Von Wright's *"The Logic of Preference"* Revisited

Fenrong Liu*

Abstract. Preference is a key area where analytic philosophy meets philosophical logic. I start with two related issues: reasons for preference, and changes in preference, first mentioned in von Wright's book *The Logic of Preference* but not thoroughly explored there. I show how these two issues can be handled together in one dynamic logical framework, working with structured two-level models, and we investigate the resulting dynamics of reason-based preference in some detail. Next, we study the foundational issue of entanglement between preference and beliefs, and relate the resulting richer logics to belief revision theory and decision theory.

Analytic philosophy is well-known for its emphasis on clarity in philosophical formulation and argument, and modern formal logic and informal analysis of language go together well as tools for this purpose. Philosophical logics are often constructed based on philosophical analysis of a certain concept, but then in return, new logics become new tools suggesting new philosophical notions. I am interested in a harmonious interplay between analytic philosophy and philosophical logic, where ideas can flow both ways. In this paper, I will take the case of preference as a typical example, a crucial notion guiding rational choice and action. Starting with von Wright's work, *The Logic of Preference*, where he gave a philosophical analysis of preference and also a first logical system that became very influential, I hope to show how modern logic can add further technical sophistication, while also raising new philosophical questions.

1. Introduction: two issues left out by von Wright

Preference is what colors our view of the world, and it drives the actions that we take in it. Moreover, we influence each other's preferences all the time by making evaluative statements, uttering requests, commands, and statements of fact that exclude or open up the possibility of certain actions. Naturally, a phenomenon of this wide importance has been studied in many disciplines, especially in philosophy and the social sciences. This paper takes a further formal point of view, being devoted to logical systems that describe both the structure of preferences, and how they may change. Our starting point is in the days when

* Department of Philosophy, School of Humanities and Social Science, Tsinghua University, China; Institute for Logic, Language and Computation, University of Amsterdam, Netherlands.

preference was first fully discussed by the famous philosopher-logician Georg Henrik von Wright.

In his seminal book *The Logic of Preference: An Essay* from 1963, von Wright started with a major distinction among the concepts that interest moral philosophers. He divided them into the following three categories (though, of course, there may be border-line cases):

- *deontological* or *normative*: right and duty, command, permission and prohibition,
- *axiological*: good and evil, the comparative notion of betterness,
- *anthropological*: need and want, decision and choice, motive, and action.

The intuitive concept of preference itself was said to 'stand between the latter two groups of notions'. It is related to the axiological notion of betterness, but also to the anthropological notion of choice.

Considering the relationship between preference and betterness, von Wright distinguished two kinds of relation: *extrinsic* and *intrinsic* preference. He explains the difference with a nice example:

> "... a person says [...] that he prefers claret to hock, because his doctor has told him, or he has found from experience that the first wine is better for his stomach or health in general. In this case a *judgement of betterness serves as a ground or reason* for a preference. I shall call preferences which hold this relationship to betterness *extrinsic*.
>
> It could, however, also be the case that a person prefers claret to hock, not because he thinks (opines) that the first wine is better for him, but simply because he likes the first better (more). Then his liking the one wine better is not a reason for his preference. ..."
>
> (Von Wright 1963, p.14)

Simply stated, the difference is principally that p is preferred *extrinsically* to q if it is preferred *because* it is better in some explicit respect. If there is no such reason, the preference is intrinsic.

Instead of making the notion of betterness the starting-point of his inquiry,[1] von Wright took a more "primitive" intrinsic notion of preference as 'the point of departure', providing a formal system for it which has generated a whole subsequent literature (cf. (Han01)).

I am by no means claiming that the division between intrinsic and extrinsic preference is the only natural way of distinguishing preferences. One can also study varieties of moral preference, aesthetic

[1] (Hal57) did propose logic systems for the notion of betterness.

preference, economic preference, etc. However, in this context, I will follow von Wright. Our first main goal is to extend the literature on intrinsic preferences with formal logical systems for the *extrinsic notion of preference*, allowing us to spell out the reasons for a preference. On the way there, I will also make new contributions to the literature on intrinsic preferences.

Besides the extrinsic notion of preference that was removed from von Wright's agenda, there is another important issue which he decided to leave open. More precisely, he writes the following:

> "The preferences which we shall study are a subject's intrinsic preferences on one occasion only. Thus we exclude both *reasons* for preferences and the possibility of *changes* in preferences."
>
> (Von Wright 1963, p.23)

But clearly, our preferences are not static! One may *revise* one's preferences for many legitimate (and non-legitimate) reasons. Thus, the second main issue dealt with here is how to model preference change in formal logics. This leads to new dynamic versions of existing preference logics, and interesting connections with belief revision theory.

Following von Wright's work, formal investigations on preference have been mainly carried out by philosophical logicians. The best survey up to 2001 can be found in the Chapter *Preference Logic* by Sven Ove Hansson in the *Handbook of Philosophical Logic*. In addition, the notion of preference has been extensively discussed in decision theory, game theory and Artificial Intelligence. Even so, over the past decades, the above two major issues that (Wri63) left out, viz. *reason-based extrinsic preference*, and the *dynamics of preference change*, have received little attention. It seems fair to say that most authors took the notion of intrinsic preference only, and concentrated on its properties.[2] Moreover, we only found few papers treating changes in preference as such. (BEF93) is a first attempt at using dynamic logic for this purpose. Also, influenced by *AGM*-style belief revision theory, (Han95) proposed postulates for four basic operations in preference change.

Against this background, the present paper will show how these two crucial aspects of reasoning with preference can be treated in a uniform logical framework. Two recent proposals will be reviewed. Along the way, I will develop a new structured model for extrinsic preference and I will show how to relate dynamics at different levels in such models. Preference does not live alone, it has an intimate relationship with beliefs, which will be illustrated too. In what follows, I will concentrate on the main ideas, only providing technical details that are needed for

[2] Still, 'reasons for preference' are an acknowledged theme in decision theory and economics, witness the brief survey in (HGY06).

our discussion. For proofs, and many other related issues, I refer to the recent works (BL07), (JL08), (Liu08a) and (Ben08).[3]

2. Basic modal preference logic

Preference in (Wri63) is a relation between states of affairs, and (BL07) follows this tradition. In formal logical languages, states of affairs are typically represented as propositions, with the latter viewed as sets of possible worlds in standard Kripke models. Reinterpreting the usual accessibility relations over possible worlds as 'betterness' relations, we immediately obtain a model for a modal preference logic. More precisely, we have:

DEFINITION 2.1. A *betterness model* is a triple $\mathcal{M} = (W, \preceq, V)$, where W is a set of worlds, $\preceq$ is a reflexive and transitive binary 'betterness' relation $\preceq$ ('at least as good as'), and V is a valuation function for proposition letters. $\lhd$

Here, we may think of the 'reason' for preference as the betterness relation between possible worlds, which is given as primitive - though we will give a more finely-grained take on 'reasons' for preference below. Let us now define the simplest possible formalism for making considerations about preference fully explicit.

DEFINITION 2.2. Let P be a set of propositional variables P ($p \in P$). The modal *betterness language* is given by this inductive syntax:

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid [bett]\varphi \mid U\varphi.$$

$\lhd$

The semantic interpretation is as usual for such a language. In particular, the modality $[bett]\varphi$ says that 'locally', all worlds that are at least as good as the current one satisfy φ. [4] U is an auxiliary 'universal modality' quantifying over all worlds.

By standard techniques, there is a complete axiomatization for this system, based on the standard $S4$-axioms for the betterness modality:

- $[bett]\varphi \to \varphi$.
- $[bett]\varphi \to [bett][bett]\varphi$.

[3] The latter paper surveys recent dissertations and papers in the area.

[4] Incidentally, a formally similar operator has been introduced in belief revision theory under the name of 'stable' or 'safe' belief: cf. (BS06) and and (SLB08).

One can argue whether these properties are intuitively reasonable. There has been much discussion in the literature about what are the best conditions on the betterness relation: some authors want to add connectedness in addition to reflexivity and transitivity, others want to drop even the transitivity. This is not our concern here, since we see these issues as orthogonal to our main purposes in what follows.

A more fundamental issue concerns the relation between betterness and preference. As we said above, the intuitive notion of preference is perhaps best viewed as a relation between propositions, i.e., *types of situations* (modeled by sets of possible worlds), while betterness is a relation over possible worlds (the way it functions in decision theory or game theory where one compares *single situations*). Hence there is an issue of moving systematically from the betterness level in our models to the level of preference, a process called *lifting* in the literature. There are various ways of lifting, determined typically by quantifier combinations (cf. Section 5 below). Each has its own intuitions, and earlier discussions can be found in (Wri63), (Hal97). For instance, when you say "I prefer φ over ψ", it could mean "*all* possible worlds where φ holds are better than (or maybe: at least as good as) *all* possible worlds where ψ holds"; but it could also mean "for *all* possible worlds where φ holds, there is *some* better possible world where ψ holds". This lifting can still be expressed in our modal language, as illustrated in the definition for the second '∀∃-lift':

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

Essentially this is a comparison between ψ-worlds and φ-worlds in the current model, and our logic provides a complete system for reasoning about this notion.

Having laid this groundwork, I now move on to the two new issues from the previous section, and show how they fit into a logical framework. This will happen step-by-step in the next two sections.

3. Reasons for preference

In many situations, it is quite natural to ask for a reason when someone states her preference to you. It may be a matter of justification to herself, but also, you simply want more explanation or information (say, in order to judge whether it is rational for her, or for you, to have that preference). So preference can come with a reason, and this is what von Wright called 'extrinsic preference'. Let us return to the example used by (Wri63) to explain this notion:

EXAMPLE 3.1. A person prefers claret to hock, *because* his doctor has told him, or he has found from experience that the first wine is better for his stomach or health in general.

Here, the first wine being better for his health is the reason for his preference of claret to hock. Similar examples abound in real life: one prefers some house over another *because* the first is cheaper, and/or of better quality than the second.

Conceptually, reasons stand at a different level from preferences, and they form a usual information-based *ground* for their justification. Reasons can be of various kinds: from general principles to more 'object-oriented' facts. In many cases, one can combine more than one reason to justify one single preference. Thus, in the house example, not only the price of the house matters, but also the quality. In such cases, reasons may have their own structure, and different considerations may be ordered according to their importance. One may think for instance that the quality of a house is more important than its price. All these ideas and intuitions will govern the proposal we are now going to introduce.

The structured model proposed in (JL08) starts from a given order of importance among propositions (viewed as 'priorities' in making comparisons), and then derives an order among the underlying objects.[5] Priorities are relevant properties of the objects, and preference between objects is determined by the properties they have. The core notion here is that of a *priority sequence*, a strict order of properties:

$$P_1(x) \gg P_2(x) \gg \cdots \gg P_n(x)$$

On the basis of this structure, a preference order $\preceq$ over objects can be *derived*. As with earlier discussion of 'lifting', there are various ways to perform this 'lowering', and I will discuss a few options in Section 5 below. Here I state only one option, inspired by linguistic Optimality Theory (OT), saying that earlier priorities in the given sequence count strictly more than later ones:

DEFINITION 3.2. Given a priority sequence of length n, strict preference between two objects x and y, *Pref(x,y)*, is defined as follows:

$Pref_1(x, y) ::= C_1(x) \wedge \neg C_1(y)$,
$Pref_{k+1}(x, y) ::= Pref_k(x, y) \vee (Eq_k(x, y) \wedge C_{k+1}(x) \wedge \neg C_{k+1}(y)), k < n$,
$Pref(x, y) ::= Pref_n(x, y)$,

[5] These objects may be compared to the 'worlds' in our earlier setting.

where the auxiliary binary predicate $Eq_k(x, y)$ is the conjunction of equivalences $(C_1(x) \leftrightarrow C_1(y)) \wedge \cdots \wedge (C_k(x) \leftrightarrow C_k(y))$.[6] $\lhd$

To understand the definition, let us consider the following example:

EXAMPLE 3.3. Alice is going to buy a house. In doing so, she has several things to consider: the cost, the quality, and the neighborhood. She has the following priority sequence:

$$C(x) \gg Q(x) \gg N(x),$$

with $C(x)$, $Q(x)$ and $N(x)$ for 'x has low cost', 'x is of good quality' and 'x has a nice neighborhood'. Consider two houses d_1 and d_2 with the following properties: $C(d_1), C(d_2), \neg Q(d_1)$, $\neg Q(d_2)$, $N(d_1)$ and $\neg N(d_2)$. According to the OT-definition, Alice prefers d_1 over d_2.

OT-derived preference is a quasi-linear order: reflexive, transitive, and also 'connected', as described in the following axioms:

(a) $\underline{Pref}(d_i, d_i)$,
(b) $\underline{Pref}(d_i, d_j) \vee \underline{Pref}(d_j, d_i)$,
(c) $\underline{Pref}(d_i, d_j) \wedge \underline{Pref}(d_j, d_k) \rightarrow \underline{Pref}(d_i, d_k)$.

Note that we are using a (fragment of a) *first-order language* here, with a binary relation symbol $Pref$. This seems a natural formalism here, rather than the earlier modal language, though the two are still closely related (cf. (BRV01), (Liu08a)). The above axioms are complete, and (JL08) has many further technical results, including the following 'representation theorem':

THEOREM 3.4 (Representation theorem for reason-based preference). $\vdash_{\mathbf{P}} \varphi$ *iff* φ *is valid in all models OT-derived from priority sequences.*

This theorem nails down the relationship between preference and priorities. If we have a model in which the preference relation between objects behaves in a certain manner, then this preference may be derived from some priority sequence. This formal result has a direct connection to our philosophical intuitions: an agent can always *find a reason* for her preferences, given that they are 'rationally ordered'.

4. Dynamics of preference

But agents change, reasons change, and so do their preferences. To discuss these dynamics, let us continue our Example 3.1. With a twist of the imagination, we make it more realistic:

[6] This way of deriving an ordering from a priority sequence also occurs in earlier literature, and it is called the *leximin ordering* in (CMLLM04).

> Suppose that, before he sees the doctor, he *preferred hock to claret.* Now the doctor tells him "the first wine is better for your health". He then *changes* his preference, and will now *prefer claret to hock*!

Such scenarios are common in real life. Suggestions, commands and other triggers continually change our preferences on the basis of new information that they bring. A formal study of this phenomenon requires a logical system which can deal with *changing models* of the standard 'static' sort we had before. We will now show how this can be done, using some recent logical techniques.

4.1. Betterness relation change and preference change

Let us first say a few words on 'logical dynamics' in general. There is a growing tradition in logic modeling changes in epistemic or doxastic states. Agents receive new information and update their knowledge or beliefs accordingly. This style of thinking can be traced back to the early 1980s, e.g., the well-known *AGM* postulates handling belief change ((AGM85)). In what follows, we will work with a recent approach called *dynamic epistemic logic DEL*(cf. (Pla89), (Vel96), (Ben96), (BMS98), (Ger99), (Liu04) and (DHK07), as well as recent work on belief revision by (Ben07) and (BS08)). Preference change can be dealt with by using the same methodology of modeling information dynamics, as we will now show, following (BL07).

First, to speak about preference dynamics, we extend the static language of the previous section with dynamic modalities as follows:

DEFINITION 4.1. Let P be a set of proposition letters, with $p \in P$. The *dynamic betterness language* is given by the following syntax rule:

$$\varphi ::= \bot \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid [bett]\varphi \mid U\varphi \mid [\pi]\varphi$$
$$\pi ::= \sharp\varphi.$$

$\triangleleft$

Here $\sharp\varphi$ is one new dynamic action of "suggesting φ", which we take as our running example - even though our approach in this paper is much more general.[7] First, we define the new model $\mathcal{M}_{\sharp\varphi}$, as the result of a *relation change* over the current domain of worlds or objects (the difference is immaterial, as we have observed before):

DEFINITION 4.2. Given any preference model $(\mathcal{M}, s)$, the *upgraded model* $(\mathcal{M}_{\sharp\varphi}, s)$ is defined as follows:

[7] We can also add the usual program operations of composition, choice, and iteration from propositional dynamic logic to create complex betterness-changing actions - but we have no special use for these in this paper.

(a) $(\mathcal{M}_{\sharp\varphi}, s)$ has the same domain of worlds, propositional valuation, and actual world as $(\mathcal{M}, s)$, but

(b) the new preference relations are now
$\preceq_i^* = \preceq_i - \{(s,t) \mid \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, t \models \neg\varphi\}$.

$\lhd$

As we said, "suggestion A" ($\sharp A$) is just one way how a betterness relation may change: we drop cases where a $\neg\varphi$-world was preferred over a φ-world. Greater generality is achieved by observing that this can be described in a standard propositional dynamic language PDL:

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

Many further relation changes have this format: cf. Section 5.2 below.

We can now interpret the new operator $[\sharp\varphi]\psi$ as follows:

DEFINITION 4.3. Given a betterness model $\mathcal{M}$, the *truth definition* for formulas is as before, but with the following additional key clause for the action modality, 'looking ahead' at the changed model:

$(\mathcal{M}, s) \models [\sharp\varphi]\psi$ iff $\mathcal{M}_{\sharp\varphi}, s \models \psi$.

$\lhd$

The logical theory of this dynamic model for betterness change is developed in (BL07), using DEL-style methods of definable model change and matching complete axiomatizations. The key idea here is the recursive analysis of what agents know, or believe, or in our case: find better after some model change has taken place. The relevant 'recursion equations' are captured by so called *reduction axioms*.

In the case of preference, the key is the following reduction axiom for changes in our betterness modalities after 'suggesting that φ':

$$\langle\sharp\varphi\rangle\langle bett\rangle\psi \leftrightarrow (\neg\varphi \wedge \langle bett\rangle\langle\sharp\varphi\rangle\psi) \vee (\langle bett\rangle(\varphi \wedge \langle\sharp\varphi\rangle\psi)).$$

This axiom shows the exact betterness changes after a suggestion comes in, stated in terms of betterness relations that held before. Similar principles can be stated for the universal modality under suggestions, and also, for betterness changes that occur under other triggers, say, when new information comes in the form of a public announcement or a public observation.

As an additional benefit, this analysis automatically extends to our earlier 'lifted preference' between propositions. Once we have a reduction axiom for the operator $[bett]$, plus one for the universal modality

U, we can just derive a reduction axiom for the propositional preference operator $Pref^{\forall\exists}$. To illustrate this, we give the calculation:

$$\langle\sharp A\rangle Pref^{\forall\exists}(\varphi,\psi) \leftrightarrow \langle\sharp A\rangle U(\psi \to \langle bett\rangle\varphi)$$
$$\leftrightarrow U(\langle\sharp A\rangle(\psi \to \langle bett\rangle\varphi)) \leftrightarrow U(\langle\sharp A\rangle\psi \to \langle\sharp A\rangle\langle bett\rangle\varphi)$$
$$\leftrightarrow U(\langle\sharp A\rangle\psi \to (\neg A \wedge \langle bett\rangle\langle\sharp A\rangle\varphi) \vee (\langle bett\rangle(A \wedge \langle\sharp A\rangle\varphi)))$$
$$\leftrightarrow U(\langle\sharp A\rangle\psi\wedge\neg A \to \langle bett\rangle\langle\sharp A\rangle\varphi)\wedge U(\langle\sharp A\rangle\psi\wedge A \to \langle bett\rangle(\langle\sharp A\rangle\varphi\wedge A))$$
$$\leftrightarrow Pref^{\forall\exists}(\langle\sharp A\rangle\varphi, \langle\sharp A\rangle\psi) \wedge Pref^{\forall\exists}((\langle\sharp A\rangle\varphi \wedge A), (\langle\sharp A\rangle\psi \wedge A)).$$

A final benefit of finding the right recursion principles for betterness change is this. Working inside out, the reduction axioms allow us to reduce dynamic preference formulas to equivalent static preference formulas in our original modal language, and using the latter's completeness, we get the following completeness result:

THEOREM 4.4. *Dynamic preference upgrade logic is complete.*

Once again, we emphasize that 'suggestion' was just our running example here. Our analysis works equally well for other betterness-changing triggers that have been proposed in the literature. One concrete example would be the "radical revision A"($\Uparrow A$) studied in belief revision theory, which makes all A-worlds better than all non-A worlds, while preserving any betterness relations that hold inside these two zones. In terms of preference, one could see this as a very strong action to start loving the A-worlds in the model.

4.2. Priority change leads to preference change

Having secured an understanding of betterness dynamics, we now move to the combination with our earlier way of introducing reasons for preferences. Recall that in the priority-based model, object-level preference was derived from a given priority sequence of propositions or properties, representing some order of importance. Naturally, there are dynamics at this level, too. New properties may become important, others may lose status. Any such change in a priority sequence may cause a change in derived preference. (JL08) consider the following four dynamic operations:

- $[^+C]$: adding C to the right of a priority sequence,
- $[C^+]$: adding C to the left,
- $[-]$: dropping the last element of a priority sequence,
- $[i \leftrightarrow i+1]$: interchanging the i-th and i+1-th elements.

These operations are natural, when considering new 'top priorities' or switches in relative importance. Following the same methodology as before, we obtain a complete set of four recursive reduction axioms:

$[^+C]Pref(x,y) \leftrightarrow Pref(x,y) \vee (Eq(x,y) \wedge C(x) \wedge \neg C(y))$,
$[C^+]Pref(x,y) \leftrightarrow ((C(x) \wedge \neg C(y)) \vee ((C(x) \leftrightarrow C(y)) \wedge Pref(x,y)))$,
$[-]Pref(x,y) \leftrightarrow Pref_{n-1}(x,y)$,
$[i \leftrightarrow i+1]Pref(x,y) \leftrightarrow Pref_{i-1}(x,y) \vee (Eq_{i-1}(x,y) \wedge C_{i+1}(x) \wedge \neg C_{i+1}(y)) \vee (Pref_i(x,y) \wedge (C_{i+1}(x) \leftrightarrow C_{i+1}(y))) \vee (Eq_{i+1}(x,y) \wedge Pref(x,y))$.

Here, either of the first two, plus the third and the fourth are sufficient to represent any change whatsoever in a priority sequence. Note also that operator $[C^+]$ has exactly the same effects on a model as the operator $[\sharp C]$ in the betterness relation-based models. Connections like this will be explored later in Section 5.2.

Our 'revisiting' of von Wright's work has achieved the following. We have shown that we can handle the two main issues he left out in a precise logical manner, viz. reasons for preference and changes in preference. Moreover, the resulting systems allow us to think much more clearly and systematically about how reasoning will proceed in such a structured dynamic environment, including the interplay of preference and new information. We feel this is the proper setting for exploring the detailed behaviour of rational agents who must act on incoming information, while driven by changing preferences and updated goals.

5. Structured models

The proposals in the two preceding sections are based on different intuitions, viz. primitive betterness relations versus priority-derived ones. Both views occur in the literature, and both seem plausible and attractive. Nevertheless, our next task is clear. In order to arrive at a truly coherent logical account of preference, we must further compare the two approaches. Indeed, we shall merge them.

5.1. Relating order at the two levels

From now on, we move to a slightly more abstract level to talk about preference and its reasons, using the following identification. Modal betterness models have an ordering of possible worlds, with propositions as sets of possible worlds, while our priority-based model was about ordering individual objects using their properties. In what follows, we will take all 'objects' to be 'worlds' in modal models - but this is just a vivid manner of speaking, and nothing would be lost if the reader were to think of 'points' and 'properties' instead of worlds and propositions. [8] With this out of the way, we start with two as yet independent orderings at different levels. One is the betterness relation over possible worlds, written as $(W, \preceq)$, the other a preference or priority relation over propositions, viewed as sets of possible worlds, denoted by $(\mathscr{P}, <)$. We bring them together as follows:

DEFINITION 5.1. A *structured model* $\mathcal{M}$ is a tuple $(W, \preceq, V, (\mathscr{P}, <))$ where W is a set of possible worlds, $\preceq$ a preference relation over W, V a valuation function for proposition letters, and $(\mathscr{P}, <)$ an ordered set of propositions, the 'important properties' or priorities. $\lhd$

Structured models simply combine the previous two approaches. In particular, the syntactic priorities are now moved directly into the models, and sit together with an order over worlds. We will see how these two layers are connected, asking the following immediate questions:

- How to *derive* a preference order of 'betterness' over possible worlds from an ordered priority sequence?

- Vice versa, how to *lift* a betterness relation on worlds to an ordering over propositions?

The following diagram shows these two complementary directions:

$$\begin{array}{c} (\mathscr{P}, <) \\ \text{lift} \uparrow \quad \downarrow \text{derive} \\ (W, \preceq) \end{array}$$

From priorities to possible worlds For the direction of deriving, there are many proposals in the literature. Section 3 considered only finite linearly ordered priority sequences, and object order was derived via an *OT*-style lexicographic stipulation. But besides the *OT*-definition, there are other ways of deriving object preferences from a priority base. E.g., if the set of priorities $\mathscr{P}$ is *flat*, the following definition gives a natural order (we call it the '$*$-definition'):

$$y \preceq^*_{\mathscr{P}} x ::= \forall P \in \mathscr{P}(Py \to Px).$$

[8] Of course, it also makes sense to have a 'modal predicate logic' view of worlds containing objects, including different betterness relations at both levels - but for more on that, we refer the reader to (Liu08a).

This is used, e.g. in the theory of default reasoning of Veltman ((Vel96)), or in the topological order theory of Chu Spaces ((Ben00)). More sophisticated options for getting preference from a priority sequence include this following definition was proposed in (ARS02):

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

This has the additional virtue that it works quite generally on *partial orders* of important propositions, not just linear priority sequences. [9] We will use the OT-, $*$-, and AR-definitions to achieve greater generality in our coming discussion of dynamics.

From possible worlds to priorities For the converse direction of lifting, we already said that many quantifier combinations are available taking world orders $x \preceq y$ to proposition or set orders $P \preceq Q$. (BRG07)) lists four obvious ones:

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

One can argue for any of these. We used the $\forall\exists$ combination to define preference from betterness in Section 3. (BOR06) claims that $\forall\forall$ is the notion of preference intended by von Wright in (Wri63), and it provides an axiomatization. But the tradition is much older and (modal) logics for preference relations over sets of possible worlds have been studied by (Lew73), (Bou94), (Hal97), and other authors.

In the preceding sections, we have modeled changes in the betterness relation over possible worlds, and also changes in priority sequences. With our structured models, we can now relate changes at these two levels in a systematic manner.

5.2. Relating dynamics at the two levels

Operations at the two levels We start with a few specific operations on both levels which have been proposed in the literature. At the possible worlds level, we look at the following two operations "suggestion A" ($\sharp A$) and the earlier-mentioned "radical revision A"($\Uparrow A$), which can both be defined in a perspicuous PDL-style language.

- $\sharp A(R) ::= (?A; R; ?A) \cup (?\neg A; R; ?\neg A) \cup (?\neg A; R; ?A)$.
- $\Uparrow A(R) ::= (?A; R; ?A) \cup (?\neg A; R; ?\neg A) \cup (?\neg A; \top; ?A)$.

[9] One more option is the *best-out ordering* of (BCD$^+$93).

Here, $\top$ is the universal relation between all pairs of worlds. The reader can easily see how these definitions spell out the earlier informal definitions of these relation changing operations.

If we can define an operation in such a dynamic format with only atomic test $?A$, $?\neg A$, and with single occurrences of the relation R, we will loosely say that it 'has a PDL-format'.

Next, at the level of priority sequences, or general priority graphs, we mention just two operations that generalize earlier ones:

– $\mathscr{P}; A$ (cf. the 'postfixing' operator $[^+A]$).

– $A; \mathscr{P}$ (cf. the 'prefixing' operator $[A^+]$).

In the special case of flat sets $\mathscr{P}$, both collapse to one set-theoretic union operator $\mathscr{P} + A$. [10]

A few correspondence results To state more uniform connections between transformations at the two levels, we give the following notion about relation-transforming functions.

DEFINITION 5.2. Let F: $(\mathscr{P}, A) \to \mathscr{P}'$, where $\mathscr{P}$ and $\mathscr{P}'$ are ordered sets of propositions, and A is a new proposition. Let σ: $(\preceq, A) \to \preceq'$, where $\preceq$ and $\preceq'$ are relations over possible worlds, and A is a new proposition. We say that *the map F induces the map σ*, given a definition of deriving object preferences from propositions, if, for any set of propositions $\mathscr{P}$ and new proposition A, we have

$$\sigma(\preceq_{\mathscr{P}}, A) = \preceq_{F(\mathscr{P}, A)} .$$

$\lhd$

This notion looks technical, but it really expresses a natural harmony between changes at the two levels of our structured models. Let us first look at how this connection works out for the simplest case, viz. adding a new proposition to a flat priority set:

FACT 5.3. *Given the $(*)$-definition, taking a suggestion A given some relation over possible worlds is induced by the following operation at the propositional level: adding a new proposition A to a flat set $\mathscr{P}$. Stated more precisely, the following diagram commutes:*

[10] (Gir08) and (Liu08a) study further natural operations, such as *disjoint union* of graphs, modeling the case of merging considerations of 'equal importance'.

$$\begin{array}{ccc} \langle W, \mathscr{P} \rangle & \xrightarrow{+A} & \langle W, \mathscr{P} \cup A \rangle \\ {\scriptstyle *}\downarrow & & \downarrow{\scriptstyle *} \\ \langle W, \preceq \rangle & \xrightarrow{\sharp A} & \langle W, \sharp A(\preceq) \rangle \end{array}$$

This means that a change in preference of this sort can be implemented either at the priority level, or at the possible worlds level, provided we use the right ($*$-) notion of 'derivation' between the two levels of the diagram.

Next we consider the case of an *ordered* set $(\mathscr{P}, <)$, where a new proposition A is added in front as a 'top priority'. Now the two-level harmony works out as follows:

FACT 5.4. *Given the OT-definition, upgrade $\Uparrow A$ over possible worlds is induced by the following operation on propositional priority orders: prefixing a new A to an ordered propositional set $(\mathscr{P}, <)$. Stated more precisely, the following diagram commutes:*

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

Thus, adding a new priority in front of a priority sequence is the same as executing "radical revision" at the possible world level.

General discussion The above two results have a similar format, linking operations at the possible world level to operations at the priority level. This leads to general questions about a combined framework:

(i) Given any priority-level preference transformer, can we define a matching world-level relation transformer?

(ii) Given any world-level relation preference transformer, can we define a matching priority-level transformer?

It is not easy to answer this in general, and we merely provide some discussion showing what interesting structures lie beneath the surface here. As a trickier instance of Direction (i), consider the also natural operation $\mathscr{P}; A$ of postfixing a proposition to an ordered set of propositions (the 'last consideration'). Now there is no simple corresponding operation at the possible world level. We need some relational algebra beyond the earlier PDL-format, witness the following observation, which employs the ARS-format extending our earlier OT-definition:

FACT 5.5. $y \preceq^{ARS}_{\mathscr{P};A} x$ *iff* $y \prec^{ARS}_{\mathscr{P}} x \vee (y \preceq^{ARS}_{\mathscr{P}} x \wedge (Ay \to Ax))$.

As for Direction (ii), given some arbitrary relation change at the level of possible worlds, it is by no means the case that there must be a simple corresponding priority operation. Indeed, our innocent 'suggestion' provides a counter-example:

FACT 5.6. *Working with linearly ordered sets of propositions* $(\mathscr{P}, <)$, $\sharp A$ *is not induced by any* F *that preserves linearity.*

Linearity of the betterness order can get lost when we execute the operation $\sharp A$. However, if we allow partially ordered priority sets, with a disjoint union operation over these, we get a positive result:

FACT 5.7. *The operation* $\sharp A$ *is induced by this operation* F *on* $\mathscr{P}$*:*

$$F(A, \mathscr{P}) = (A; \mathscr{P}) \uplus (\mathscr{P}; A).$$

For more discussion of natural operations on priority sets, and their interpretation as a sort of 'agenda algebra' of importance, we refer to (Liu08a), (Gir08) and (Liu08c).

The preceding results show there is an elegant combined perspective for reason-based preference and its dynamics. So, our earlier two amendments to von Wright merge harmoniously in one structured view of agency at interconnected levels. In doing so, the above logical investigation has resulted in a new model for new significant conceptual issues concerning preference, which naturally extends the usual discussion of preference found with analytical philosophers. Finally, the technical issues that emerged in this conceptual analysis establish further logical interest, rather than mere formal nuisance.

6. Preference and beliefs

So far, I have dealt with two crucial aspects of preference enriching the original framework of von Wright. To test my approach, I now go one step further, and discuss one more philosophical issue which is not really considered by von Wright, namely, the uncertainties involved in preference. For instance, when someone tries to give a reason for her preference, in some situations, she may not have precise information to offer. Instead, she may say things like 'I believe that it is going to rain, so I prefer bringing my umbrella'. Under such circumstances, one's preference relies on one's *beliefs*, and beliefs come in as extra reasons for preference. Likewise, it has been argued in (LTW03), and (BRG07) that 'ceteris paribus' aspects of preference (observed by von Wright, but left aside) involve a restriction to the doxastically 'most plausible' or 'normal' worlds encoding our beliefs.

In this sense, preference is partly an epistemic or doxastic notion, conceptually 'entangled' with our knowledge or belief. In particular, then, it makes perfect sense to ask whether preference admits of epistemic or doxastic introspection. But such questions can only be made precise when we extend the logics we developed so far. This entanglement can happen to various 'degrees' in our systems, if we also order objects/worlds by plausibility, and then consider combinations of that with betterness. In the following, we will discuss a few 'degrees of entanglement' offered by the logical perspective.

6.1. Entanglement through composition

First recall the global lifted definition of preference we gave before:

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

Intuitively, $(Ubett)$ says the following:

> for any ψ-world in the model, there exists a world which is at least as good as that world, where φ is true.

Essentially this is a comparison between ψ-worlds and φ-worlds in the whole space of possible worlds for an agent, with no subjective attitude involved yet, as pictured below:

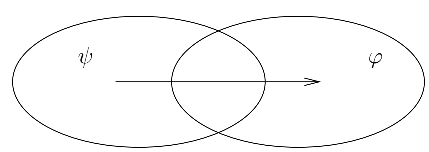

Figure 1. Preference defined by U and betterness relations

But epistemic agents may be only interested in that part of the model which is relevant to them. Thus, we might want to express that:

> for any ψ-world that is most plausible to agent i in the model, there exists a world which is as good as that world, where φ is true.

Here is a definition for this preference using belief modalities:

$$Pref^{\forall\exists}(\varphi,\psi) ::= B_i(\psi \rightarrow \langle bett \rangle \varphi). \quad (Bbett)^{11}$$

To interpret such beliefs formally, we turn to doxastic logic, and extend our modal betterness models with plausibility relations to obtain tuples $\mathcal{M} = (W, \leq, \preceq, V)$, where W is a set of possible worlds, $\leq$ is a doxastic relation 'at least as plausible as', and $\preceq$ the earlier relation 'at least as good as', with V again a valuation for the proposition letters. The new definition is illustrated in the following figure:

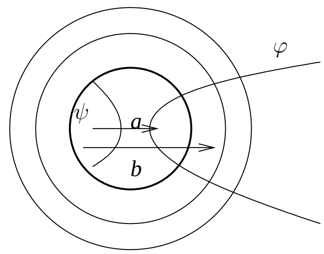

Figure 2. Preference defined by B and betterness relations

In this picture, worlds lie ordered according to their plausibility, as in Lewis' spheres for conditional logic. The part inside the black circle depicts the most plausible worlds. We consider the ψ-worlds in this area, and then distinguish two sorts of preference relation: links of 'type a' stay inside the most plausible region, links of 'type b' go outwards to the less plausible, or even implausible region of the model.

The resulting logical language can even speak about subtle but realistic operator combinations like 'believing that one prefers' or conversely, 'preferring to believe'. (Liu08a) sketches a complete doxastic dynamic preference logic for reasoning in this setting. Its static base system consists of the usual modal axioms for (conditional) beliefs and for betterness. Dynamic reduction axioms take care of both betterness change as described earlier in this paper, epistemic update by world elimination, and belief revision through plausibility change on the models. In other words, combining different dynamic logics for different aspects of agency poses no problems here.

[11] An alternative version would replacing belief by knowledge:

$$Pref^{\forall\exists}(\varphi,\psi) ::= K_i(\psi \rightarrow \langle bett \rangle \varphi). \quad (Kbett)$$

Also, there are natural related notions using *conditional beliefs*: cf. (Liu08a).

6.2. Entanglement through intersection

But things can be more complicated than this. Note that Definitions (*Bbett*) has a specific feature: a better world can lie *outside of the most plausible part* of the model, witness the type b link in Figure 2. The intuition behind this phenomenon is clear, and reasonable in many cases. It may well be that there exist better worlds, which the agent does not view as most plausible. (BL07) consider scenarios where this happens, with 'regret' about what might have been. But *if* we want to have the two base relations entangled more intimately, and stick to more epistemically realistic action-oriented preferences, we might want to just look at better alternatives *inside* the relevant epistemic or doxastic zone. Such considerations are found in the study of normative reasoning in (LTW03) where a normality relation and a preference relation live in one model. Likewise, (BRG07) discuss the 'normality sense' of ceteris paribus preference, restricting preference relations to just the normal worlds for the agents. More discussion of this second more intimate form of preferential-doxastic entanglement can be found in (Ben08). We explore this line a bit further.

Instead of simple composition, we now need to merge the two relations, viz. their intersection. Here is how:

DEFINITION 6.1. A *merged preference model* is a tuple $\mathcal{M} = (W, \leq, \preceq, \leq\cap\preceq, V)$, with W a set of possible worlds with doxastic and betterness relations, but also a new relation $\leq\cap\preceq$ as the intersection of the relations 'at least as plausible as' and 'at least as good as', with V again a valuation for proposition letters. $\triangleleft$

The original language had separate modal operators B and $[bett]$, but now, the new relation allows us to extend it with a new modality H, which we read as 'hopefully'. The semantic truth clause for such formulas is simply as follows:

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

Now we get one more natural notion of preference over propositions:

$$Pref^{\forall\exists}(\varphi, \psi) ::= B(\psi \to \langle H \rangle \varphi). \quad (BH)$$

This entangled notion of belief (also found in (JL08)) says that:

> For any most plausible ψ-world in the model, there exists a world which is *as good as* this world, and at the same time, *as plausible as* this world, where φ is true.

Obviously, we can now talk about preferences that are totally restricted to the most plausible part of the model, keeping only the a-type betterness arrows in Figure 2. As far as a complete calculus of reasoning is concerned, the following result holds:

THEOREM 6.2. *The logic for a combined doxastic preference language with an added 'Hopefully' operator is completely axiomatizable.*

A complete proof can be found in (Liu08b).

This poses a new challenge to our dynamic approach: what happens to the new operator H under our three characteristic actions: standard public announcements $A!$ that updates the domain of worlds, suggestions $\sharp A$ that change the betterness relations, and also radical revisions $\Uparrow A$ changing the plausibility relations? Fortunately, there are still valid reduction axioms, even for the intersective H-operator (see (Liu08a)), which we just display here for concreteness:

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. $\langle A!\rangle\langle H\rangle\varphi \leftrightarrow A \wedge \langle H\rangle\langle A!\rangle\varphi$.

Once we have found these reduction axioms, it is easier to get a complete dynamic logic in which the above operators are involved.

THEOREM 6.3. *The dynamic logic for the intersection of preference and belief is completely axiomatizable.*

Again, I refer to (Liu08b) for precise statements and proofs.

6.3. Other forms of entanglement

There are also other entanglements of preference and beliefs. For instance, decision theory has a tradition in modeling decision making under uncertainty ((Sav54), (Jef65)). Here most models rely on a numerical representation where utility and uncertainty are commensurate. For instance, an agent may not know the outcomes of her actions, but now using a probability distribution over outcomes. The *expected value* of an action can then be computed from utility and probability, as explained in any textbook, which deeply entangles the agents' betterness relations and her beliefs about possible outcomes. By contrast, our logical approach is qualitative, and we must leave a comparison with probabilistic systems for another occasion.

Even so, it will be clear that the style of dynamic preference analysis advocated in this paper does extend naturally to deal with the epistemic and doxastic entanglements of preference, forming a natural complement to von Wright's original setting.

7. Conclusion and further perspectives

Starting from a basic preference logic in modal style close to von Wright's original system for preferences related to obligations and decisions, I have shown how two further essential aspects can naturally be incorporated, viz. reasons for preference, and changes in preference. This resulted in a much more finely-structured priority-based view of reasons for preferences, and the processes that change them, based on a new type of structured two-layer models providing a more structured view of agents as partaking in various kinds of dynamics at the same time. Moreover, this view of agency also ties up naturally with a doxastic setting, where preferences are tightly linked with agents' beliefs. Thus the paper may be seen also as a contribution to epistemology, and the philosophy of action.

Once at this level, one can also see connections with other areas of philosophy, such as ethics or social philosophy. For instance, I feel that the logics proposed here also make sense in the *deontic* study of changing obligations and norms (cf. (Yam06) and (Yam07)). Moreover, they admit easily of a *multi-agent* perspective: which is what they were designed for anyway in the epistemic case. Indeed, preference change is usually a multi-agent process, where both factual information and information about other people's preferences change our own. Thus, with the current apparatus, we are now at a stage where we could do a more detailed dynamic logic of deliberation, which ties up with ideas of 'procedural justice'. Finally, a natural extension of all the above systems would also assign preferences to *group actors*, merging common knowledge and belief with group preferences as studied in social choice theory (cf. (ARS02) and (Ben08)).

Going back to the interface with analytic philosophy, I see the combination of issues concerning rational agency in this paper as one more virtue of our formal analysis. Logical methods often cut across philosophical boundaries, making ideas flow in surprising channels. In particular, our new reason-based dynamic logics provide a richer view of what preference really is, including its entanglements with belief. In doing so, they extend existing philosophical discussions, and also, provide the philosopher with tools for carrying this perspective further.

Acknowledgement I thank Tongdong Bai for his efforts in putting together this volume. I thank Johan van Benthem for his helpful comments. Finally, I thank Esther Rosario for her English corrections.

References

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

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

S. Benferhat, C. Cayol, D. Dubois, J. Lang, and H. Prade. Inconsistency management and prioritized syntax-based entailment. In *Proceedings of IJCAI'93*, pages 640–645, 1993.

J. van Benthem, J. van Eijck, and A. Frolova. Changing preferences. Technical Report, CS-93-10, Centre for Mathematics and Computer Science, Amsterdam, 1993.

J. van Benthem. *Exploring Logical Dynamics*. CSLI Publication, Stanford, 1996.

J. van Benthem. Information transfer across Chu spaces. *Logic Journal of the IGPL*, 8:719–731, 2000.

J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logic*, 17:129–156, 2007. Technical Report, PP-2006-11, ILLC, University of Amsterdam.

J. van Benthem. For better or for worse: Dynamic logic of preference. To appear in T.Grune-Yanoff and S.O. Hansson eds, *Preference Change: Approaches from Philosophy, Economics and Psychology*, Theory and Decision Library, 2008.

J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logic*, 17:157–182, 2007. Technical Report, PP-2005-29, ILLC, University of Amsterdam.

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.

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.

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

J. van Benthem, O. Roy, and P. Girard. Everything else being equal: A modal logic approach to ceteris paribus preferences. Technical Report, PP-2007-09, ILLC, University of Amsterdam, 2007.

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

A. Baltag and S. Smets. Dynamic belief revision over multi-agent plausibility models. In *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 06)*, Liverpool, 2006.

A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. To appear in G. Bonanno, W. van der Hoek, M. Wooldridge, editors, *Texts in Logic and Games*, 2008.

S. Coste-Marquis, J. Lang, P. Liberatore, and P. Marquis. Expressive power and succinctness of propositional languages for preference representation. In *Proceedings of KR 2004*. AAAI Press, 2004.

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

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

P. Girard. *Modal logics for belief and preference change*. PhD thesis, Stanford University, 2008. To appear.

S. Halldén. *On the Logic of "Better"*. Lund, 1957.

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

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

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

S.O. Hansson and T. Grüne-Yanoff. Preferences. In *Stanford Encyclopedia of Philosophy*. Stanford, 2006. http://plato.stanford.edu/entries/preferences/.

R.C. Jeffrey. *The Logic of Decision*. Chicago: University of Chicago Press, 1965.

D. de Jongh and F. Liu. Preference, priorities and belief. To appear in T. Grune-Yanoff and S. O. Hansson eds., Preference Change: Approaches from Philosophy, Economics and Psychology, 2008.

D. Lewis. *Counterfactuals*. Oxford: Blackwell, 1973.

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

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

F. Liu. Logics for interaction between preference and belief. Manuscript, Department of Philosophy, Tsinghua University, 2008.

F. Liu. A two-level perspective on preference. Manuscript, Department of Philosophy, Tsinghua University, 2008.

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)*, 2003.

J.A. Plaza. Logics of public announcements. In *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, 1989.

L.J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

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

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

G.H. von Wright. *The Logic of Preference*. Edinburgh, 1963.

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.

T. Yamada. Logical dynamics of some speech acts that affect obligations and preferences. In J. van Benthem, S. Ju, and F. Veltman, editors, *A Meeting of the Minds–Proceedings of the Workshop on Logic, Rationality and Interaction*, pages 275–289. King's College Publications, 2007.