats of ordering change supporting our sort of dynamic logics, we refer to the 'priority update' with event models in [10], the order merge perspective of [21], as well as the still more general 'dynamic dynamic logic' of [55].

In our view, the practical and theoretical theoretical variety of ordering changes for plausibility and preference is not a nuisance, but a feature. It matches the wealth of evaluative actions that we encounter in daily life.

6 Reason-based preferences

Primitive betterness relations among worlds or objects reflect what are called 'intrinsic preferences'. But very often, our preferences have an underlying structure, and we compare according to criteria: our preferences are then reason-based, or 'extrinsic'. In this section we develop the latter view, that has motivations in linguistic Optimality Theory, cf. [112], and belief revision based on entrenchment, cf. [116]. This view also occurs in reason-based deontic logic, cf. [48], [56] and [81], as we shall see in Section 9.

A simplest illustration of our approach, that suffices for many natural scenarios, starts with linear orders of relevant properties that serve as criteria for determining our evaluation of objects or worlds.

6.1 Priority based preference

The following proposal has many ancestors, among which we mention the treatment in [49], [116]. We follow [38], that starts from a given primitive ordering among propositions ('priorities' among properties of objects or worlds), and then derives a preference among objects themselves.

Definition 21. *A* priority sequence *is a finite linear sequence of formulas written as follows:* $C_1 \gg C_2 \cdots \gg C_n \quad (n \in \mathbb{N})$*, where the* C_m *come from a language describing objects, with one free variable* x *in each* C_m*.*

Definition 22. *Given a priority sequence and objects* x *and* y*, Pref*(x, y) *is defined lexicographically: at the first property* C_i *in the given sequence where* x, y *have a different truth value,* $C_i(x)$ *holds, but* $C_i(y)$ *fails.*

The logic of this framework is analyzed in [38], while applications to deontic logic are developed in [25]. Still, this is only one of many ways of deriving a preference from a priority sequence. A good overview of existing approaches is found in [36].

6.2 Pre-orders

In general, comparison order need not be connected, and then the preceding needs a significant generalization. This was done, in a setting of social choice and belief merge, in the seminal paper [2], which we adapt slightly here to the notion of 'priority graphs', based on the treatment in [54], [95].

The following definitions contain a free parameter for a *language* L that can be interpreted in the earlier modal betterness models $\mathfrak{M}$. For simplicity only, we will take this to be a simple propositional language of properties.

Definition 23. *A* priority graph $\mathscr{G} = \langle P, < \rangle$ *is a strictly partially ordered set of propositions in the relevant language of properties* L.

Here is how one derives a betterness order from a priority graph:

Definition 24. *Let* $\mathscr{G} = \langle P, < \rangle$ *be a priority graph, and* $\mathfrak{M}$ *a model in which the language* L *defines properties of objects. The* induced betterness relation $\preceq_{\mathscr{G}}$ *between objects or worlds is defined as follows:*

$$y \preceq_{\mathscr{G}} x := \forall P{\in}\mathscr{G}((Py \to Px) \vee \exists P'{<}P(P'x \wedge \neg P'y)).$$

Here, in principle, $y \preceq_{\mathscr{G}} x$ requires that x has every property in the graph that y has. But there is a possibility of 'compensation': if y has P while x does not, this is admissible, provided there is some property P' with higher priority in the graph where x does better: x has P' while y lacks it. Clearly, this stipulation subsumes the earlier priority sequences: linear priority graphs lead to lexicographic order.

One can think of priority graphs of propositions in many ways that are relevant to this paper. In the informational realm, they are hierarchically ordered information sources, structuring the evidence for agents' beliefs. In the normative realm, they can stand for complex hierarchies of laws, or of norm givers with relative authority.

6.3 Static logic and representation theorem

In what follows, we immediately state a crucial technical property of this framework, cf. [49], [95].

Theorem 25. *Let* $\mathfrak{M} = (W, \preceq, V)$ *be any modal preference model, without constraints on its relation. The following two statements are equivalent:*

(*a*) *The relation* $y \preceq x$ *is a reflexive and transitive order,*
(*b*) *There is a priority graph* $\mathscr{G} = (P, <)$ *such that, for all worlds* $x, y \in W$, $y \preceq x$ *iff* $y \preceq_{\mathscr{G}} x$.

This representation theorem says that the general logic of derived extrinsic betterness orderings is still just that of pre-orders. But it also tells us that any intrinsic pre-order can be rationalized as an extrinsic reason-based one by adding structure without disturbing the base model as it is.

6.4 Priority dynamics and graph algebra

Now, we have a new locus for more fine-grained preference change: the family of underlying reasons, which brings its own logical structure. For linear priority sequences, relevant changes involve the obvious operations $[^{+}C]$ of adding a new proposition C to the right, $[C^{+}]$ of adding C to the left, and various functions $[-]$ dropping first, last or intermediate elements of a priority sequence. [38] give complete dynamic logics for these. Here is one typical valid principe:

$$[^{+}C]Pref(x,y) \leftrightarrow Pref(x,y) \vee (Eq(x,y) \wedge C(x) \wedge \neg C(y))$$

This set of natural operations for changing preferences becomes even richer in the realm of priority graphs, due to their possibly non-linear structure. However, in this setting an elegant mathematical alternative arises, in terms of merely two fundamental operations that combine arbitrary graphs:

- $\mathscr{G}_1;\mathscr{G}_2$ adding a graph to another in top position
- $\mathscr{G}_1\|\mathscr{G}_2$ adding two graphs in parallel.

One can think of this as the obvious counterparts of 'sequential' versus 'parallel' composition. Here the very special case where one of the graphs consists of just one proposition models simple update actions.

This graph calculus has been axiomatized completely in [2] by algebraic means, while [54] presents a further modal-style axiomatization. We display its major modal principles here, since they express the essential recursion underlying priority graph dynamics. Here is one case where, as mentioned earlier, a slight language extension is helpful: in what follows, the proposition letter n is a 'nominal' from hybrid logic denoting one single world.

$$\langle \mathscr{G}_1\|\mathscr{G}_2\rangle^{\leq} n \;\leftrightarrow\; \langle \mathscr{G}_1\rangle^{\leq} n \wedge \langle \mathscr{G}_2\rangle^{\leq} n.$$

$$\langle \mathscr{G}_1\|\mathscr{G}_2\rangle^{<} n \;\leftrightarrow\; (\langle \mathscr{G}_1\rangle^{<} n \wedge \langle \mathscr{G}_2\rangle^{\leq} n) \vee (\langle \mathscr{G}_1\rangle^{\leq} n \wedge \langle \mathscr{G}_2\rangle^{<} n).$$

$$\langle \mathscr{G}_1;\mathscr{G}_2\rangle^{\leq} n \;\leftrightarrow\; (\langle \mathscr{G}_1\rangle^{\leq} n \wedge \langle \mathscr{G}_2\rangle^{\leq} n) \vee \langle \mathscr{G}_1\rangle^{<} n.$$

$$\langle \mathscr{G}_1;\mathscr{G}_2\rangle^{<} n \;\leftrightarrow\; (\langle \mathscr{G}_1\rangle^{\leq} n \wedge \langle \mathscr{G}_2\rangle^{<} n) \vee \langle \mathscr{G}_1\rangle^{<} n.$$

These axioms reduce complex priority relations to simple ones, after which the whole language reduces to the modal logic of weak and strict atomic betterness orders. In particular, this modal graph logic is decidable.

Thus, we have shown how putting reasons underneath agents' preferences (or, for that matter, their beliefs) admits of precise logical treatment, while still supporting the systematic dynamics that we are after.

7 A two-level view of preference

Now we have two ways of looking at preference: one through intrinsic betterness order on modal models, the other through priority structure inducing extrinsic betterness orders. One might see this as calling for a reduction from one level to another, but instead, *combining* the two perspectives seems the more attractive option, as providing a richer modeling tool for preference-driven agency.

7.1 Harmony of world order and reasons

In many cases, the two modeling levels are in close harmony, allowing for easy switches from one to the other (cf. [91]):

Definition 26. *Let α: $(\mathscr{G}, A) \to \mathscr{G}'$, with $\mathscr{G}$, $\mathscr{G}'$ priority graphs, and let A be a new proposition. Let σ be a map from $(\preceq, A)$ to $\preceq'$, where $\preceq$ and $\preceq'$ are betterness relations over worlds. We say that α* induces *σ, if always:*

$$\sigma(\preceq_{\mathscr{G}}, A) \quad = \quad \preceq_{\alpha(\mathscr{G}, A)}$$

Here are two results that elaborate the resulting harmony between two levels for our earlier major betterness transformers:

Fact 27. *Taking a suggestion A is the map induced by the priority graph update $\mathscr{G} \| A$. More precisely, the following diagram commutes:*

$$\begin{array}{ccc}
\langle \mathscr{G}, < \rangle & \xrightarrow{\|A} & \langle (\mathscr{G} \| A), < \rangle \\
\downarrow & & \downarrow \\
\langle W, \preceq \rangle & \xrightarrow{\sharp A} & \langle W, \sharp A(\preceq) \rangle
\end{array}$$

For a second telling illustration of such harmony in terms of our earlier themes, consider a priority graph $(\mathscr{G}, <)$ with a new proposition A added on top. The logical dynamics at the two levels is now correlated as follows:

Fact 28. *Placing a new proposition* A *on top of a priority graph* $(\mathscr{G}, <)$ *induces the radical upgrade operation* $\Uparrow A$ *on possible worlds ordering models. More precisely, the following diagram commutes:*

$$\begin{array}{ccc} \langle \mathscr{G}, < \rangle & \xrightarrow{A;\mathscr{G}} & \langle (A;\mathscr{G}), < \rangle \\ \downarrow & & \downarrow \\ \langle W, \preceq \rangle & \xrightarrow{\Uparrow A} & \langle W, \Uparrow A(\preceq) \rangle \end{array}$$

Thus the two kinds of preference dynamics dovetail well: [94] has details.

7.2 Correlated dynamics

There are several advantages to working at both levels without reductions. For a start, not all natural operations on graphs have matching betterness transformers at all. An example from [95] is *deletion* of the topmost elements from a given priority graph. This syntactic operation of removing criteria is not invariant for replacing graph arguments by other graphs inducing the same betterness order, and hence it is a genuine extension of preference change.

But also conversely, there is no general match. Not all PDL-definable betterness changers from Section 5.3 are graph-definable. In particular, not all PDL transformers preserve the basic order properties of reflexivity and transitivity guaranteed by priority graphs. For a concrete illustration, consider the program

$?A; R$: ‘keep the old relation only from where A is true’.

This change does not preserve reflexivity of an order relation R, because the $\neg A$-worlds now have no outgoing relation arrows any more.[25]

All this argues for a more general policy of modeling both intrinsic and extrinsic preference for agents, with reasons for the latter encoded in priority graphs that are an explicit part of the modeling.

Still, one might think that intrinsic betterness relations merely reflect an agent’s raw feelings or prejudices. But the intrinsic-extrinsic contrast is relative, not absolute. If I obey the command of a higher moral authority, I may acquire an extrinsic preference, whose reason is obeying a superior. But for that higher agent, the same preference may be intrinsic: “The king’s whim is my law”. This observation suggests a further theme: transitioning from one perspective to the other.

[25] Intuitively, the operation $?A; R$ amounts to a refusal to make betterness comparisons at worlds that lack property A. Though idiosyncratic, this seems a bona fide mind change for an agent.

7.3 Additional dynamics: language change

Technically, intrinsic betterness can become extrinsic through a dynamics that has been largely outside the scope of dynamic-epistemic logic so far, that of *language change*. One mechanism here is the proof of the earlier representation result stated in Theorem 25. It partitions the given betterness pre-order into clusters, and if these are viewed as new relevant reasons or criteria, the resulting strict order of clusters is a priority graph inducing the given order. This may look like mere formal rationalization, but in practice, one often observes agents' preferences between objects, and then postulates reasons for them. A relevant source is the notion of 'revealed preference' from the economics literature: cf. [79].

Thus, our richer view of preference also suggests a new kind of dynamics beyond what we have considered so far. In general, reasons for given preferences may have to come from some other, richer language than the one that we started with: we are witnessing a dynamic act of *language creation*.[26]

8 Combining evaluation and information

We have now completed our exposition of information dynamics as well as preference dynamics, which brought its own further topics. What must have become abundantly clear is that there are strong formal similarities in the logic of order and order change in the two realms. We have not even enumerated all of these similarities, but, for instance, all of our earlier ideas and results about reason-based preference also make sense when analyzing evidence-based belief.

This compatibility helps with the next natural step we must take. As we said right at the start of this paper, the major agency systems of information and evaluation do not live in isolation: they interact all the time. A rational agent can process information well in the sense of proof or observation, but is also 'reasonable' in a broader sense of being guided by goals. This *entanglement* of knowledge, belief, and preference shows in many specific settings. We will look at a few cases, and in particular, their impact on the dynamics of preference change.[27] Though we will mainly discuss how information dynamics influences preference and deontic notions, the opposite influence is equally real. In particular, information flow depends on *trust* and *authority*: which are clearly deontic notions.[28]

[26]For a study of language change in the setting for belief revision, cf. [108].

[27]For a more general discussion, we refer to [109].

[28]Following Wittgenstein, Brandom (cf. [34]) has even argued that language use can only be fully understood in terms of commitments that carry rights and obligations.

8.1 Generic preference with knowledge

In Section 4.2, we defined one basic generic preference as follows:

$Pref^{\forall\exists}(\psi, \varphi) := U(\psi \to \langle\leq\rangle\varphi)$.

This refers to possibilities in the whole model, including even those that an agent might know to be excluded. [26] defend this scenario in terms of 'regret', but still, there is also a reasonable intuition that preference only runs among situations that are epistemically possible.

This suggests the entangled notion that, for any ψ-world that is *epistemically accessible* to agent a in the model, there is a world which is at least as good where φ is true. This can be written with an epistemic modality:

$$Pref^{\forall\exists}(\varphi, \psi) ::= K_a(\psi \to \langle\leq\rangle\varphi). \qquad (K_{bett})$$

But this is not yet what we are after, since we want the 'better world' to be epistemically accessible itself. [92] shows how this cannot be defined in a simple combined language of knowledge and betterness, and that instead, a richer preference formalism is needed with a new *intersection modality* for epistemic accessibility and betterness. The latter entangled notion can be axiomatized, and it also supports a dynamic logic of preference change as before.[29]

8.2 Generic preference with belief

Issues of entanglement become even more appealing with generic preference and belief, where the two relational styles of modeling were very similar to begin with. Again, we might start with a mere combination formula

$$Pref^{\forall\exists}(\varphi, \psi) ::= B_a(\psi \to \langle\leq\rangle\varphi). \qquad (B_{bett})$$

This says that, among the most plausible worlds for the agent, for any ψ-world, there exists a world which is at least as good where φ is true.[30]

Again, this seems not quite right in many cases, since we often want the better worlds relevant to preference to stay inside the most plausible part of the model, being 'informational realists' in our desires. To express this, we again need a stronger

[29] An alternative approach would be to impose *additional modal axioms* that require betterness alternatives to be epistemic alternatives via frame correspondence. However, this puts constraints on our dynamic operations transforming models that we have not investigated. We leave this alternative line as a topic for further investigation.

[30] One might also think here of using a *conditional belief* $B^{\psi}\langle\leq\rangle\varphi$, but to us, this seems an intuitively less plausible form of entanglement.

merge of the two relations by intersection. The key clause for a corresponding new modality then reads like a 'wishful safe belief':

$$\mathfrak{M}, s \models H\varphi \text{ iff } \textit{for all } t \textit{ with both } s \leq t \textit{ and } s \preceq t,\ \mathfrak{M}, t \models \varphi.$$

As before, the static and dynamic logic of this entangled notion yield to the general dynamic-epistemic methodology explained in earlier sections.

8.3 Other entanglements of preference and normality

Entangled versions of plausibility and betterness abound in the literature. For instance, [33] has models $\mathfrak{M} = (W, \leq_P, \leq_N, V)$ with W a set of possible worlds, V a valuation function and $\leq_P$, $\leq_N$ two transitive connected relations $x \leq_P y$ ('y is as good as x) and $x \leq_N y$ ('y is as normal as x). These models support an operator of *conditional ideal goal* (IG):

$$\mathfrak{M} \models IG^{\psi}\varphi \text{ iff } Max(\leq_P, Max(\leq_N, Mod(\psi))) \subseteq Mod(\varphi)$$

This says that the best of the most normal ψ worlds satisfy φ. Such entangled notions are still expressible in the modal systems of this chapter.

Fact 29. $IG^{\psi}\varphi ::= U(\psi \wedge \neg\langle B^{<}\rangle\psi) \wedge \neg\langle<\rangle(\psi \wedge \neg\langle B^{<}\rangle\psi) \rightarrow \varphi)$.[31]

Following up on this tradition in agency studies in computer science, the paper [87] defines the following entangled notion of preference:

Definition 30. $\mathfrak{M} \models Pref^{*}(\varphi, \psi)$ *iff for all* $w' \in Max(\leq_N, Mod(\psi))$, *there exists* $w \in Max(\leq_N, Mod(\varphi))$ *such that* $w' <_P w$.

This reflects the earlier-mentioned 'ceteris paribus' sense of preference, where one compares only the normal worlds of the relevant kinds.[32] Intriguingly, a source of similar ideas is the semantics of expressions like "want" and "desire" in natural language, cf. [128], [74], [37].

The preceding notions are similar to our earlier one with an intersection modality, but not quite. They only compare the two most plausible parts for each proposition.

We give no deeper analysis of all these entangled notions here, but as one small appetizer, we note that we are still within the bounds of this paper.

Fact 31. *$Pref^{*}$ is definable in a modal doxastic preference language.*

[31] Here, $B^{<}$ is an earlier-mentioned modality of *strong belief* that we do not define.

[32] This makes sense in epistemic game theory, where 'rationality' means comparing moves by their most plausible consequences according to the player's beliefs and then choosing the best.

8.4 Preference change and belief revision

As we have observed already, our treatment of the statics and dynamics of belief and preference shows many similarities. It is an interesting test, then, if the earlier dynamic logic methods transfer to entangled notions of preference. Intuitively, entangled preferences can change because of two kinds of trigger: evaluative acts like suggestions or commands, and informative acts changing our beliefs. As a positive illustration, we quote one result from [91]:

Theorem 32. *The dynamic logic of the above intersective preference H is axiomatizable, with the following essential recursion axioms:*

1. $\langle \sharp A\rangle\langle H\rangle\varphi \leftrightarrow (A \wedge \langle H\rangle(A \wedge \langle \sharp A\rangle\varphi)) \vee (\neg A \wedge \langle H\rangle\langle \sharp A\rangle\varphi)$.
2. $\langle \Uparrow A\rangle\langle H\rangle\varphi \leftrightarrow (A \wedge \langle H\rangle(A \wedge \langle \Uparrow A\rangle\varphi)) \vee (\neg A \wedge \langle H\rangle(\neg A \wedge \langle \Uparrow A\rangle\varphi)) \vee$ $(\neg A \wedge \langle bett\rangle(A \wedge \langle \Uparrow A\rangle\varphi))$.
3. $\langle A!\rangle\langle H\rangle\varphi \leftrightarrow A \wedge \langle H\rangle\langle A!\rangle\varphi$.

Having intersection modalities may not be all that is needed, though, since there may also be *entangled triggering events* that do not easily reduce to purely informational or purely evaluative actions.[33]

Trade-offs between preference change and information change. Finally, as often in logic, distinctions can get blurred through redefinition. For instance, sometimes, the same scenario may be modeled either in terms of preference change, or as information change. Two concrete examples of such redescription are "Buying a House" in [38] and "Visit by the Queen" in [88]. Important though it is, we leave the study of precise connections between different representations of dynamic entangled scenarios to another occasion.

9 Deontic reasoning, changing norms and obligations

Our analysis of information and preference can itself be viewed as a study of normative discourse and reasoning. However, in this section, we turn to explicit deontic scenarios, and take a look at some major issues concerning obligations and norms from the standpoint of dynamic systems for preference change.[34]

[33] For an analogy, see the question scenarios involving conversational triggers for parallel information and issue change in [19].

[34] Our treatment largely follows the papers [25], [17].

Perhaps the most immediate concrete task at hand is charting the large variety of deontic actions in daily life that affect normative betterness orderings. These normative triggers range from commands to promises and permissions. We will not undertake such a survey here, but the examples in this paper will hopefully convince the reader that a dynamic action perspective on deontic issues is natural, and that much can be done with the tools presented here. Instead, we consider four general topics that have roots in the deontic literature.

9.1 Unary and dyadic obligation on ordering models

Our static logics heavily relied on binary ordering relations. In fact, deontic logic was first with this approach, building on observations from ethics that the deontic notions of obligation, permission and prohibition can be naturally made sense of in terms of an *ideality ordering* $\preceq$ on possible worlds. Here is an early quote from [101], found in [48], p.6.

> " [...] to assert that a certain line of conduct is [...] absolutely right or obligatory, is obviously to assert that more good or less evil will exist in the world, if it is adopted, than if anything else be done instead."

In this line, the pioneering study [67] interpreted dyadic obligations of the type 'it is obligatory that φ under condition ψ' on semantic models like ours, using a notion of maximality as in our study of belief:

$$\mathfrak{M}, s \models \mathbf{O}(\varphi \mid \psi) \iff Max(||\psi||_{\mathfrak{M}}) \subseteq ||\varphi||_{\mathfrak{M}}$$

Depending on the properties of the relation $\preceq$, different deontic logics are obtained here: [67] starts with a $\preceq$ which is only reflexive, moving then to total pre-orders. This is of course the same idea that has also emerged in conditional logic, belief revision, and the linguistic semantics of generic expressions.[35] Variations of this modeling have given rise to various preference-based semantics of deontic logic: see [134] for an overview.

In this light, our paper has taken up an old idea in the semantics of deontic reasoning, and then added some recent themes concerning preference: criterion-based priority structure, dynamics of evaluative acts and events, and extended logical languages making these explicit. This seems a natural continuation of deontic logic, while also linking it up with developments in other fields.

[35] One deontic criticism of this account has been that it made conditional obligation lack the property of antecedent strengthening: [132]. This, however, makes perfect sense in our view, as it reflects precisely the non-monotonicity inherent in the dynamics of information change, where the most ideal worlds can change during update.

9.2 Reasons and dynamics in deontic paradoxes

The dynamic emphasis in this paper on changes and their triggering events has thrown fresh light on the study of information and preference-based agency. Deontic logic proves to be no exception, if we also bring in our treatment of reason-based preference – as we shall see with a few examples.

The Gentle Murder scenario from [46], p.194, is a classic of deontic logic that illustrates the basic problem of 'contrary-to-duty' obligations (CTDs).

Example 33. *"Let us suppose a legal system which forbids all kinds of murder, but which considers murdering violently to be a worse crime than murdering gently. [...] The system then captures its views about murder by means of a number of rules, including these two:*

1. *It is obligatory under the law that Smith not murder Jones.*
2. *It is obligatory that, if Smith murders Jones, Smith* [does so] *gently."*

The priority format of Section 6.1, even just linear sequences, can represent this scenario in a natural way. Recall that a linear priority sequence $P_1, \ldots, P_n$ combines bipartitions $\{\mathcal{I}(p_i), -\mathcal{I}(p_i)\}$ of the domain of discourse S. Moving towards the right direction of the sequence, ever more atoms p_i are falsified. In a deontic reading, this means that, the more we move towards the right side of the sequence, the more violations hold of morally desirable properties.

Concretely, in the Gentle Murder scenario, the result is two classes of ideality: one class I_1 in which Smith does not murder Jones, i.e., $\mathsf{I}_1 := \neg m$; and another I_2 in which either Smith does not murder Jones or he murders him gently, i.e., $\mathsf{I}_2 := \neg m \vee (m \wedge g)$. The relevant priority sequence $\mathcal{B}$ has $\mathsf{I}_2 \prec \mathsf{I}_1$. Such a sequence orders the worlds via its induced relation $\preceq^{IM}_{\mathcal{B}}$ in three clusters. The most ideal states are those satisfying I_1, worse but not worst states satisfy $\mathsf{V}_1 := \neg\mathsf{I}_1$ but at the same time I_2, and, finally, the worst states satisfy $\mathsf{V}_2 := \neg\mathsf{I}_2$.

With this representation, we can take the scenario one step further.

Example 34. *Consider the priority sequence for Gentle Murder from the preceding Example: $\mathcal{B} = (\mathsf{I}_1, \mathsf{I}_2)$. We can naturally* restrict *$\mathcal{B}$ to an occurrence of the first violation by intersecting all formulas in the sequence with V_1. Then the first proposition becomes a contradiction, distinguishing no worlds. The best among the still available worlds are those with $Max^+(\mathcal{B}^{\mathsf{V}_1}) = \mathsf{I}_2 \wedge \mathsf{V}_1$. A next interesting restriction is $\mathcal{B}^{\mathsf{V}_2}$, which represents what the original priority sequence prescribes under the assumption that also the CTD obligation "kill gently" has been violated. In this case we end up in a set of states that are all equally bad.*

This brief sketch may suffice to show our approach provides a simple perspective on the deontic robustness of norms and laws viewed as *CTD* structures: they can still function when transgressions have taken place.[36]

Other major puzzles in the deontic literature, such as the Chisholm Paradox, are given similar reason-based representations in [17].

9.3 Typology of change at two levels

We have shown how two-level structure of preference provides a natural medium for modeling deontic notions. Likewise, it yields a rich account of deontic changes. In Section 7, we developed a theory of both informational and evaluative changes, either directly on possible world order, or on priority structure underlying such orders. This also makes sense here.

As an illustration, we add a temporal twist to the above deontic scenario, by 'dynamifying' Gentle Murder.

Example 35. *We start with a priority sequence $\mathcal{B} = (\neg m)$. This current deontic state of affairs generates a total pre-order where all $\neg m$-states are above all m-states: "It is obligatory under the law that Smith not murder Jones". Now, we refine this order so as to introduce the sub-ideal obligation to kill gently: "it is obligatory that, if Smith murders Jones, Smith murders Jones gently". In other words, we want to model the process of refining legal codes, by introducing a contrary-to-duty obligation.*

Intuitively, this change can happen in one of two ways:

1. *We refine the given betterness ordering 'on the go' by requesting a further bipartition of the violation states, putting the $m \wedge g$-states above the $m \wedge \neg g$-states. This can be seen as the successful execution of a command of the sort "if you murder, then murder gently".*

2. *We introduce a new law 'from scratch', where $m \to g$ is now explicitly formulated as a class of possibly sub-ideal states. This can be seen as the enactment of a new priority sequence $(\neg m, m \to g)$.*[37]

The example illustrates how a *CTD* sequence can be dynamically created either by uttering a sequence of commands stating what ought to be the case in a sub-ideal situation, or by enacting a new priority sequence.

[36]Representing *CTD* structures as finite chains of properties already occurs informally in [48]. The first formal account is in [57], where an elegant Gentzen calculus is developed for handling formulae of the type $\varphi_1 @ \ldots @ \varphi_n$ with @ a connective representing a sort of 'sub-ideality' relation. It is an interesting open problem if such a proof calculus can be embedded in the modal logics of this chapter.

[37]We have encountered this before, since $m \to g$ is equivalent to $\neg m \vee (m \wedge g)$.

But in this setting, Theorem 27 from Section 7 applies: in terms of betterness among worlds, the two instructions amount to the same thing! In other words, in this scenario, the same deontic change can be obtained both by refining the order dictated by a given law, and by enacting a new law.

Of course, this is just a start, and not everything is smooth application. Our discussion of two-level dynamics in Section 7.2 also suggests that some well-known changes in laws, such as *abrogation* (a counterpart to the earlier operation of 'graph deletion') have no obvious counterpart at the pure worlds level.

9.4 Norm change

The preceding discussion leads up to a more general theme of global dynamics. The problem of *norm change* has recently gained attention from researchers in deontic logic, legal theory, as well as multi-agent systems.

Approaches to norm change fall into two groups. In syntactic approaches—inspired by legal practice—norm change is an operation performed directly on the explicit provisions in the code of the normative system [58], [59], [31]. In semantic approaches, however, norm change follows deontic preference order (cf. also [6]). Our initial betterness dynamics on models belonged to the latter group, but our priority methods tie it to the former.[38]

More drastic changes of norms and moral codes can be modeled, too, in our framework, using the calculus of priority graphs that we have sketched in Section 6. For details, we refer again to [17].

9.5 Entangled changes

Finally, as observed already in Section 8 on entanglement (cf. [87] for a deontic discussion), the dynamic logic connection allows for a unified treatment of *t*wo kinds of change that mix harmoniously in deontic reasoning: information change given a fixed normative order, and evaluation change modifying such an order. Deontic scenarios can have deeply intertwined combinations of obligation, knowledge and belief (cf. [94]). Some sophisticated moral scenarios in [106] include natural dynamic issues that we have ignored here, such as the subtle, but real difference between 'knowing one's duty' versus 'having a duty to know'.

Many further dynamic deontic themes can be analyzed along the above lines. We refer to [25], [17] for a detailed treatment of the Chisholm Paradox, and concrete ways in which priority graph calculus models norm change.

[38] The bridge here is our earlier analysis: obligations defined via ideality and maximality are special kinds of classifications of an Andersonian-Kangerian type.

10 Further directions

Collecting points from earlier sections, here are a few further directions where deontic logic meets with current trends in dynamic logics of agency.

10.1 Language, speech acts, and agency

Events that drive information or preference change are often *speech acts* of telling, asking, and so on. Natural language has a sophisticated repertoire of speech acts with a deontic flavour (commanding, promising, allowing, and so on) that invite further logical study, taking earlier studies in meta-ethics and Speech Act Theory (cf. [119]) to the next level. In particular, such studies will also need a more fine-grained account of the *multi-agency* in dynamic triggers, that has been ignored in this chapter. For instance, things are said by someone to someone, and their uptake depends on relations of authority or trust. Likewise, promises, commands, or permissions are given by someone to someone, and their normative effect depends in subtle ways on who does, and is, what. [148] is a pioneering study of this fine-structure of normative action using dynamic-epistemic logic.

10.2 Multi-agency and groups

A conspicuous turn in studies of information dynamics has been a strong emphasis on social scenarios with more than one agent: [12], [121], [7], [66]. After all, the natural paradigm for language use is communication between different agents, a major historical source for logic is argumentation between different parties, social behaviour is kept in place by mutual expectations, and so on.

In the logics for knowledge, belief, and preference of this paper, part of this multi-agent turn can be represented by mere iteration of single-agent modalities, as in a's knowing that b does, or does not, knows some fact. But the next stage is the introduction of *groups* as agents, where logics have been devised for notions such as 'common knowledge' or 'distributed knowledge' in groups, and likewise for beliefs (cf. [45], [100]), or the group-level preferences underlying Social Choice Theory (cf. [44]). All these logics also have dynamic-epistemic extensions in the style of this chapter, although systematic extensions to, say, social choice or judgement aggregation remain to be developed.

The social turn is highly relevant to deontic logic. From the start, deontic notions and morality seems all about *others*: my duties are usually toward other people, my norms come from outside sources: my boss, or a lawgiver.[39] In principle, the methods

[39]This social aspect has been clearly acknowledged by computer scientists working on multi-agent

of this paper can deal with social multi-agent structure in deontic settings, though much remains to be investigated. For instance, it is easy to interpret informational iterations $K_a K_b p$, but what, for instance, is the meaning of an iterated obligation $O_a O_b p$? And beyond this, what is a group-based 'common obligation': is this more like common belief, or like a demand for joint action of the group? Other relevant issues are the entanglement of informational and evaluative acts for groups: cf. [73], [84], [83], and [75] on morality as held together by social expectations such as trust. An account of deontically relevant actions for groups will also have to include new operations reminiscent of social choice, such as *belief merge* and *preference merge*, where the priority structures of Section 6 may find a new use: this time, as a model for institutions (cf. [61]).

10.3 Games and dependent behaviour

Multi-agency is tied together not just by social knowledge or beliefs, but also by dependent individual and collective *action*. Thus, logics of agency have close connections with game theory ([125], [12]) and the study of strategic behavior and its equilibria. In deontic practice, dependent action is crucial (think of sanctions or rewards), and games are a congenial paradigm. Many topics in this paper suggest game-theoretic analogies. We already saw how belief-entangled set lifting is crucial to player's choices and their rationality, making preference logics a natural tool in the analysis of games (cf. [118], [39], [13]). Conversely, ideas from game theory have entered deontic logic, witness the use of game solution methods as moral deliberation procedures in [96]. One might even argue that dependent behaviour is the source of morality, and in that sense, games would be the really natural next stage after the single-episode driven dynamic logics of this paper.

10.4 Temporal perspective

Games are one longer-term activity, but deontic agency involves many different processes, some even infinite. The general logical setting here are temporal logics (cf. [45], [110]) where new phenomena come to the fore. Deontics and morality is not just about single episodes, but about action and interaction over time. Early work in deontic logic already used temporal logics: cf. the pioneering dissertation [43]) where events happen in infinite histories, and obligations come and go. Likewise, in the multi-agent community, logics have been proposed for preferences between complete histories, and planning behaviour leading toward most desired histories (cf. [98], [122]). Such temporal logics mesh well with dynamic-epistemic logics (cf.

systems: cf. [99], [145], and [114].

[14]), with an interesting role for *protocols* as a new object of study, i.e., available procedures that both provide and constrain available actions for reaching goals. Plans and protocols have a clear normative dimension as well, and one would wish to incorporate them into the preference dynamics of this paper.

10.5 Syntax and fine-structure

Most dynamic logics for agency, whether about information dynamics or evaluation dynamics, are semantic in nature. The states changed by the process are semantic models. Still, in philosophical logic, there is a continuing debate about the right representation of *information*. Semantic information, though common to many areas, including decision theory and game theory, is coarse-grained, identifying logically equivalent propositions, suppressing the very act of logical analysis as an information-producing process. Zooming in on the latter, agents engage in many activities, such as inference, memory retrieval, introspection, or other forms of 'awareness management' that require a more fine-grained notion of information, closer to syntax. Several dynamic logics of this kind have been proposed in recent years (cf. [18], [82], [137]).

The same issues of grain level for information make sense in the deontic realm. For instance, our priority graphs were syntactic objects than get manipulated by insertions, deletions, permutations, and the like. But also, deontic logic has its own counterpart to the epistemological problem of 'omniscience'. My moral obligations to you cannot reasonably be based on my foreseeing every consequence of my commitments. I owe you careful deliberation, not omniscience.[40] Here too, there is a need for more fine-grained dynamic representations, closer to deontic syntax.

10.6 Numerical strength

While the main theme of this chapter is qualitative approaches, there are also numerical approaches to preferences, employing utilities (cf. [115], [133]) or more abstract 'grades' for worlds (cf. [127]). Dynamic ideas work in this setting, too, witness the modal logic with graded modalities indicating the strength of preference in [5], which also defines product update for numerical plausibility models. A stream-lined version in [90] uses propositional constants q_a^m saying that agent a assigns the current world a value of at most m. Our earlier ordering models, both for plausibility and for preference, now get numerical graded versions, with more finely-grained statements of strength of belief and of preference. Dynamic updates can still be defined,

[40] Likewise, citizens are supposed to know the law, but they need not be professional lawyers in seeing every relevant deductive consequence.

where we assign values to actions or events, using numerical stipulations in terms of 'product update' from the cited references.[41] More complex numerical evaluation uses *utility* as a fine-structure of preference, and its dynamics can also be dealt with in this style: cf. [90], [93].

While the technical details of these approaches are not relevant here, systems like this do address two issues that seem of great deontic relevance. One is the possibility of comparing not just worlds qua preference, but also actions, making sense of the principled distinction in ethics between outcome-oriented and deontological views of obligations and commitments. The other major feature is that we can now study the more quantitative logic of *how much good* an action does, and the extent to which we can *improve* current situations by our actions.

10.7 Probability

Another obvious quantitative addition to our analysis would be *probability.* Probabilities measure strengths of beliefs, thereby providing fine-structure to the plausibility orderings that we have worked with. But they can also indicate information that we have about a current process, or a reliability we assign to our observation of a current event.[42] Finally, the numerical factors in probability theory also allow us to mix and weigh various factors in the entangled versions of preference and deontic notions that we have discussed in Section 8. A striking entangled notion is *expected value* in probability theory, whose definition mixes beliefs and evaluation. A treatment of such notions in our current framework remains a desideratum.

11 Appendix: relevant strands in the literature

The themes of this paper have a long history, with many proposals in the literature for combining and 'dynamifying' preferences, beliefs, and obligations. In addition to those cited already, here are some other relevant lines of work.

Computation and agency. [99] is a pioneering study of deontics from a dynamic viewpoint, reducing deontic logics to suitable dynamic logics. In the same tradition, [98] takes the deontic logic/dynamic logic interface a step further, studying 'free choice permission' with a new dynamic logic where preferences can hold between actions. Completeness theorems for this enriched semantics then result for several systems. [113] provide a dynamified logic of permission that builds action policies for

[41] The resulting dynamic logic of numerical evaluation can be axiomatized in the same recursive style as the qualitative systems that we have discussed in this paper.

[42] See [15] for a rich dynamic-epistemic logic of reasoning with and updating probability.

agents by adding or deleting transitions. [40] reduces an extension of van der Meyden's logic to PDL, yielding an EXPTIME decision procedure, and showing how PDL can deal with agents' policies. Preference semantics has also been widely used in AI tasks: e.g., [144] gives a preference-based semantics for goals in decision theory. This provides criteria for verifying the design of goal-based planning strategies, and a new framework for knowledge-level analysis of planning systems. [78] studies commonsense normative reasoning, arguing that techniques of non-monotonic logic provide a better framework than the usual modal treatments. The paper has applications to conflicting obligations and conditional obligations. [87] propose a logic of desires whose semantics contains two ordering relations of preference and normality, and then interpret "in context A, I desire B" as 'the best among the most normal $A \wedge B$ worlds are preferred to the most normal $A \wedge \neg B$ worlds', providing a new entanglement of preference and normality.

Semantics of natural language. In a line going back to [127], [139] presents an update semantics for default rules, locating their meaning in the way in which they modify expectation patterns. This is part of a general program of 'update semantics' for conditionals and other key expressions in natural language. [135] use ideas from update semantics to formalize deontic reasoning about obligations. In their view, the meaning of a normative sentence resides in the changes it brings about in the 'ideality relations' of agents to whom a norm applies. [149] uses a simple dynamic update logic to formalize natural language imperatives of the form $FIAT\ \varphi$, which can be used in describing the search for solutions of planning problems. [97] extends the update semantic analysis of imperatives to include third person and past tense imperatives, while also applying it to the notion of free choice permission. [107] outlines a preference-based account of communication, which brings the dynamics of changing obligations for language users to the fore. [147] distinguishes the illocutionary acts of commanding from the perlocutionary acts that affect preferences of addressees, proposes a new dynamic logic which combines preference upgrade and deontic update, and discusses some deontic dilemmas in this setting.

Philosophical logic. The philosophical study of agency has many themes that are relevant to this paper, often inspired by topics in epistemology or by the philosophy of action. In a direction that is complementary to ours, with belief change as a starting point, [70] identifies four types of changes in preference, namely revision, contraction, addition and subtraction, and shows that they satisfy plausible postulates for rational changes. The collection [63] brings together the latest approaches on preference change from philosophy, economics and psychology. Following Hansson's work, [1] defines minimal preference change in the spirit of AGM framework and characterises minimal contraction by a set of postulates. A linear time algorithm is proposed for computing preference changes. In addition, going far beyond what we have discussed in this paper, Hansson has written a series of seminal papers combining ideas from preference logic and deontic logic, see e.g. [69], [68] and [72].

Rational choice theory. Preference is at the heart of decision and rational choice. In recent work at the interface of preference logic, philosophy, and social science, themes from our chapter such as reason-based and belief-entangled preference have come to the fore, with further lines of their own. [42] and [41] point out that, though existing decision theory gives a good account of how agents make choices given their preferences, issues of where these essential preferences come from and how they can change are rarely studied.[43] The authors propose a model in which agents' preferences are based on 'motivationally salient properties' of alternatives, consistent sets of which can be compared using a 'weighing relation'. Two intuitive axioms are identified in this setting that precisely characterize the property-based preference relations. Starting from similar motivations, [102] studies reason-based preference in more complex doxastic settings, drawing on ideas from similarity-based semantics for conditional logic. Essentially, preference results here from agents' comparing two worlds, one having some property and the other lacking it, close to their actual world, and comparing these based on relevant aspects of utility. The framework supports extensive analysis in modal logic, including illuminating results on frame correspondence and axiomatization. [103] gives an extension of this approach to preference in the presence of quantifiers, while [104] makes a link between these preference models and deontic logic. A detailed comparison of the two mentioned recent approaches with the one in this paper remains to be undertaken.

12 Conclusion

We have shown how dynamic logics of agency can deal with information, criteria, and preference change. In doing so, we obtained a suggestive framework for the analysis of deontic notions that connects many strands in the literature on agency.

References

[1] N. Alechina, F. Liu, and B. Logan. Minimal preference change. In D. Grossi, O. Roy, and H. Huang, editors, *Proceedings of the 4th International Workshop on Logic, Rationality and Interaction (LORI 2013)*, volume 8196 of *FoLLI-LNCS*, pages 15–26. Springer, 2013.

[43] These are of course precisely the two main topics of this paper: cf. also [94].

[2] H. Andréka, M. Ryan, and P-Y. Schobbens. Operators and laws for combining preferential relations. *Journal of Logic and Computation*, 12:12–53, 2002.

[3] L. Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*. Naples: Bibliopolis, 1987.

[4] L. Åqvist. Deontic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 605–714. Dordrecht: Kluwer, 1994.

[5] G. Aucher. A combined system for update logic and belief revision. Master's thesis, MoL-2003-03. ILLC, University of Amsterdam, 2003.

[6] G. Aucher, D. Grossi, A. Herzig, and E. Lorini. Dynamic context logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 15–26. Springer, 2009.

[7] A Baltag, Z. Christoff, J.U. Hansen, and S. Smets. Logical models of informational cascades. In J. van Benthem and F. Liu, editors, *Logic Across the University: Foundations and Application*, pages 405–432. College Publications, London, 2013.

[8] A. Baltag, N. Gierasimczuk, and S. Smets. Belief revision as a truth tracking process. In *Proceedings of Theoretical Aspects of Rationality and Knowledge(TARK 2011)*, 2011.

[9] A. Baltag, L.S. Moss, and S. Solecki. The logic of common knowledge, public announcements, and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 98)*, pages 43–56, 1998.

[10] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In M. Wooldridge G. Bonanno, W. van der Hoek, editor, *Logic and the Foundations of Game and Decision Theory*, volume 3 of *Texts in Logic and Games*. Amsterdam: Amsterdam University Press, 2008.

[11] A. Baltag, S. Smets, and J. Zvesper. Keep 'hoping' for rationality: A solution to the backward induction paradox. *Synthese*, 169(2):301–333, 2009.

[12] J. van Benthem. *Logical Dynamics Information And Interaction*. Cambridge University Press, 2011.

[13] J. van Benthem. *Logic in Games*. The MIT Press, 2014.

[14] J. van Benthem, J. Gerbrandy, T. Hoshi, and E. Pacuit. Merging frameworks for interaction. *Journal of Philosophical Logic*, 38(5):491–526, 2009.

[15] J. van Benthem, J. Gerbrandy, and B. Kooi. Dynamic update with probabilities. *Studia Logica*, 93(1):67–96, 2009.

[16] J. van Benthem, P. Girard, and O. Roy. Everything else being equal: A modal logic approach for ceteris paribus preferences. *Journal of Philosophical Logic*, 38(1):83–125, 2009.

[17] J. van Benthem, D. Grossi, and F. Liu. Priority structures in deontic logic. *Theoria*, 80(2):116–152, 2014.

[18] J. van Benthem and M. Martínez. The stories of logic and information. In J. van Ben-

them and P. Adriaans, editors, *Handbook of Philosophy of Information.* Amsterdam: Elsevier, 2008.

[19] J. van Benthem and S. Minica. Toward a dynamic logic of questions. In X. He, J. F. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 27–41. Springer, 2009.

[20] J. van Benthem, S. van Otterloo, and O. Roy. Preference logic, conditionals and solution concepts in games. In H. Lagerlund, S. Lindström, and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, pages 61–77. Uppsala Philosophical Studies 53, 2006.

[21] J.van Benthem. Belief update as social choice. In P. Girard, O. Roy, and M. Marion, editors, *Dynamic Formal Epistemology*, pages 151–160. Springer, Dordrecht, 2006.

[22] J.van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, 17:129–156, 2007.

[23] J.van Benthem. *Modal Logic for Open Minds.* Stanford: CSLI Publications, 2010.

[24] J.van Benthem. Talking about knowledge. To appear in C. Baskent, L. Moss and R. Ramanujam, eds., Volume in Honor of Rohit Parikh, Outstanding Contributions to Logic, Springer, Dordrecht, 2014.

[25] J.van Benthem, D. Grossi, and F. Liu. Deontics = betterness + priority. In G. Governatori and G. Sartor, editors, *Deontic Logic in Computer Science, 10th International Conference, DEON 2010*, volume 6181 of *LNAI*, pages 50–65. Springer, 2010.

[26] J.van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17:157–182, 2007.

[27] J.van Benthem and E. Pacuit. Dynamic logics of evidence-based beliefs. *Studia Logica*, 99:61–92, 2011.

[28] J.van Benthem and S. Smets. Dynamic logic of belief change. To appear in H. van Ditmarsch, J. Halpern, W. van der Hoek & B. Kooi, eds., Handbook of Logics of Knowledge and Belief, College Publications, London, 2014.

[29] P. Blackburn, J. van Benthem, and F. Wolter. *Handbook of Modal Logic.* Elsevier, 2007.

[30] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic.* Cambridge: Cambridge University Press, 2001.

[31] G. Boella, G. Pigozzi, and L. van der Torre. Normative framework for normative system change. In P.Decker, J.Sichman, C.Sierra, and C.Castelfranchi, editors, *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 169–176, 2009.

[32] C. Boutilier. *Conditional Logics for Default Reasoning and Belief Revision.* PhD thesis, University of Toronto, 1992.

[33] C. Boutilier. Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68:87–154, 1994.

[34] R. Brandom. *Making it Explicit: Reasoning, Representing, and Discursive Commit-*

ment. Harvard University Press, 1994.

[35] J. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 89–133. Dordrecht: D. Reidel, 1984.

[36] S. Coste-Marquis, J. Lang, P. Liberatore, and P. Marquis. Expressive power and succinctness of propositional languages for preference representation. In *Proceedings of the 9th International Conference on Principles of Knowledge Representation and Reasoning (KR 2004)*. Menlo Park, CA: AAAI Press, 2004.

[37] S. Dandelet. Partial desires, blinkered beliefs. Manuscript, January 22, 2014.

[38] D. de Jongh and F. Liu. Preference, priorities and belief. In T.Grune-Yanoff and S.O. Hansson, editors, *Preference Change: Approaches from Philosophy, Economics and Psychology*, Theory and Decision Library, pages 85–108. Springer, 2009.

[39] C. Dégremont. *The Temporal Mind. Observations on the Logic of Belief Change in Interactive Systems*. PhD thesis, ILLC, University of Amsterdam, 2010.

[40] S. Demri. A reduction from DLP to PDL. *Journal of Logic and Computation*, 15:767–785, 2005.

[41] F. Dietrich and C. List. A reason-based theory of rational choice. *Nous*, 47(1):104–134, 2013.

[42] F. Dietrich and C. List. Where do preferences come from? *International Journal of Game Theory*, 42(3):613–637, 2013.

[43] J. van Eck. *A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications*. PhD thesis, University of Groningen, 1981.

[44] U. Endriss. Applications of logic in social choice theory. In *Proceedings of the 12th International Workshop on Computational Logic in Multiagent Systems (CLIMA-2011)*, volume 6814 of *LNAI*, pages 88–91. Springer-Verlag, July 2011. Extended abstract corresponding to an invited talk.

[45] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. Cambridge, MA: The MIT Press, 1995.

[46] J. Forrester. Gentle murder, or the adverbial samaritan. *Journal of Philosophy*, 81:193–197, 1984.

[47] B. van Fraassen. The logic of conditional obligation. *Journal of Philosophical Logic*, 1:417–438, 1972.

[48] B. van Fraassen. Values and the heart's command. *The Journal of Philosophy*, 70(1):5–19, 1973.

[49] N. Friedman and J. Halpern. Plausibility measures: A user's guide. In *Proc. Eleventh Conf. on Uncertainty in Artificial Intelligence (UAI 95)*, pages 175–184, 1995.

[50] N. Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems, part I: Foundations. *Artificial Intelligence*, 95(2):257–316, 1997.

[51] N. Friedman and Joseph Y. Halpern. Modeling belief in dynamic systems. part II: Revision and update. *Journal of Artificial Intelligence Research*, 10:117–167, 1999.

[52] D. Gabbay, J.Horty, R. van der Meyden, and L. van der Torre. Handbook on Deontic Logic and Normative Systems, 2014. To appear.

[53] J. Gerbrandy. *Bisimulation on Planet Kripke.* PhD thesis, ILLC, University of Amsterdam, 1999.

[54] P. Girard. *Modal Logics for Belief and Preference Change.* PhD thesis, Stanford University, 2008.

[55] P. Girard, J. Seligman, and F. Liu. General dynamic dynamic logic. In Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence S. Moss, editors, *Advances in Modal Logic*, pages 239–260. College Publications, 2012.

[56] L. Goble. Multiplex semantics for deontic logic. *Nordic Journal of Philosophical Logic*, 5(2):113–134, 2000.

[57] G. Governatori and A. Rotolo. Logic of violations: A gentzen system for reasoning with contrary-to-duty obligations. *Australasian Journal of Logic*, 3:193–215, 2005.

[58] G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. part 1: Revision and defeasible theories. In R. van der Meyden and L. van der Torre, editors, *Proceedings of the 9th International Conference on Deontic Logic in Computer Science (DEON 2008)*, volume 5076 of *LNAI*, pages 3–18. Springer, 2008.

[59] G. Governatori and A. Rotolo. Changing legal systems: Abrogation and annulment. part 2: Temporalised defeasible logic. In G. Boella, G. Pigozzi, M. P. Singh, and H. Verhagen, editors, *Proceedings of the 3rd International Workshop on Normative Multiagent System (NorMAS 2008), Luxembourg, Luxembourg, July 14-15, 2008*, pages 112–127, 2008.

[60] U. Grandi and U. Endriss. First-order logic formalisation of arrow's theorem. In *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI-2009)*, volume 5834 of *LNAI*, pages 133–146. Springer-Verlag, October 2009. Also presented at DGL-2009.

[61] D. Grossi. *Desigining Invisible Hancuffs. Formal Investigations in Institutions and Organizations for Multi-Agent Systems.* PhD thesis, Utrecht University, 2007. SIKS Dissertation Series 2007-16.

[62] A. Grove. Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[63] T. Grune-Yanoff and S.O. Hansson, editors. *Preference Change: Approaches from Philosophy, Economics and Psychology.* Theory and Decision Library. Springer, 2009.

[64] M. Guo and J. Sliegman. Making choices in social situations. In *Logic and Interactive Rationalityok*, pages 176–202. ILLC, University of Amsterdam, 2011.

[65] J.Y. Halpern. Defining relative likelihood in partially-ordered preferential structure. *Journal of Artificial Intelligence Research*, 7:1–24, 1997.

[66] P. G. Hansen and V. F. Hendricks. *Infostorms: How to Take Information Punches and Save Democracy.* Copernicus Books / Springer, 2014.

[67] B. Hansson. An analysis of some deontic logics. *Nous*, 3:373–398, 1969.

[68] S.O Hansson. Defining 'good' and 'bad' in terms of 'better'. *Notre Dame of Journal of Formal Logic*, 31:136–149, 1990.

[69] S.O Hansson. Preference-based deontic logic. *Journal of Philosophical Logic*, 19:75–93,

1990.

[70] S.O Hansson. Changes in preference. *Theory and Decision*, 38:1–28, 1995.

[71] S.O Hansson. Preference logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 319–393. Dordrecht: Kluwer, 2001.

[72] S.O Hansson. *The Structure of Values and Norms.* Cambridge: Cambridge University Press, 2001.

[73] G. den Hartog. *Wederkerige Verwachtingen [Mutual Expectations].* PhD thesis, University of Amsterdam, 1985.

[74] I. Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9(3):183–221, 1992.

[75] W. H. Holliday. Dynamic testimonial logic. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality, and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 161–179. Springer, 2009.

[76] W. H. Holliday. *Knowing what follows: epistemic closure and epistemic logic.* PhD thesis, Stanford University, 2012.

[77] W. H. Holliday. Epistemic closure and epistemic logic i: Relevant alternatives and subjunctivism. *Journal of Philosophical Logic*, 2014.

[78] J. Horty. Deontic logic as founded on nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, 9:69–91, 1993.

[79] D. Houser and R. Kurzban. Revealed preference, belief, and game theory. *Economics and Philosophy*, 16:99–115, 2002.

[80] T. Icard III, E. Pacuit, and Y. Shoham. Joint revision of beliefs and intentions. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)*, pages 572 – 574. AAAI Publications, 2010.

[81] F. Jackson. On the semantics and logic of obligation. *Mind*, XCIV:177–195, 1985.

[82] M. Jago. *Logics for Resource-Bounded Agents.* PhD thesis, University of Nottingham, 2006.

[83] J. Konkka. Funk games: Approaching collective rationality. E-Thesis, University of Helsinki, 2000.

[84] B. Kooi and A. Tamminga. Conflicting obligations in multi-agent deontic logic. In John-Jules Ch. Meyer and Lou Goble, editors, *Deontic Logic and Artificial Normative Systems: 8th International Workshop on Deontic Logic in Computer Science*, volume 4048 of *LNCS*, pages 175–186. Springer, 2006.

[85] P. Lamarre. S4 as the conditional logic of nonmonotonicity. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 357–367, 1991.

[86] P. Lamarre and Y. Shoham. Knowledge, certainty, belief, and conditionalisation abbreviated version. In *KR'94*, pages 415–424, 1994.

[87] J. Lang, L. van der Torre, and E. Weydert. Hidden uncertainty in the logical representation of desires. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, pages 189–231, 2003.

[88] J. Lang and L. van der Torre. From belief change to preference change. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI-2008)*, pages 351–355, 2008.

[89] D. Lewis. *Counterfactuals.* Oxford: Blackwell, 1973.

[90] F. Liu. Dynamic variations: Update and revision for diverse agents. Master's thesis, MoL-2004-05. ILLC, University of Amsterdam, 2004.

[91] F. Liu. *Changing for the Better: Preference Dynamics and Agent Diversity.* PhD thesis, ILLC, University of Amsterdam, 2008.

[92] F. Liu. Logics for interaction between preference and belief. Manuscript, Department of Philosophy, Tsinghua University, 2009.

[93] F Liu. Preference change: A quantitative approach. *Studies in Logic*, 2(3):12–27, 2009.

[94] F. Liu. *Reasoning about Preference Dynamics*, volume 354 of *Synthese Library.* Springer, 2011.

[95] F. Liu. A two-level perspective on preference. *Journal of Philosophical Logic*, 40:421–439, 2011.

[96] L. O. Loohuis. Obligations in a responsible world. In X. He, J. Horty, and E. Pacuit, editors, *Proceedings of the 2nd International Workshop on Logic, Rationality and Interaction (LORI 2009)*, volume 5834 of *FoLLI-LNAI*, pages 251–262. Springer, 2009.

[97] R. Mastop. *What Can You Do? Imperative Mood in Semantic Theory.* PhD thesis, ILLC, University of Amsterdam, 2005.

[98] R. van der Meyden. The dynamic logic of permission. *Journal of Logic and Computation*, 6:465–479, 1996.

[99] J-J.Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

[100] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for Computer Science and Artificial Intelligence.* Cambridge: Cambridge University Press, 1995.

[101] G. E. Moore. *Principia Ethica.* Cambridge University Press, 1903.

[102] D. Osherson and S. Weinstein. Preference based on reasons. *The Review of Symbolic Logic*, 5:122–147, 3 2012.

[103] D. Osherson and S. Weinstein. quantified preference logic. Manuscript, October 5, 2012.

[104] D. Osherson and S. Weinstein. Deontic modality based on preference. Manuscript, July 4, 2014.

[105] S. van Otterloo. *A Strategic Analysis of Multi-agent Protocols.* PhD thesis, Liverpool University, UK, 2005.

[106] E. Pacuit, R. Parikh, and E. Cogan. The logic of knowledge based on obligation. *Synthese*, 149:311–341, 2006.

[107] X. Parent. Remedial interchange, contrary-to-duty obligation and commutation. *Journal of Applied Non-Classical Logics*, 13(3/4):345–375, 2003.

[108] R. Parikh. Beliefs, belief revision, and splitting languages. In J. Ginzburg, L. Moss, and M. de Rijke, editors, *Logic, Language and Computation*, volume 2, pages 266–278. Center for the Study of Language and Information Stanford, CA, 1999.

[109] R. Parikh, L. O. Loohuis, and C. Baskent. Epistemic norms. To appear in Dov Gabbay, John Horty, Ron van der Meyden and Leon van der Torre eds, Handbook on Deontic Logic and Normative Systems, College Publications, London, 2011.

[110] R. Parikh and R. Ramanujam. A knowledge-based semantics of messages. *Journal of Logic, Language and Information*, 12:453–467, 2003.

[111] J.A. Plaza. Logics of public communications. In M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems: Poster Session Program*, pages 201–216, 1989.

[112] A. Prince and P. Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell, 2004.

[113] R. Pucella and V. Weissmann. Reasoning about dynamic policies. In *Proceedings FoSSaCS-7*, Lecture Notes in Computer Science 2987, pages 453–467, 2004.

[114] A.S. Rao and M.P. Georgeff. Modeling rational agents within a BDI-architecture. In J. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484. San Mateo, CA: Morgan Kaufmann, 1991.

[115] N. Rescher. Notes on preference, utility, and cost. *Synthese*, 16:332–343, 1966.

[116] H. Rott. Basic entrenchment. *Studia Logica*, 73:257–280, 2003.

[117] H. Rott. Shifting priorities: Simple representations for 27 iterated theory change operators. In H. Langerlund, S. Lindström, and R. Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, pages 359–384. Uppsala Philosophical Studies 53, 2006.

[118] O. Roy. *Thinking before Acting: Intentions, Logic and Rational Choice*. PhD thesis, ILLC, University of Amsterdam, 2008.

[119] J. R. Searle and D. van der Veken. *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press, 1985.

[120] K. Segerberg. The basic dynamic doxastic logic of AGM. In M.-A. Williams and H. Rott, editors, *Frontiers in Belief Revision*, pages 57–84. Kluwer Academic Publishers, 2001.

[121] J. Seligman, F. Liu, and P. Girard. Facebook and the epistemic logic of friendship. In Burkhard. C. Schipper, editor, *Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 229–238, 2013.

[122] M. Sergot. $(C+)^{++}$: An action language for modelling norms and institutions. Technical Report 8, Department of Computing, Imperial College, London, 2004.

[123] C. Shi. Logics of evidence-based belief and knowledge. Master's thesis, Tsinghua University, 2014.

[124] Y. Shoham. *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge, MA: The MIT Press, 1988.

[125] Y. Shoham and K. Leyton-Brown. *Multiagent Systems: Algorithmic, Game Theoretic and Logical Foundations.* Cambridge: Cambridge University Press, 2008.

[126] J. Snyder. Product update for agents with bounded memory. Manuscript, Department of Philosophy, Stanford University, 2004.

[127] W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W.L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change and Statistics II*, pages 105–134. Dordrecht: Kluwer, 1988.

[128] R. Stalnaker. *Inquiry.* Cambridge University Press: Cambridge, United Kingdom, 1984.

[129] R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12(2):133–163, 1996.

[130] R. Stalnaker. Extensive and strategic forms: Games and models for games. *Research in Economics*, 53:293–319, 1999.

[131] R. Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128(1):169–199, 2006.

[132] Y.-H. Tan and L. van der Torre. How to combine ordering and minimizing in a deontic logic based on preferences. In M. Brown and J. Carmo, editors, *Deontic Logic, Agency and Normative Systems, DEON '96: The 3rd International Workshop on Deontic Logic in Computer Science*, pages 216–232, 1996.

[133] R.W. Trapp. Utility theory and preference logic. *Erkenntnis*, 22:301–339, 1985.

[134] L. van der Torre. *Reasoning about Obligations: Defeasibility in Preference-based Deontic Logic.* PhD thesis, Rotterdam, 1997.

[135] L. van der Torre and Y.-H. Tan. An update semantics for deontic reasoning. In P. McNamara and H. Prakken, editors, *Norms, Logics and Information Systems*, pages 73–90. Amsterdam: IOS Press, 1999.

[136] H. Van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic.* Berlin: Springer, 2007.

[137] F. R. Velazquez-Quesada. Inference and update. *Synthese*, 169:283–300, 2009.

[138] F. Veltman. *Logics for Conditionals.* PhD thesis, University of Amsterdam, 1985.

[139] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.

[140] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[141] G. H. von Wright. A note on deontic logic and derived ogligation. *Mind*, 65:507–509, 1956.

[142] G. H. von Wright. *The Logic of Preference.* Edinburgh: Edinburgh University Press, 1963.

[143] G. H. von Wright. A new system of deontic logic. In *Danish Yearbook of Philosophy*, pages 173–182. Museum Tusculanum Press, Copenhagen, 1964.

[144] M. Wellman and J. Doyle. Preferential semantics for goals. In *Proceedings of the National Conference on Artificial Intelligence*, pages 698–703, 1991.

[145] M. Wooldridge. *Reasoning about Rational Agents.* Cambridge, MA: The MIT Press, 2000.

[146] T. Yamada. Acts of commands and changing obligations. In K. Inoue, K. Satoh, and F. Toni, editors, *Proceedings of the 7th Workshop on Computational Logic in Multi-Agent Systems (CLIMA VII)*, 2006. Revised version appeared in LNAI 4371, pages 1-19, Springer-Verlag, 2007.

[147] T. Yamada. Logical dynamics of some speech acts that affect obligations and preferences. *Synthese*, 165(2):295–315, 2008.

[148] T. Yamada. Scorekeeping and dynamic logics of speech acts. Manuscript, Hokkaido University, 2010.

[149] B. Zarnic. Imperative change and obligation to do. In K. Segerberg and R. Sliwinski, editors, *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Aqvist*, pages 79–95. Uppsala philosophical studies 51. Uppsala: Department of Philosophy, Uppsala University, 2003.

Received July 2014