The Original Position: A Logical Analysis

Thijs De Coninck and Frederik Van De Putte [1]

Ghent University
University of Bayreuth

Abstract

Rawls famously claimed that choices based on the Difference Principle coincide with the choices of any rational individual in the Original Position. In this paper, we develop a logic in which we can express and prove Rawls' thesis in its object language. Starting from a standard semantics of choice under uncertainty, we enrich our models in order to represent uncertainty about one's position. We then introduce a sound and strongly complete logic that allows us to speak about agents' positions and their derived utilities, and that can express changes in the uncertainty about those positions using dynamic operators. Finally, we show how this logic allows us to define various types of obligation based on a Rawlsian notion of procedural fairness.

Keywords: The Original Position, Choice under uncertainty, Deontic logic, Fairness.

1 Introduction

In his *Theory of Justice*, John Rawls puts forward principles of justice that he argues should be used to determine the basic structure of society [16]. What made Rawls innovative, however, were not the principles themselves, but the way in which he argued for them [15]. Famously, he makes use of a methodological device known as the *Original Position*, which he describes as

> [...] a purely hypothetical situation characterized so as to lead to a certain conception of justice. Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength, and the like. [16, p. 11]

So Rawls' basic idea is to conceive of a situation in which a person is deprived of morally irrelevant knowledge and to ask: what would such a person choose?

One of the principles that would characterize the resulting choices, according to Rawls, is the *Difference Principle*, which states that "social and economic inequalities are to be arranged so that they are to the greatest benefit of the

[1] Thijs De Coninck is a PhD fellow of the Research Foundation – Flanders supported by a fundamental research grant (1167619N). Frederik Van De Putte is a Marie Skłodowska-Curie Fellow at the University of Bayreuth (grant agreement ID: 795329) and a Postdoctoral Fellow of the Research Foundation - Flanders (grant 12Q1918N).

least advantaged." [16, p. 266]. Rawls claims that what one ought to choose according to the Difference Principle coincides with what a rational individual would choose, if it were fully uncertain about the position it occupies in society. Henceforth we refer to this claim as *Rawls' thesis.* [2]

In this paper, we provide a logical analysis of Rawls' thesis. We first set out general models of choice under uncertainty and define notions of individually rational choices and fair choices (Section 2). Next, we refine these models in such a way that we can verify Rawls' thesis (Section 3). In Section 4 we introduce a logic that can express key features of those models and has Rawls' thesis as a validity. Finally, we show how our logic can handle various deontic operators based on Rawls' notion of procedural fairness (Section 5).

Existing Formal Models Rawls' publication of *A Theory of Justice* has spawned research both of informal and formal nature. Most of the formal literature is focused on the Rawls/Harsanyi dispute over how exactly to characterize the Original Position and how agents would choose, once placed in such a situation. John Harsanyi [9] conceives of the situation as one of choice under risk where, for any given outcome, the agent can reason based on some probability estimate of how likely it is that that outcome occurs. With this in place, Harsanyi argues that a rational individual would choose according to the principles of expected utility theory. In contrast, Rawls thinks of the situation as one of choice under uncertainty, where no such probabilities are given [16, p. 134].

Given a great deal of uncertainty and the risks associated with choosing suboptimal options, Rawls argues that individuals would choose according to the *Maximin rule* (cf. Section 2.2). In contrast, most of the formal work on the Original Position follows Harsanyi's characterization by relying on a uniform probability distribution that assigns a chance of $\frac{1}{n}$ to an individual ending up in one of the n possible positions (see e.g. [6,8,13,17]).

In the present paper, we bracket the Rawls/Harsanyi dispute and stay as close as possible to Rawls' conception of the Original Position as a situation of non-strategic choice under uncertainty (cf. [7]).

2 Choice Under Uncertainty

In this section we present general models of choice under uncertainty and introduce a formal language that can express some of Rawls's fundamental concepts.

[2] Rawls states: "To say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion. If necessary, the argument to this result could be set out more formally." [16, p. 120]. What we call Rawls' thesis is thus the more concrete version of this claim where the Difference Principle is put forward as the conception of justice in question.

2.1 Models of Choice Under Uncertainty

Our models are inspired by the tradition of STIT logics, i.e. logics that feature modal operators of the type "agent i sees to it that", which are interpreted in terms of the states of affairs that are guaranteed by the (past or current) choice(s) of i. The classic exposition of STIT logic is Belnap et al. [4]. In [10], Horty shows that this framework can be combined with utilitarian ideas, in order to interpret various deontic notions such as individual and group oughts. Kooi and Tamminga [14,18] use STIT models without a temporal component, but including agent-relative utilities. Here, we further simplify the models of [14] by working with a single set of choices and a finite [3] set of utility values.

Fix a finite set Agt of agents, a finite set $N = \{1, 2, \ldots\} \subset \mathbb{N}$ of utilities, and a countable set $Q = \{q, q', \ldots\}$ of propositional variables. We use i, j and n, m as metavariables for agents and values respectively.

Definition 2.1 A *model of choice under uncertainty* is a tuple $\mathfrak{M} = \langle S, U, C, V\rangle$, where $S \neq \emptyset$ is a set of *states*, $U : S \times Agt \to N$ is a *utility function*, C is a partition of S into *choices*, and $V : Q \to \wp(S)$ is a *valuation function*.

Each state $s \in S$ can be seen as a possible outcome of the choice situation. The utility function U specifies, for each state s and agent i, the utility $U(s, i)$ that i receives at s. Note that choices are *sets* of states $X \in C$. This means that, as in the traditional STIT-based accounts, we identify choices with the set of states they leave open. Unlike in STIT, we do not attribute choices to (a) particular (group of) agents. The focus is rather on how choices *affect* agents, not on who is choosing or acting. Depending on the particular perspective we take, e.g. that of an individual or that of society at large, some of the choices will be better or worse than others. Correspondingly, one may interpret the choices as those of a social planner or policy-maker, even if that person is herself a member of Agt.

If $s \in X$, then s is a possible outcome of choosing X. We write $C(s)$ for the unique choice $X \in C$ such that $s \in X$. Figure 1 represents a simple model of choice under uncertainty for two agents i and j, with two choices $X = \{s_1, s_2\}$ and $Y = \{s_3, s_4\}$. Here, the couples (n, m) represent the utility function, where $n = U(s, i)$ and $m = U(s, j)$. For instance, at state s_1, agent i receives a utility of 3 whereas agent j receives a utility of 1.

X		Y	
s_1	s_2	s_3	s_4
$(3, 1)$	$(2, 4)$	$(2, 2)$	$(1, 4)$

Fig. 1. A model of choice under uncertainty.

[3] The generalization to an infinite set of utility values is left for future work.

2.2 Two Standards of Admissibility

Given a model of choice under uncertainty $\mathfrak{M}$, the utility function U induces agent-relative preferences over outcomes: i weakly prefers s over s' iff $U(s,i) \geq U(s',i)$. For example, in Figure 1 agent i weakly prefers s_1 over s_2 since $U(s_1,i) \geq U(s_2,i)$. However, since we are considering what choices a rational agent should make, we should specify how preferences over states induce preferences over choices. In other words, we need to specify a *lifting criterion*. Four such lifting criteria are given in Table 1, which is based on [19]. Each of these lifting criteria give us a weak preference relation over the set of choices in a model.

Where $l \in \{\forall\forall, \forall\exists, \exists\forall, \exists\exists\}$ the strict preference relation $\sqsupset_i^l$ is defined as: $X \sqsupset_i^l Y$ iff $X \sqsupseteq_i^l Y$ and $Y \not\sqsupseteq_i^l X$. In words, X being strictly preferred to Y means that X is preferred to Y while Y is not preferred to X. Following common practice, we assume that it is rational to choose X for an agent i iff there is no other choice Y such that i strictly prefers Y to X. We call such rational choices *admissible* for the agent in question, and treat rationality and admissibility as interchangeable notions.

Definition 2.2 Where $\mathfrak{M} = \langle S, U, C, V\rangle$ is a model of choice under uncertainty, $i \in Agt$, and $l \in \{\forall\forall, \forall\exists, \exists\forall, \exists\exists\}$: the set of *$i$-admissible^l choices in* $\mathfrak{M}$ is

$$Adm_i^l(\mathfrak{M}) =_{\mathsf{df}} \{X \in C \mid \text{for no } Y \in C : Y \sqsupset_i^l X\}.$$

$l=$	**Preference Relation**
$\forall\forall$	$X \sqsupseteq_i^{\forall\forall} Y \;\; =_{\mathsf{df}} \;\; \forall s \in X, \forall s' \in Y : U(s,i) \geq U(s',i)$
$\forall\exists$	$X \sqsupseteq_i^{\forall\exists} Y \;\; =_{\mathsf{df}} \;\; \forall s \in X, \exists s' \in Y : U(s,i) \geq U(s',i)$
$\exists\forall$	$X \sqsupseteq_i^{\exists\forall} Y \;\; =_{\mathsf{df}} \;\; \exists s \in X, \forall s' \in Y : U(s,i) \geq U(s',i)$
$\exists\exists$	$X \sqsupseteq_i^{\exists\exists} Y \;\; =_{\mathsf{df}} \;\; \exists s \in X, \exists s' \in Y : U(s,i) \geq U(s',i)$

Table 1
Lifting criteria. Here, X and Y are sets of states.

Maximin In what follows, we focus on the *Maximin* criterion, i.e. the lifting criterion denoted by $\forall\exists$. We hence take $Adm_i^{\forall\exists}$ as defining rational choice under uncertainty. We return to the other lifting criteria in Section 5. Until then, we omit the superscript l in notation.

The Maximin principle is usually considered typical for risk-averse agents. Rawls states that "the maximin rule is not, in general, a suitable guide for choices under uncertainty" while he does defend Maximin in situations "marked by certain special features" [16, pp. 133]. These features are:

- knowledge of likelihoods is impossible, or at best extremely insecure;
- the person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that

he can, in fact, be sure of by following the maximin rule;

- the rejected alternatives have outcomes that one can hardly accept.

Rawls concludes that because the Original Position has these three features, the Maximin criterion is the most appropriate one in this context.

The Difference Principle The principle of justice that we focus on in this paper is the Difference Principle. Informally, it states that we should maximize the gains of the least well-off. Rawls warns us that the Difference Principle should not be mistaken for the Maximin rule [16, p. 72]:

> The maximin criterion is generally understood as a rule for choice under great uncertainty, whereas the difference principle is a principle of justice. It is undesirable to use the same name for two things that are so distinct.

To define the Difference Principle in exact terms, we need some more notation. For any state s in a given model, let $U(s,*)$ denote the smallest $n \in N$ such that, for some $i \in Agt$, $U(s,i) = n$. Intuitively, $U(s,*)$ is the utility of the agent that is the least well-off at state s. One may say that, according to the Difference Principle, a state s is at least as good as a state s' if and only if $U(s,*) \geq U(s',*)$, i.e. whenever the least well-off at state s is at least as well-off as the least well-off at state s'.

Just as before, we need to lift this preference relation on states in order to obtain preferences over choices. In line with the preceding, we use the Maximin criterion. [4] This gives us the following definitions:

Definition 2.3 Where $\mathfrak{M} = \langle S, U, C, V\rangle$ is a model of choice under uncertainty and $X, Y \in C$: $X \sqsupseteq_*^{\forall\exists} Y$ iff $\forall s \in X, \exists s' \in Y$: $U(s,*) \geq U(s',*)$.

Definition 2.4 Where $\mathfrak{M} = \langle S, U, C, V\rangle$ is a model of choice under uncertainty, the set of *Difference admissible choices in* $\mathfrak{M}$ is

$$Adm_*(\mathfrak{M}) =_{\mathsf{df}} \{X \in C \mid \text{For no } Y \in C : Y \sqsupset_*^{\forall\exists} X\}.$$

In our example from Figure 1, it can be easily verified that $X \sqsupset_i^{\forall\exists} Y$, $Y \sqsupset_j^{\forall\exists} X$, $X \sqsupseteq_*^{\forall\exists} Y$, and $Y \sqsupseteq_*^{\forall\exists} X$. Hence, $Adm_i(\mathfrak{M}) = \{X\}, Adm_j = \{Y\}$, and $Adm_* = \{X, Y\}$. In other words, both X and Y are difference admissible in this model, while only X is admissible for i and only Y is admissible for j.

In what follows, we will sometimes use "$*$" to denote "the least well-off" (at a given state in a given model). This convention allows us to present our results in a compact way.

2.3 Expressing Admissibility in a Formal Language

Here, we introduce a formal language that allows us to express i.a. that the current choice is i-admissible and/or difference admissible. Let $\mathfrak{L}$ be defined by the following Backus-Naur Form (BNF):

$$\varphi := q \mid \mathsf{u}_i^n \mid \neg\varphi \mid \varphi \vee \varphi \mid \Box\varphi \mid \boxed{\mathsf{c}}\varphi$$

[4] One can define alternative "fairness rankings", using the other lifting criteria from Table 1. We leave the study of such rankings for future work.

where q ranges over Q, i over Agt, and n over N. The constant u_i^n expresses that agent i receives a utility of n. $\Box$ is a universal modality: $\Box\varphi$ means that φ is true at all states in the model; $\Diamond$ denotes its dual. $\boxed{c}\varphi$ expresses that the current choice guarantees that φ is the case; the dual of $\boxed{c}$ is denoted by $\langle c \rangle$. $\boxed{c}$ is a normal modal operator, similar in spirit to the "Chellas STIT" [5,11]. Both $\Box$ and $\boxed{c}$ are modal operators of type S5.

Definition 2.5 Where $\mathfrak{M} = \langle S, C, U, V\rangle$ is a model of choice under uncertainty and $s \in S$:

(SC1) $\mathfrak{M}, s \models q$ iff $s \in V(q)$

(SC2) $\mathfrak{M}, s \models \neg\varphi$ iff $\mathfrak{M}, s \not\models \varphi$

(SC3) $\mathfrak{M}, s \models \varphi \vee \psi$ iff $\mathfrak{M}, s \models \varphi$ or $\mathfrak{M}, s \models \psi$

(SC4) $\mathfrak{M}, s \models \boxed{c}\varphi$ iff for all $s' \in C(s)$, $\mathfrak{M}, s' \models \varphi$

(SC5) $\mathfrak{M}, s \models \Box\varphi$ iff for all $s' \in S$, $\mathfrak{M}, s' \models \varphi$

(SC6) $\mathfrak{M}, s \models \mathsf{u}_i^n$ iff $U(s, i) = n$.

Let us use the example from Figure 1 to illustrate some of these semantic clauses. Let $\mathfrak{M}$ correspond to the model in Figure 1 with $V(q) = \{s_1, s_3, s_4\}$. Since $U(s_1, i) = 3$ and by applying (SC6), we obtain that $\mathfrak{M}, s_1 \models \mathsf{u}_i^3$. In view of (SC4) and since $U(s_2, i) = 2$, $\mathfrak{M}, s_1 \not\models \boxed{c}\mathsf{u}_i^3$. Likewise, since q is false at s_2, $\mathfrak{M}, s_1 \not\models \boxed{c}q$. However, by (SC6) and since q is true at both s_3 and s_4, we have $\mathfrak{M}, s_1 \models \Diamond\boxed{c}q$.

With the language $\mathfrak{L}$, we can express the notions of individual admissibility and difference admissibility that were introduced in Section 2.2. In order to explain this, we need some preparatory work. Where $\dagger \in Agt \cup \{*\}$ and where s is a state in some model $\mathfrak{M}$, let $G^{\mathfrak{M}}(s, \dagger)$ be the set of all $n \in N$ such that for all $s' \in C(s)$, $U(s', \dagger) \geq n$. When $n \in G^{\mathfrak{M}}(s, \dagger)$, we say that utility n is *guaranteed* for $\dagger$ at s. A little reflection on the Maximin criterion and our definitions of admissibility gives us:

Lemma 2.6 *$C(s) \in Adm_\dagger(\mathfrak{M})$ iff for all $s' \in S$: $G^{\mathfrak{M}}(s', \dagger) \subseteq G^{\mathfrak{M}}(s, \dagger)$.*

Let s be a state in some model of choice under uncertainty, and let $X = C(s)$. Using the formal language $\mathfrak{L}$, we can express that, for any utility $n \in N$ and for any other choice Y in the model, if Y guarantees n, then so does X – see Table 2. Relying on Lemma 2.6, we immediately obtain:

Theorem 2.7 *Where $\mathfrak{M} = \langle S, U, C, V\rangle$ is a model of choice under uncertainty, $s \in S$, and $\dagger \in Agt \cup \{*\}$: $C(s) \in Adm_\dagger(\mathfrak{M})$ iff $\mathfrak{M}, s \models \mathsf{Adm}_\dagger$.*

Proof. $C(s) \in Adm_\dagger(\mathfrak{M})$ iff [by Lemma 2.6] for all $s' \in S$, $G^{\mathfrak{M}}(s', \dagger) \subseteq G^{\mathfrak{M}}(s, \dagger)$ iff for all $s' \in S$, for all $n \in N$, if $n \in G^{\mathfrak{M}}(s', \dagger)$ then $n \in G^{\mathfrak{M}}(s, \dagger)$ iff [in view of Table 2] for all $n \in N$, for all $s' \in S$, if $\mathfrak{M}, s' \models \mathsf{g}_\dagger^n$, then $\mathfrak{M}, s \models \mathsf{g}_\dagger^n$ iff for all $n \in N$, if there is an $s' \in S$ such that $\mathfrak{M}, s' \models \mathsf{g}_\dagger^n$, then $\mathfrak{M}, s \models \mathsf{g}_\dagger^n$ iff [by the semantic clauses] for all $n \in N$, $\mathfrak{M}, s \models \Diamond\mathsf{g}_\dagger^n \rightarrow \mathsf{g}_\dagger^n$ iff $\mathfrak{M}, s \models \mathsf{Adm}_\dagger$. $\Box$

Theorem 2.7 tells us that we can express that a given option is individually admissible (according to the Maximin criterion) or difference admissible in $\mathfrak{L}$.

Abbr.	Definition	Interpretation
$\mathsf{u}_i^{\geq_n}$	$\bigvee_{m \geq n} \mathsf{u}_i^m$	The utility of i is at least n.
u_*^n	$(\bigvee_{i \in Agt} \mathsf{u}_i^n) \wedge (\bigwedge_{j \in Agt} \mathsf{u}_j^{\geq n})$	The utility of the least well-off is n.
$\mathsf{u}_*^{\geq_n}$	$\bigvee_{m \geq n} \mathsf{u}_*^m$	The utility of the least well-off is at least n.
g_i^n	$\boxed{\mathsf{c}}\mathsf{u}_i^{\geq_n}$	A utility of n is guaranteed for i.
g_*^n	$\boxed{\mathsf{c}}\mathsf{u}_*^{\geq_n}$	A utility of n is guaranteed for the least well-off.
Adm_i	$\bigwedge_{n \in N}(\Diamond \mathsf{g}_i^n \to \mathsf{g}_i^n)$	The given choice is i-admissible.
Adm_*	$\bigwedge_{n \in N}(\Diamond \mathsf{g}_*^n \to \mathsf{g}_*^n)$	The given choice is difference admissible.

Table 2
Some useful abbreviations. Here, † ranges over $Agt \cup \{*\}$.

Recall however that, according to Rawls' thesis, these two notions are only related given a specific type of uncertainty, viz. uncertainty about the position one occupies in society. In what follows, we show how our semantics and formal language can be refined in order to represent such uncertainty.

3 A Semantics for Rawls' Thesis

The kind of uncertainty we are dealing with in the Original Position is, at bottom, uncertainty about who gets which position; from that, one then derives uncertainty about the agent's utilities. To make this idea precise, we introduce a more specific class of models of choice under uncertainty in Section 3.1. Next, we define a type of updates on those models, which capture changes in position uncertainty (Section 3.2). Finally, we show how, with the formal instrumentarium thus introduced, we can make Rawls' thesis exact (Section 3.3).

3.1 Models of Choice Under Position Uncertainty

Fix a finite, non-empty set of positions $P = \{p, p', \ldots\}$, with $|P| \leq |Agt|$.[5] Here, one should think of a position in rather abstract terms: a position is simply that which determines the utility of the agent at a given state.

Definition 3.1 A *model of choice under position uncertainty* is a tuple $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0 \rangle$ where $W \neq \emptyset$ is the set of *worlds*, Π is a non-empty set of *position assignments* $\pi : Agt \to P$ that are surjective, C^0 is a partition of W, $U^0 : W \times P \to N$ is a *position-utility function*, and $V^0 : Q \to \wp(W)$ is a *valuation function*.

[5] We require that the number of positions does not exceed that of agents because we will need the presupposition that every position is occupied by at least one agent for Rawls' thesis to hold – see also footnote 7.

Definition 3.2 Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0 \rangle$ is a model of choice under position uncertainty, the corresponding model of choice under uncertainty is $\mathfrak{M} = \langle S, C, U, V \rangle$, where:

- $S =_{\mathsf{df}} W \times \Pi$
- for all $(w, \pi) \in W \times \Pi : U((w, \pi), i) =_{\mathsf{df}} U^0(w, \pi(i))$
- $C =_{\mathsf{df}} \{\{(w, \pi) \mid w \in X, \pi \in \Pi\} \mid X \in C^0\}$
- $V(q) =_{\mathsf{df}} \{(w, \pi) \mid w \in V^0(q), \pi \in \Pi\}$

In a model of choice under position uncertainty, states are made up of two components: a world w that determines what factual states of affairs obtain and what utilities each position gets, and a position assignment π that determines the position of each agent.[6] Note that we require the position assignment functions to be surjective. This means that every position in society is occupied by at least one agent.[7]

This in turn allows us to decompose the utility function U from Section 2 into two parts. First, U^0 specifies the utilities of every position, for every way the world may end up being. So $U^0(w, p) = n$ means that at world w, any agent with position p receives a utility of n. Second, the position assignment π specifies the position an agent gets in society. The *agent-utility of i at a state $s = (w, \pi)$* is then defined as $U^0(w, \pi(i))$: it is the position-utility at w of the position to which i is assigned at s.

In view of Definition 3.2, each model of choice under position uncertainty corresponds to a model of choice under uncertainty. Given this, we can apply our earlier definitions of individual and difference admissibility to models of choice under position uncertainty.

Figure 2 represents two models of choice under position uncertainty. In $\mathfrak{M}^0_1$, Π is a singleton $\{\pi_1\}$. In $\mathfrak{M}^0_2$, Π consists of two position assignments. Note that this difference affects which choices are admissible for each of the agents, though it does not affect which choices are difference admissible. In particular, $Adm_i(\mathfrak{M}^0_1) = \{X\}, Adm_j(\mathfrak{M}^0_1) = \{Y\}$, and $Adm_*(\mathfrak{M}^0_1) = \{X, Y\}$, while $Adm_i(\mathfrak{M}^0_2) = Adm_j(\mathfrak{M}^0_2) = Adm_*(\mathfrak{M}^0_2) = \{X, Y\}$.

3.2 Updates of Position Uncertainty

Given a model $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0 \rangle$, the parameter Π specifies our uncertainty about who gets what position in society. Importantly, and in line with our agent-independent notion of choice, this uncertainty is agent-independent. For instance, if there are $\pi, \pi' \in \Pi$ and $p, p' \in P$ such that $\pi(i) = p$ and

---

[6] Here, a warning is in place: since π determines which agent gets which utility, the "factual states of affairs" are limited to those statements that do not depend, logically speaking, on who gets what. For instance, "agent 2 gets a utility of 5" is not a "factual state of affairs" on this reading. In principle, we could also make the truth of propositional variables dependent on both w and π. This would not affect our main results in this paper.

[7] This presupposition is necessary for Rawls' thesis. Indeed, otherwise the "worst-off" agent given the current position assignment may be guaranteed to get a higher utility than what some agents could have in positions that are currently not occupied.

	X		Y	
	w_1	w_2	w_3	w_4
U^0	$(\mathbf{3},\mathbf{1})$	$(\mathbf{2},\mathbf{4})$	$(\mathbf{2},\mathbf{2})$	$(\mathbf{1},\mathbf{4})$
π_1	$(3,1)$	$(2,4)$	$(2,2)$	$(1,4)$

(a) The model $\mathfrak{M}^0_1$ with $\Pi = \{\pi_1\}$.

	X		Y	
	w_1	w_2	w_3	w_4
U^0	$(\mathbf{3},\mathbf{1})$	$(\mathbf{2},\mathbf{4})$	$(\mathbf{2},\mathbf{2})$	$(\mathbf{1},\mathbf{4})$
π_1	$(3,1)$	$(2,4)$	$(2,2)$	$(1,4)$
π_2	$(1,3)$	$(4,2)$	$(2,2)$	$(4,1)$

(b) The model $\mathfrak{M}^0_2$ with $\Pi = \{\pi_1, \pi_2\}$.

Fig. 2. Two models of choice under position uncertainty. Where $(\mathbf{n}, \mathbf{m}) \in N \times N$, $\mathbf{n}$ denotes the utility of p_1, and $\mathbf{m}$ denotes the utility of p_2 at the given world. The two position assignments are: $\pi_1(i) = p_1, \pi_1(j) = p_2$ and $\pi_2(i) = p_2, \pi_2(j) = p_1$.

$\pi'(i) = p'$ (with $p \neq p'$), then this means that whoever is choosing does not know whether i occupies position p, or rather position p'.

Consequently, a change in position uncertainty amounts to an update of the parameter Π. We will define such updates in general, after which we apply them to Rawls' thesis. In what follows, let Π_* denote the set of *all* position assignments, i.e. all surjective functions $\pi : Agt \to P$.

Definition 3.3 Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ is a model of choice under position uncertainty and where $\emptyset \neq \Pi' \subseteq \Pi_*$, $\mathfrak{M}^0_{\Pi'} = \langle W, \Pi', C^0, U^0, V^0\rangle$.

In other words, all that is changed by an update (if anything) is the set of position assignments that are considered possible. With this general type of update, we can model both increasing and decreasing uncertainty about position assignments. At one end of the spectrum, updates with a singleton $\{\pi\}$ amount to restricting the model to a single position assignment. At the other end, updates with Π_* amount to making every position assignment possible.

Returning to our example in Figure 2, it can be easily observed that the model on the right hand side is obtained by updating the model on the left hand side with $\{\pi_1, \pi_2\}$, and conversely, the model on the left hand side is obtained by updating the model on the right hand side with $\{\pi_1\}$.

3.3 Rawls' Thesis

Recall that in the Original Position, we do not know anything about our position in society. So if, for a given model $\mathfrak{M}^0$ of choice under position uncertainty, we ask what an agent i would choose in the Original Position, we are in fact asking what i would choose in the updated model $\mathfrak{M}^0_{\Pi_*}$. On this analysis, Rawls' thesis says that a given choice is difference admissible in $\mathfrak{M}^0$ if and only if the "corresponding" choice in $\mathfrak{M}^0_{\Pi_*}$ is i-admissible in $\mathfrak{M}^0_{\Pi_*}$.

In order to make this notion of correspondence precise we need some extra notation. Given any model $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$, we let $\mathfrak{M}^0_* = \mathfrak{M}^0_{\Pi_*} = \langle W, \Pi_*, C^0, U^0, V^0\rangle$, and we use C_*, U_*, and V_* to refer to the set of choices, the agent-utility function, and the valuation function of the model $\mathfrak{M}_*$ of choice

under uncertainty that corresponds to $\mathfrak{M}^0_*$ (cf. Definition 3.2).

Our proof of Rawls' thesis crucially relies on the observation that the set of guaranteed utilities for the least well-off at a given state in the original model equals the set of guaranteed utilities for any individual i in the corresponding state in the Original Position. Formally:

Lemma 3.4 *Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ is a model of choice under position uncertainty, $s \in W \times \Pi$, and $i \in Agt$: $G^{\mathfrak{M}^0}(s, *) = G^{\mathfrak{M}^0_*}(s, i)$.*

Proof. Let $i \in Agt$ and $n \in N$. We have: $n \in G^{\mathfrak{M}^0}(s, *)$ iff [by the definition of $G^{\mathfrak{M}^0}(s, *)$] for all $s' \in C(s)$, $U(s', *) \geq n$ iff [by the definition of $U(*, s)$] for all $i \in Agt$ and $s' \in C(s)$, $U(s', i) \geq n$ iff [since every position assignment is surjective] for all $p \in P$ and all $w' \in C^0(w)$, $U^0(w', p) \geq n$ iff [by the definition of Π_*] for all $s' \in C_*(s)$, $U_*(s, i) \geq n$ iff [by the definition of $G^{\mathfrak{M}^0_*}(s, i)$] $n \in G^{\mathfrak{M}^0_*}(s, i)$. □

Note also that, whatever utility is guaranteed for i at a state (w, π) in a model $\mathfrak{M}^0$, is also guaranteed for i at every state (w, π') in $\mathfrak{M}^0$. Formally:

Fact 3.5 *Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ is a model of choice under position uncertainty, $w \in W$, $\pi, \pi' \in \Pi$, and $i \in Agt \cup \{*\}$: $G^{\mathfrak{M}^0}((w, \pi), i) = G^{\mathfrak{M}^0}((w, \pi'), i)$.*

Theorem 3.6 *Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ is a model of choice under position uncertainty, $(w, \pi) \in W \times \Pi$, and $i \in Agt$: $C(w, \pi) \in Adm_*(\mathfrak{M}^0)$ iff $C_*(w, \pi) \in Adm_i(\mathfrak{M}^0_*)$. (Rawls' Thesis)*

Proof. $C(w, \pi) \in Adm_*(\mathfrak{M}^0)$ iff [by Lemma 2.6] for all $(w', \pi') \in W \times \Pi$, $G^{\mathfrak{M}^0}((w', \pi'), *) \subseteq G^{\mathfrak{M}^0}((w, \pi), *)$ iff [by Lemma 3.4] for all $(w', \pi') \in W \times \Pi$, $G^{\mathfrak{M}^0_*}((w', \pi'), i) \subseteq G^{\mathfrak{M}^0_*}((w, \pi), i)$ iff [by Fact 3.5] for all $(w', \pi') \in W \times \Pi_*$, $G^{\mathfrak{M}^0_*}((w', \pi'), i) \subseteq G^{\mathfrak{M}^0_*}((w, \pi), i)$ iff [by Lemma 2.6] $C_*(w, \pi) \in Adm_i(\mathfrak{M}^0_*)$. □

4 A Logic of Choice Under Position Uncertainty

In order to express Rawls' thesis syntactically, we enrich the formal language $\mathfrak{L}$ from Section 2. First, we define a static modal language in which we can express position utilities and position assignments, and provide an axiomatization for the resulting logic (Section 4.1). Next, we add dynamic operators that can express changes in position uncertainty and give reduction axioms for them (Section 4.2). After this preparatory work, we show that Rawls' thesis corresponds to a validity of the resulting logic (Section 4.3).

4.1 Static Part

Formal Language Let $\mathfrak{L}^+$ be defined by the BNF:

$$\varphi := q \mid \mathsf{a}_{ip} \mid \mathsf{u}^n_p \mid \neg\varphi \mid \varphi \vee \varphi \mid \Box\varphi \mid \boxdot\varphi \mid \boxplus\varphi$$

where q ranges over Q, i over N, p over P, and n over N. The constant a_{ip} expresses that agent i occupies position p, while u^n_p expresses that any agent

with position p gets utility n. The only new modality is $\boxplus$. This operator allows us to talk about all states that have the same world component (see (SC9) below). In other words, $\boxplus\varphi$ expresses that "φ is the case, no matter which position the agents occupy".

The following definition gives the semantic clauses for a_{ip}, u_p^n, and $\boxplus$; note that the clauses for the variables, connectives, and other operators are exactly as in Definition 2.5, relying on the fact that every model of choice under position uncertainty is also a model of choice under uncertainty (cf. Definition 3.2).

Definition 4.1 Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ is a model of choice under position uncertainty and $s = (w, \pi) \in W \times \Pi$,

(SC7) $\mathfrak{M}^0, s \models \mathsf{a}_{ip}$ iff $\pi(i) = p$

(SC8) $\mathfrak{M}^0, s \models \mathsf{u}_p^n$ iff $U^0(w, p) = n$

(SC9) $\mathfrak{M}^0, s \models \boxplus\varphi$ iff for all $\pi \in \Pi$, $\mathfrak{M}^0, (w, \pi) \models \varphi$

The formal language introduced above is an extension of $\mathfrak{L}$. The constants that expressed agent-utilities in $\mathfrak{L}$ can now be defined:

$$\mathsf{u}_i^n \quad =_{\mathsf{df}} \bigvee_{p\in P} (\mathsf{a}_{ip} \wedge \mathsf{u}_p^n)$$

Consequently, we can reuse all the definitions from Table 2 to express that a given choice is i-admissible or difference admissible in $\mathfrak{L}^+$. However, we now also have the additional expressive power that allows us to talk about position uncertainty, which is crucial for Rawls' thesis.

Axiomatization The set of validities in $\mathfrak{L}^+$ is axiomatized by the axioms and rules in Table 3. Axiom QW (resp. PW) expresses that the truth of a propositional variable (resp. the utility of a position) depends only on the world-component of a state. I1-I3 capture interactions between the various modalities. I1 is an immediate result of the fact that $\Box$ is a universal modality. I2 follows from the fact that choices are defined in terms of the world-components of states, and hence one cannot choose between two states with the same world-component. I3 captures the property that, if a certain position assignment π is possible in the model at hand, then there is some state with the same world component as the current state and the possition assignment π. Finally, PA1 and PA2 (PU1 and PU2) express that every π (U) is a function; PA3 expresses that every π is surjective.

Theorem 4.2 $\vdash \varphi$ *iff* $\models \varphi$. *(Soundness and Completeness)*

Proof. Soundness is a matter of routine. For completeness, observe that every model $\mathfrak{M}^0$ of choice under position uncertainty can be rewritten as a Kripke-model of the type $\mathfrak{M}^K = \langle S, \sim^{\boxdot}, \sim^{\boxplus}, V\rangle$, where $S \neq \emptyset$ is a set of states, $\sim^{\boxdot}$ is the equivalence relation that corresponds to the choices in $\mathfrak{M}^0$, $\sim^{\boxplus}$ is the equivalence relation that corresponds to the worlds in $\mathfrak{M}^0$, and $V : Q \cup \{\mathsf{a}_{ip} \mid i \in Agt, p \in P\} \cup \{u_p = n \mid p \in P, n \in N\} \to \wp(S)$ is a valuation function. Conversely, given suitable conditions on such Kripke-models, we can

CL	Classical Logic	
S5	S5 for $\blacksquare \in \{\Box, \boxdot, \boxplus\}$	
QW	$\boxplus q \vee \boxplus \neg q$	$(q \in Q)$
PW	$\boxplus \mathsf{u}_p^n \vee \boxplus \neg \mathsf{u}_p^n$	$(p \in P, n \in N)$
I1	$\Box\varphi \rightarrow \boxdot\varphi$	
I2	$\boxdot\varphi \rightarrow \boxplus\varphi$	
I3	$\boxplus \bigwedge_{i \in Agt, \pi(i)=p} \mathsf{a}_{ip} \rightarrow \Box \bigwedge_{i \in Agt, \pi(i)=p} \mathsf{a}_{ip}$	$(\pi \in \Pi_*)$
PA1	$\bigvee_{p \in P} \mathsf{a}_{ip}$	$(i \in Agt)$
PA2	$\mathsf{a}_{ip} \rightarrow \neg \mathsf{a}_{ip'}$	$(i \in Agt, p, p' \in P, p \neq p)$
PA3	$\bigvee_{i \in Agt} \mathsf{a}_{ip}$	$(p \in P)$
PU1	$\bigvee_{n \in N} \mathsf{u}_p^n$	$(p \in P)$
PU2	$\mathsf{u}_p^n \rightarrow \neg \mathsf{u}_p^m$	$(p \in P, n, m \in N, n \neq m)$
MP	if $\vdash \varphi \rightarrow \psi$ and $\vdash \varphi$ then $\vdash \psi$	
NEC	if $\vdash \varphi$ then $\vdash \Box\varphi$	

Table 3

(C)	if $s \sim^{\boxplus} s'$ and $s, s' \in \bigcap_{i \in Agt, \pi(i)=p} V(\mathsf{a}_{ip})$, then $s = s'$	
(CQW)	if $s \sim^{\boxplus} s'$, then $s \in V(q)$ iff $s' \in V(q)$	
(CPW)	if $s \sim^{\boxplus} s'$, then $s \in V(\mathsf{u}_p^n)$ iff $s' \in V(\mathsf{u}_p^n)$	
(CI2)	$\sim^{\boxplus} \subseteq \sim^{\boxdot}$	
(CI3)	if $s \in \bigcap_{i \in Agt, \pi(i)=p} V(\mathsf{a}_{ip})$, then $\forall s' \in S$, $\exists s'' \in S$: $s' \sim^{\boxplus} s''$ and $s'' \in \bigcap_{i \in Agt, \pi(i)=p} V(\mathsf{a}_{ip})$	$(\pi \in \Pi_*)$
(CPA1)	$\forall i \in Agt, \exists p \in P$: $s \in V(\mathsf{a}_{ip})$	
(CPA2)	if $s \in V(\mathsf{a}_{ip})$, then $s \notin V(\mathsf{a}_{ip'})$	$(p \neq p')$
(CPA3)	$\forall p \in P, \exists i \in Agt$: $s \in V(\mathsf{a}_{ip})$	
(CPU1)	$\forall p \in P, \exists n \in N$: $s \in V(\mathsf{u}_p^n)$	
(CPU2)	if $s \in V(\mathsf{u}_p^n)$, then $s \notin V(\mathsf{u}_p^m)$	$(n \neq m)$

Table 4

rewrite them as models of position uncertainty — cf. Table 4. Proving that, taken jointly, these conditions ensure translatability to a model of choice under position uncertainty is a tedious but routine job, which we omit for reasons of space.

Let MCS be the set of all maximal consistent subsets of $\mathfrak{L}^+$. Where $\blacksquare \in \{\Box, \boxdot, \boxplus\}$ and $\Delta \in \mathsf{MCS}$, let $\Delta^{\blacksquare} = \{\blacksquare\varphi \in \mathfrak{L}^+ \mid \blacksquare\varphi \in \Delta\}$. Fix a $\Gamma \in \mathsf{MCS}$. Let $\mathfrak{M}_\Gamma^K = \langle S_\Gamma, \sim_\Gamma^{\boxdot}, \sim_\Gamma^{\boxplus}, V_\Gamma \rangle$, where

1. S_Γ is the set of all maximal consistent sets Δ such that $\Delta^{\Box} = \Gamma^{\Box}$

2. Where $\Delta, \Theta \in S_\Gamma$, $\Delta \sim_\Gamma^{\boxed{\mathbb{C}}} \Theta$ iff $\Delta^{\boxed{\mathbb{C}}} = \Theta^{\boxed{\mathbb{C}}}$
3. Where $\Delta, \Theta \in S_\Gamma$, $\Delta \sim_\Gamma^{\boxplus} \Theta$ iff $\Delta^{\boxplus} = \Theta^{\boxplus}$
4. $V_\Gamma(\varphi) = \{\Delta \in S_\Gamma \mid \varphi \in \Delta\}$ for all $\varphi \in Q \cup \{\mathsf{a}_{ip} \mid i \in Agt, p \in P\} \cup \{u_p = n \mid p \in P, n \in N\}$

The truth lemma is proven for $\mathfrak{M}_\Gamma^K$ in the standard way. By an induction on the complexity of formulas, we can moreover prove that for any $s, s' \in S_\Gamma$, condition (C) is satisfied. For the other conditions, we can rely on the corresponding axioms to prove they hold for $\mathfrak{M}_\Gamma^K$. In sum, $\mathfrak{M}_\Gamma^K$ satisfies all the conditions from Table 4. □

4.2 Dynamic Operators

In order to express what holds given an update of the set of position assignments, we rely on well-known ideas from dynamic epistemic logic [3,20]. In particular, we consider *pointed* updates of *pointed* models. As we will show in Section 4.3, we can use the resulting dynamic operators to express what holds in the Original Position.

Henceforth, an *update model* is a couple (Π, π), where $\Pi \subseteq \Pi_*$ and $\pi \in \Pi$. Intuitively, the update model expresses the new set of position assignments that become possible, and the specific position assignment that becomes actual. Update models are used to change a given pointed model of position uncertainty, i.e. a model together with a given state (w, π) in that model:

Definition 4.3 Where $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$, $(w, \pi) \in W \times \Pi$, and (Π', π') is an update model: the *update of* $(\mathfrak{M}^0, (w, \pi))$ *with* (Π', π') is $(\mathfrak{M}^0, (w, \pi)) \circ (\Pi', \pi') =_{\mathsf{df}} (\mathfrak{M}^0_{\Pi'}, (w, \pi'))$.

Given these conventions, we can introduce dynamic operators $[\Pi, \pi]$ for every update model (Π, π), and interpret them using the following standard clause:

(SC10) $\mathfrak{M}^0, s \models [\Pi, \pi]\varphi$ iff $(\mathfrak{M}^0, s) \circ (\Pi, \pi) \models \varphi$

In dynamic epistemic logic terminology, our updates are a specific type of (finitary) ontic updates with an empty precondition. Relying on this observation, we can easily find reduction axioms for the dynamic operators. These are listed in Table 5. Given these reduction axioms and Theorem 4.2, we obtain a sound and strongly complete axiomatization for the extension of $\mathfrak{L}^+$ with all dynamic operators of the type $[\Pi, \pi]$.

4.3 Rawls' Thesis in $\mathfrak{L}^+$

Recall that Π_* denotes the set of all position assignments. By means of the dynamic operators $[\Pi_*, \pi]$ we can define an operator that expresses what holds in the Original Position:

$$\boxtimes\varphi =_{\mathsf{df}} \bigwedge_{\pi \in \Pi_*} \Big(\bigwedge_{i \in Agt, \pi(i) = p} \mathsf{a}_{ip} \rightarrow [\Pi_*, \pi]\varphi \Big)$$

RA1	$[\Pi, \pi]q \leftrightarrow q$ (for all $q \in Q$)
RA2	$[\Pi, \pi]\mathsf{u}_p^n \leftrightarrow \mathsf{u}_p^n$ (for all $p \in P$ and $n \in N$)
RA3	$[\Pi, \pi]\mathsf{a}_{ip} \leftrightarrow \top$ if $\pi(i) = p$
RA4	$[\Pi, \pi]\mathsf{a}_{ip} \leftrightarrow \bot$ if $\pi(i) \neq p$
RA5	$[\Pi, \pi]\neg\varphi \leftrightarrow \neg[\Pi, \pi]\varphi$
RA6	$[\Pi, \pi](\varphi \vee \psi) \leftrightarrow ([\Pi, \pi]\varphi \vee [\Pi, \pi]\psi)$
RA7	$[\Pi, \pi]\blacksquare\varphi \leftrightarrow \bigwedge_{\pi' \in \Pi} \blacksquare[\Pi, \pi']\varphi$ (for $\blacksquare \in \{\Box, \boxdot, \boxplus\}$)

Table 5
Reduction axioms for the dynamic operators.

Theorem 4.4 *Where $\mathfrak{M}^0$ and $\mathfrak{M}^0_*$ are models of choice under position uncertainty, we have: $\mathfrak{M}^0, (w, \pi) \models \boxtimes\varphi$ iff $\mathfrak{M}^0_*, (w, \pi) \models \varphi$.*

Proof. For all $\pi \in \Pi$, let $\mathsf{a}_\pi = \bigwedge_{i \in Agt, \pi(i)=p} \mathsf{a}_{ip}$. We have: $\mathfrak{M}^0, (w, \pi) \models \boxtimes\varphi$ iff [by the definition of $\boxtimes$] for all $\pi' \in \Pi_*$, $\mathfrak{M}^0, (w, \pi) \models \mathsf{a}_{\pi'} \to [\Pi_*, \pi']\varphi$ iff [since only a_π is true at $\mathfrak{M}^0, (w, \pi)$] $\mathfrak{M}^0, (w, \pi) \models [\Pi_*, \pi]\varphi$ iff [by the semantic clause for $[\Pi, \pi]$] $(\mathfrak{M}^0, (w, \pi)) \circ (\Pi_*, \pi) \models \varphi$ iff [by the definition of pointed updates and since $\mathfrak{M}^0_* = \mathfrak{M}^0_{\Pi_*}$] $\mathfrak{M}^0_*, (w, \pi) \models \varphi$. □

Theorem 3.6 is now expressible as a theorem in the object-language:

Theorem 4.5 $\models \mathsf{Adm}_* \leftrightarrow \boxtimes\mathsf{Adm}_i$. *(Rawls' Thesis in $\mathfrak{L}^+$)*

Proof. Let $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$ and $s \in W \times \Pi$. We have: $\mathfrak{M}^0, s \models \mathsf{Adm}_*$ iff [by Theorem 2.7] $C(s) \in Adm_*(\mathfrak{M}^0)$ iff [by Theorem 3.6] $C_*(s) \in Adm_i(\mathfrak{M}^0_*)$ iff [by Theorem 2.7] $\mathfrak{M}^0_*, s \models \mathsf{Adm}_i$ iff [by Theorem 4.4] $\mathfrak{M}^0, s \models \boxtimes\mathsf{Adm}_i$. □

5 Deontic Logics Based on Fairness

In this last, somewhat more programmatic section, we show the potential of our models and logic from the viewpoint of deontic logic. We first show how admissibility based on the other lifting criteria can be formalized in $\mathfrak{L}$ (Section 5.1). This in turn gives us a general recipe for expressing various other notions of fairness (Section 5.2), and deontic operators based on them (Section 5.3).

5.1 Other Lifting Criteria

In Section 2 we introduced four lifting criteria that can be used to determine which choices are admissible in a given choice situation. Moreover, we demonstrated that the current choice being admissible for i according to the Maximin lifting ($\forall\exists$) can be expressed in the language using object-level formulas. In Table 6, we give an overview of how admissibility of the current choice for i can be expressed for the other three lifting criteria from Section 2. The logical relations between these notions are depicted in Figure 3, where the arrows stand for logical consequence.

Abbreviation	Definition
$\mathsf{u}_i^{\leq_n}$	$\bigvee_{m \leq n} \mathsf{u}_i^m$
$\mathsf{Adm}_i^{\forall\forall}$	$\bigwedge_{n \in N}\big(\boxdot \mathsf{u}_i^{\leq_n} \to ((\Diamond \mathsf{g}_i^n \to \mathsf{g}_i^n) \wedge \Box(\mathsf{g}_i^n \to \boxdot \mathsf{u}_i^n))\big)$
$\mathsf{Adm}_i^{\exists\forall}$	$\bigwedge_{n \in N}(\Diamond \textcircled{\scriptsize c} \mathsf{u}_i^{\geq_n} \to \textcircled{\scriptsize c} \mathsf{u}_i^{\geq_n})$
$\mathsf{Adm}_i^{\exists\exists}$	$\bigwedge_{n \in N}(\Diamond \mathsf{g}_i^n \to \textcircled{\scriptsize c} \mathsf{u}_i^{\geq_n})$

Table 6

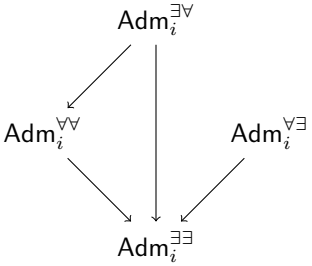

Fig. 3.

5.2 Other Notions of Fairness

Recall that the defined operator $\boxtimes$ talks about what holds in the Original Position (cf. Theorem 4.4). We can use this operator and our l-admissibility formulas to define three additional, distinct notions of fairness admissibility. That is, the formula $\boxtimes\mathsf{Adm}_i^l$ expresses that in the Original Position, if our standard of rational choice under uncertainty is determined by lifting criterion l, then the given choice is l-admissible for i. So, if one agrees with Rawls that fair choices are the choices a rational agent would make in the Original Position, then $\boxtimes\mathsf{Adm}_i^l$ expresses that the given choice is fair (modulo l).

The logical relations depicted in Figure 3 immediately transfer to the corresponding notions of fairness admissibility, in view of the following:

Theorem 5.1 *Where $i \in Agt$ and $l, l' \in \{\forall\forall, \forall\exists, \exists\forall, \exists\exists\}$: $\vdash \mathsf{Adm}_i^l \to \mathsf{Adm}_i^{l'}$ iff $\vdash \boxtimes\mathsf{Adm}_i^l \to \boxtimes\mathsf{Adm}_i^{l'}$.*

Proof. For left to right, one should show that $\boxtimes$ is a normal modal operator. For the other direction, suppose that $\nvdash \mathsf{Adm}_i^l \to \mathsf{Adm}_i^{l'}$. So there is a model $\mathfrak{M}^0$ and state s such that $\mathfrak{M}^0, s \models \mathsf{Adm}_i^l$ and $\mathfrak{M}^0, s \not\models \mathsf{Adm}_i^{l'}$. Consider the model $\mathfrak{M}_e^0$ that differs only from $\mathfrak{M}^0$ in that, at every state, *all* the agents receive the utility that i receives in the corresponding state in $\mathfrak{M}^0$. In this model, individual admissibility and fairness admissibility coincide, and hence $\mathfrak{M}_e^0, s \models \boxtimes\mathsf{Adm}_i^l$, $\mathfrak{M}_e^0, s \not\models \boxtimes\mathsf{Adm}_i^{l'}$. □

Thus, e.g. fairness admissibility using the Maximin criterion is strictly stronger than fairness admissibility using criterion $\forall\forall$, which in turn implies fairness admissibility using $\exists\exists$. In contrast, fairness using $\exists\forall$ is logically in-

comparable to fairness admissibility with Maximin or with $\forall\forall$.

5.3 Deontic Operators

By employing the familiar Kangerian reduction [1,2,12] we can use our admissibility formulas to define two types of deontic operators:

$$\mathsf{O}_i^l\varphi =_{\mathsf{df}} \Box(\mathsf{Adm}_i^l \to \varphi)$$

$$\mathsf{O}_*^l\varphi =_{\mathsf{df}} \Box(\boxtimes\mathsf{Adm}_i^l \to \varphi)$$

The formula $\mathsf{O}_i^l\varphi$ can be read as "it ought to be that φ for i" (where l determines a particular standard of rational choice under uncertainty). This contrasts with the formula $\mathsf{O}_*^l\varphi$ which can be read as "from the viewpoint of fairness, it ought to be that φ". Both O_i^l and O_*^l are normal modal operators in virtue of their definition.

Because the admissibility formulas stand in logical relations with each other, we can expect there to be logical relations between obligation statements as well. For example, we have:

Theorem 5.2 $\vdash \mathsf{Adm}_i^l \to \mathsf{Adm}_i^{l'}$ *iff* $\vdash \mathsf{O}_i^{l'}\varphi \to \mathsf{O}_i^l\varphi$.

Proof. Left to right of the equivalence is safely left to the reader. For the other direction, let $\varphi = \mathsf{Adm}_i^{l'}$. Then, the right hand side implies that $\vdash \Box(\mathsf{Adm}_i^l \to \mathsf{Adm}_i^{l'})$ and hence, by the T-axiom for $\Box$, $\vdash \mathsf{Adm}_i^l \to \mathsf{Adm}_i^{l'}$. □

Consequently, if Adm_i^l and $\mathsf{Adm}_i^{l'}$ are incomparable, then $\mathsf{O}_i^l\varphi$ and $\mathsf{O}_i^{l'}\varphi$ are incomparable as well. We can also expect there to be logical relations between the individual oughts and fairness oughts, in line with Rawls' thesis. For example, what ought to be for agent i (given the Maximin criterion) and what ought to be from the viewpoint of fairness coincide in the Original Position:

Theorem 5.3 $\vdash \bigwedge_{\pi\in\Pi_*} \Diamond \bigwedge_{i\in Agt,\pi(i)=p} \mathsf{a}_{ip} \to (\mathsf{O}_i^{\forall\exists}\varphi \leftrightarrow \mathsf{O}_*^{\forall\exists}\varphi)$.

Proof. Note that, if the left hand side of the implication is true in a model $\mathfrak{M}^0 = \langle W, \Pi, C^0, U^0, V^0\rangle$, then $\Pi = \Pi^*$. By our earlier results, individual admissibility and fairness admissibility coincide in such models, and hence so do the corresponding ought-operators. □

To summarize, by using a Kangerian reduction, we obtain various kinds of deontic logics, based on individual standards of rationality and Rawls' procedural account of fairness. All these logics are fragments of the logic presented in Section 4. Here we merely sketched the various possibilities this generates; we leave a full investigation for future work.

6 Conclusion

We have given a logical analysis of Rawls' thesis that choices motivated by the Difference Principle coincide with the choices of any rational individual in the Original Position. In particular, we presented models of choice under position uncertainty, inspired by simple models for STIT logic. With the help of these models and a suitable formal language, we showed how to capture Rawls' thesis

both in semantic and in syntactic terms. Finally, we demonstrated the potential of our logical analysis for the study of deontic notions related to fairness.

Future Work We chose to work with a finite set of utility values as this removes some complexities. However, one may ask to which extent our results still go through when working with infinite sets of utility values such as $\mathbb{N}$ or $\mathbb{R}$. While the semantic results (e.g. Theorem 3.3) seem easy to generalize to such richer settings, this is far less obvious on the syntactic side. In particular, can the language be modified in order to cope with infinite sets of values, while keeping the logic well-behaved meta-theoretically (e.g. axiomatizable and compact)?

We focused on the four lifting criteria from Table 1. An open question is whether it is possible to express more complex lifting criteria, such as e.g. lexicographic preferences. Finally, both the notion of choice and the notion of uncertainty are agent-independent in our models. A natural generalization would be to have models where the choices and/or uncertainty are agent-dependent. Here again, semantics seem relatively easy to obtain, but complexity grows rapidly at the syntactic level.

References

[1] Anderson, A. R., *Some nasty problems in the formal logic of ethics*, Noûs (1967), pp. 345–360.

[2] Åqvist, L., *Deontic logic*, in: *Handbook of philosophical logic*, Springer, 2002 pp. 147–264.

[3] Baltag, A., L. S. Moss and S. Solecki, *The logic of public announcements, common knowledge, and private suspicions*, in: *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK 98 (1998), pp. 43–56.

[4] Belnap, N. D., M. Perloff, M. Xu et al., "Facing the future: agents and choices in our indeterminist world," Oxford University Press on Demand, 2001.

[5] Chellas, B. F., "The Logical Form of Imperatives," Ph.D. thesis, Stanford University (1969).

[6] Chung, H., *Rawlss self-defeat: A formal analysis*, Erkenntnis (2018), pp. 1–29.

[7] Gaus, G. and J. Thrasher, *Rational choice and the original position: the (many) models of Rawls and Harsanyi* (2015).

[8] Giraud, G. and C. Renouard, *Is the veil of ignorance transparent?* (2010).

[9] Harsanyi, J. C., *Can the maximin principle serve as a basis for morality? a critique of john rawls's theory*, American political science review **69** (1975), pp. 594–606.

[10] Horty, J. F., "Agency and deontic logic," Oxford University Press, 2001.

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

[12] Kanger, S., *New foundations for ethical theory*, in: *Deontic logic: Introductory and systematic readings*, Springer, 1970 pp. 36–58.

[13] Kariv, S. and W. R. Zame, *Piercing the veil of ignorance* (2008).

[14] Kooi, B. and A. Tamminga, *Moral conflicts between groups of agents*, Journal of Philosophical Logic **37** (2008), pp. 1–21.

[15] Kukathas, C. and P. Pettit, "Rawls: A Theory of Justice and Its Critics," Key contemporary thinkers, Polity, 1990.

[16] Rawls, J., "A Theory of Justice," Belknap Press, 1999, rev sub edition.

[17] Svensson, L.-G., *Fairness, the veil of ignorance and social choice*, Social Choice and Welfare **6** (1989), pp. 1–17.

[18] Tamminga, A., *Deontic logic for strategic games*, Erkenntnis **78** (2013), pp. 183–200.

[19] Van Benthem, J., P. Girard and O. Roy, *Everything else being equal: A modal logic for ceteris paribus preferences*, Journal of philosophical logic **38** (2009), pp. 83–125.
[20] Van Benthem, J., J. van Eijck and B. Kooi, *Logics of communication and change*, Information and Computation **204** (2006), pp. 1620 – 1662.