Input/Output STIT Logic for Normative Systems

Xin Sun

Faculty of Science, Technology and Communication, University of Luxembourg
xin.sun@uni.lu

Abstract. In this paper we study input/output STIT logic. We introduce the semantics, proof theory and prove the completeness theorem. Input/output STIT logic has more expressive power than Makinson and van der Torre's input/output logic. We show that input/output STIT logic is decidable and free from Ross' paradox.

Key words: input/output logic, STIT, norm

1 Introduction

In recent years, normative multi-agent system [7, 2] arises as a new interdisciplinary academic area bringing together researchers from multi-agent system [22], deontic logic [9] and normative system [1, 10]. Norms play an important role in normative multi-agent system. They are heavily used in agent cooperation and coordination, group decision making, multi-agent organizations, electronic institutions, and so on.

In the first volume of the handbook of deontic logic and normative systems [9], input/output logic [14–17] appears as one of the new achievement in deontic logic in recent years. Input/output logic takes its origin in the study of conditional norms. Unlike the modal logic framework, which usually uses possible world semantics, input/output logic adopts mainly operational semantics: a normative system is conceived in input/output logic as a deductive machine, like a black box which produces normative statements as output, when we feed it descriptive statements as input.

Boella and van der Torre [6] extends input/output logic to reasoning about constitutive norms. Tosatto *et al.* [8] adapts it to represent and reason about abstract normative systems. For a comprehensive introduction to input/output logic, see Parent and van der Torre [17]. A technical toolbox to build input/output logic is developed in Sun [21].

One limitation of Makinson and van der Torre's input/output logic is that it uses propositional logic as its base logic. Such treatment restricts its expressive power. For example, concepts such as agent, action and ability which are crucial for agent theory and multi-agent system, are unable to be expressed in input/output logic. To overcome this limitation, we need a more expressive logic to be the base of input/output logic.

STIT theory or STIT logic [5], is one of the most prominent accounts of agency in philosophy of action. It is the logic of constructions of the form "agent i sees to it that ϕ holds". STIT logic has strong expressive power. Notions like agent, action and ability can be expressed in STIT logic. Therefore STIT logic is an ideal candidate to build new input/output logic. But there are various STIT logic: individual STIT and group STIT, achieve STIT and deliberative STIT. In this paper we choose *individual deliberative* STIT logic as the basis to develop input/output logic. We make this choice for the following reasons:

1. Compared to Makinson and van der Torre's input/output logic, this input/output STIT logic has more expressive power.
2. Choosing individual STIT makes our logic decidable, while if we choose group STIT we lose decidability.
3. By choosing deliberative STIT, our logic is free from a well known paradox, Ross' paradox, which is a challenge for lots of deontic logic, including Makinson and van der Torre's input/output logic. If we choose achieve STIT, we are not free from Ross' paradox.

The structure of this paper is as follows: we recap some background knowledge, including some basic concepts and results of STIT logic, in the Section 2. Then in Section 3 and 4 we study the proof theory, semantics, completeness and decidability of input/output STIT logic. We show that input/output STIT logic solves Ross' paradox in Section 5. We discuss research avenues for future work and conclude this paper in Section 6.

2 Background

Given a countable set $\mathbb{P}$ of propositional letters and a finite set Agt of agents, the language of individual STIT logic $\mathfrak{L}$ is defined by the following BNF: for every $p \in \mathbb{P}$ and $i \in Agt$,

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid [i^d]\varphi \mid \Box\varphi$$

Intuitively $[i^d]\varphi$ is read as "agent i deliberately sees to it that φ" , $\Box\varphi$ is read as "necessary φ". We use $[i]\varphi$, read as "agent i successfully sees to it that φ", as an abbreviation of $[i^d]\varphi \vee \Box\varphi$. We use $\Diamond\varphi$ to represent $\neg\Box\neg\varphi$.

In the literature $[i^d]$ is called "deliberative STIT" and $[i]$ is called "achieve STIT" (or Chellas' STIT). Intuitively, $[i]\varphi$ simply means i sees to it that φ holds, while $[i^d]\varphi$ means i not only sees to it that φ holds, but also φ can be false without the action of i. Deliberative STIT and achieve STIT are inter-definable because $[i]\phi$ is equivalent to $[i^d]\varphi \vee \Box\varphi$, while $[i^d]\phi$ is equivalent to $[i]\varphi \wedge \neg\Box\varphi$. We will introduce the semantics and axiomatic system via achieve STIT, and build our input/output logic on deliberative STIT.

In STIT logic, actions are expressed as relations between agents and effects: $[i]\phi$ is an action which means "agent i ensures the world is among those satisfying ϕ". Agent's ability is expressed by $\Diamond[i]\varphi$ meaning that agent i has the ability to ensure the world is among those satisfying ϕ.

The semantics of STIT logic is originally defined by the branching-time choice structure. A simpler possible world semantics for group STIT is proposed by Kooi and Tamminga [13]. Here we simplify it for individual STIT.

Definition 1 (Possible world semantics). *A model is a tuple* $M = (W, Choice, V)$*, where*

1. W *is a nonempty set of possible worlds,*
2. $V : \mathbb{P} \mapsto 2^W$ *is the valuation for propositional letters.*
3. $Choice$ *is a choice function which satisfies the following conditions:*
   (a) *for every* $i \in Agt$ *it holds that* $Choice(i)$ *is a partition of* W*;*
   (b) *for* $Agt = \{1, ..., n\}$*, for every* $x_1 \in Choice(1), \ldots, x_n \in Choice(n)$*,* $x_1 \cap \ldots \cap x_n \neq \emptyset$*;*

Let R_i be the equivalence relation induced by $Choice(i)$. That is, $(w, w') \in R_i$ iff there is $K \in Choice(i)$ such that $\{w, w'\} \subseteq K$. Given a model M and a world $w \in M$, formulas of $\mathfrak{L}$ is evaluated as follows:

- $M, w \models p$ iff $w \in V(p)$ for all $p \in \mathbb{P}$.
- $M, w \models \neg\varphi$ iff not $M, w \models \varphi$.
- $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$.
- $M, w \models \Box\varphi$ iff $M, w' \models \varphi$ for all $w' \in W$.
- $M, w \models [i]\varphi$ iff $M, w' \models \varphi$ for all w' such that $(w, w') \in R_i$.

A formula $\phi \in \mathfrak{L}$ is valid iff for all model M and all $w \in M$, if $M, w \models \phi$. ϕ is satisfiable iff there are some model M and some $w \in M$ such that $M, w \models \phi$. ϕ is a logical consequence of a set of formulas Φ if for all model M and all $w \in M$, if $M, w \models \psi$ for all $\psi \in \Phi$, then $M, w \models \phi$. The individual STIT logic is axiomatized by the following axioms [5, 4]:

1. all instances of propositional tautologies
2. the axiom schemas of $S5$ for $\Box$
3. the axiom schemas of $S5$ for every $[i]$
4. $\Box\varphi \rightarrow [i]\varphi$
5. $(\Diamond[0]\varphi_0 \wedge \ldots \wedge \Diamond[k]\varphi_k) \rightarrow \Diamond([0]\varphi_0 \wedge \ldots \wedge [k]\varphi_k)$

The derivation rules of STIT logic is modus ponens and necessitation for $\Box$. A formula φ is a derivable ($\vdash \varphi$) iff it is derivable via the above axiomatic system. We use $\psi \vdash \varphi$ to represent $\vdash \psi \rightarrow \varphi$.

Theorem 1 ([5]). *For every* $\phi \in \mathfrak{L}$*,* $\models \phi$ *iff* $\vdash \phi$*.*

The satisfiability problem of individual STIT logic is the following decision problem: given a formula ϕ, is ϕ satisfiable? Balbiani *et al* [4] show that this problem is solvable in exponential time by a non-deterministic Turing machine.

Theorem 2 ([4]). *The complexity of the satisfiability problem of individual STIT logic is in NEXPTIME.*

3 Input/output STIT logic

Input/output logic adopts mainly operational semantics. The procedure of operational semantics is divided into three stages. In the first stage, we have in hand a set of propositions (call it the input) as a description of the current state. We then apply logical operators to this set, say close the set by logical consequence. Then we pass this set to a deductive machine and we reach the second stage. In the second stage, the machine accepts the input and produces a set of propositions as output. In the third stage, we accept the output and apply logical operators to it. A more formal explanation relies on the following terminologies.

A normative system $N \subseteq \mathfrak{L} \times \mathfrak{L}$ is a set of ordered pairs of formulas. A pair $(\phi, \psi) \in N$, call it a norm, is read as "given ϕ, it ought to be ψ". N is viewed as a function (or a deductive machine) from $2^{\mathfrak{L}}$ to $2^{\mathfrak{L}}$ such that for a set Φ of formulas, $N(\Phi) = \{\psi \mid (\phi, \psi) \in N \text{ for some } \phi \in \Phi\}$. Let $Cn(\Phi) = \{\phi \in \mathfrak{L} : \Phi \models \phi\}$.

3.1 Simple-minded

Definition 2 (Simple-minded output). *Given a set of norms $N \subseteq \mathfrak{L} \times \mathfrak{L}$ and a set of formulas $\Phi \subseteq \mathfrak{L}$,*

$$O_1(N, \Phi) = Cn(N(Cn(\Phi))).$$

The idea behind simple-minded input/output STIT logic O_1 is: we first take a set of formulas representing facts, then we close it under logical consequence. We further pass this closed set to the deductive machine (*i.e.* the normative system). The deductive machine produces a set of formulas representing obligations. We finally close obligations under logical consequence.

Example 1. Suppose a, b, x, y are propositional letters, i, j are agents. Let $N = \{(a, [i]x), (a, [j]y), (b, x \wedge y)\}$. Then $O_1(N, \{a\}) = Cn(N(Cn(\{a\}))) = Cn(\{[i]x, [j]y\})$. □

On the proof-theoretical side, input/output STIT logics are characterized by derivation rules about norms. Given a set of norms N, a derivation system is the smallest set of norms which extends N and is closed under certain derivation rules. The following are the rules we will use:

- SI (strengthening the input): from (ϕ_1, ψ) to (ϕ_2, ψ) whenever $\models \phi_2 \rightarrow \phi_1$
- WO (weakening the output): from (ϕ, ψ_1) to (ϕ, ψ_2) whenever $\models \psi_1 \rightarrow \psi_2$
- AND (conjunction of the output): from (ϕ, ψ_1) and (ϕ, ψ_2) to $(\phi, \psi_1 \wedge \psi_2)$
- OR (disjunction of the input): from (ϕ_1, ψ) and (ϕ_2, ψ) to $(\phi_1 \vee \phi_2, \psi)$
- CT (cumulative transitivity): from (ϕ, ψ_1) and $(\phi \wedge \psi_1, \psi_2)$ to (ϕ, ψ_2)

The derivation system of simple-minded input/output STIT logic, $D_1(N)$, is decided by the rules SI, WO and AND. Adding OR to $D_1(N)$ gives $D_2(N)$, the derivation system of basic input/output STIT logic. Adding CT to $D_1(N)$ gives $D_3(N)$, the derivation system of simple-minded reusable input/output STIT logic. All the five rules together gives the derivation system of basic reusable input/output STIT logic.

Example 2. Suppose a, b, x, y are propositional letters, i, j are agents. Let $N = \{(a \vee b, [j]x)\}$, then $([i]b, [j](x \vee y)) \in D_1(N)$ because we have the following derivation

1.	$(a \vee b, [j]x)$	Assumption
2.	$([i]b, [j]x)$	1, SI
3.	$([i]b, [j](x \vee y))$	2, WO

Theorem 3. *Given $N \subseteq \mathfrak{L} \times \mathfrak{L}$, $\psi \in O_1(N, \{\phi\})$ iff $(\phi, \psi) \in D_1(N)$.*

Proof. Using technics from Sun [20], the proof is routine and here we omit it. $\dashv$

3.2 Basic

Simple-minded output O_1 is unable to process disjunctive input intelligently: from input $\Phi = \{\phi_1 \vee \phi_2\}$ and normative system $N = \{(\phi_1, \psi), (\phi_2, \psi)\}$ we don't have $\psi \in O_1(N, \Phi)$. Basic output O_2 strengthens O_1 to make up for such deficiency.

Definition 3 (Basic output). *Given a set of norms $N \subseteq \mathfrak{L} \times \mathfrak{L}$ and a set of formulas $\Phi \subseteq \mathfrak{L}$,*

$$O_2(N, \Phi) = \bigcap \{Cn(N(Cn(\Psi))) : \Phi \subseteq \Psi, \Psi \text{ is disjunctive}\}.$$

Here a set Ψ is disjunctive if for all $\phi \vee \psi \in \Psi$, either $\phi \in \Psi$ or $\psi \in \Psi$.

It can be verified that from input $\Phi = \{\phi_1 \vee \phi_2\}$ and normative system $N = \{(\phi_1, \psi), (\phi_2, \psi)\}$ we have $\psi \in O_2(N, \Phi)$. The following completeness theorem shows that O_2 corresponds to the derivation system D_2 where the rule disjunction of the input is involved.

Theorem 4. *Given $N \subseteq \mathfrak{L} \times \mathfrak{L}$, $\psi \in O_2(N, \{\phi\})$ iff $(\phi, \psi) \in D_2(N)$.*

Proof. ($\Rightarrow$) Assume $(\phi, \psi) \in D_2(N)$, we prove by induction on the length of derivation.

Base step: assume $(\phi, \psi) \in N$. Then $\psi \in N(\{\phi\}) \subseteq N(Cn(\phi)) \subseteq \bigcap \{N(Cn(\Psi)) : \phi \in \Psi, \Psi \text{ is disjunctive}\} \subseteq \bigcap \{ Cn(N(Cn(\Psi))) : \phi \in \Psi, \Psi \text{ is disjunctive}\} = O_2(N, \{\phi\})$.

Inductive step: here we only prove the case (ϕ, ψ) is derived by the OR rule in the last step of derivation. Other cases are easier. Assume there are $(\phi_1, \psi) \in D_2(N)$, $(\phi_2, \psi) \in D_2(N)$ and ϕ is $\phi_1 \vee \phi_2$. By induction hypothesis we know $\psi \in O_2(N, \phi_1)$ and $\psi \in O_2(N, \phi_2)$. Now for every set of formulas E such that $\phi \in E$ and E is disjunctive, we have $\phi_1 \vee \phi_2 \in E$ since ϕ is $\phi_1 \vee \phi_2$. Note that E is disjunctive, so we further have either $\phi_1 \in E$ or $\phi_2 \in E$. If $\phi_1 \in E$, then E is a disjunctive set contains ϕ_1. So we have $\psi \in O_2(N, \phi_1) = \bigcap \{Cn(N(Cn(B)) : \phi_1 \in B, B \text{ is disjunctive}\} \subseteq Cn(N(Cn(E)))$. Hence $\psi \in Cn(N(Cn(E)))$. If $\phi_2 \in E$, we can similarly deduce $\psi \in Cn(N(Cn(E)))$. Therefore no matter $\phi_1 \in E$ or $\phi_2 \in E$, we have $\psi \in Cn(N(Cn(E)))$. Therefore $\psi \in O_2(N, \phi)$.

($\Leftarrow$) Assume $\psi \in O_2(N, \phi)$, then $\psi \in \bigcap \{Cn(N(Cn(B)) : \phi \in B, B \text{ is disjunctive}\}$. Let $\{B_1, \ldots, B_n\}$ be the set of all minimal disjunctive extensions of $\{\phi\}$. Therefore we have $\psi \in Cn(N(Cn(B_i)))$ for each $i \in \{1, \ldots, n\}$.

Each B_i corresponds to a branch of the disjunctive parsing tree, defined in Definition 4, of ϕ. Note that formulas in B_i can be strictly ordered by their length. Let ϕ_i be the shortest formula of B_i. Then for each $\chi \in B_i$, $\models \phi_i \rightarrow \chi$.

Then we know $\psi \in Cn(N(Cn(\phi_i)))$. Hence there are $\psi_{i1}, \ldots, \psi_{ik} \in N(Cn(\phi_i))$ such that $\psi_{i1} \wedge \ldots \wedge \psi_{ik} \models \psi$. Then by SI we know $(\phi_i, \psi_{i1}), \ldots, (\phi_i, \psi_{ik}) \in D_2(N)$. Then by AND and WO we know $(\phi_i, \psi) \in D_2(N)$. Now by Lemma 2 we know $(\phi, \psi) \in D_2(N)$. ⊣

Definition 4 (disjunctive parsing tree). *Given a formula $\phi \in \mathfrak{L}$, the disjunctive parsing tree $P(\phi)$ is a tree such that:*

(a) ϕ is the root of $P(\phi)$.
(b) Every node which is not a leaf has arity 2.
(c) A node ψ has daughters ψ_1 and ψ_2 iff ψ is $\psi_1 \vee \psi_2$.
(d) We define the height for each node as follows: every leaf has height 0. If μ is a node with daughters ν_1, ν_2, then the height of μ is $max\{height(\nu_1), height(\nu_2)\} + 1$.

Lemma 1. *For every formula ϕ, every branch of $P(\phi)$ is a disjunctive set.*

Proof. Let B be an arbitrary branch of $P(\phi)$. For every $\phi_1 \vee \phi_2 \in B$, we know ϕ_1 and ϕ_2 are the only daughters of $\phi_1 \vee \phi_2$. Therefore B contains either ϕ_1 or ϕ_2. Hence B is disjunctive. ⊣

Lemma 2. *Let (ϕ, ψ) be a norm and N a normative system. If for every B_i which is a branch of $P(\phi)$, there exist $\phi_i \in B_i$ such that $(\phi_i, \psi) \in N$, then $(\phi, \psi) \in D_2(N)$.*

Proof. Since the length of ϕ is always finite, we know $P(\phi)$ is also finite. So we assume $\{B_1, \ldots, B_n\}$ is the set of all branches of $P(\phi)$.

Here we just consider the worst case, other cases are easier. In the worst case we have for every B_i, the element $\phi_i \in B_i$ such that $(\phi_i, \psi) \in N$ is of height 0. Then by applying the OR rule finitely many times we know that for every $\phi_i' \in B_i$ with $height(\phi_i') = 1$, $(\phi_i', \psi) \in D_2(N)$. Similarly we can deduce that for every $\phi_i'' \in B_i$ with $height(\phi_i'') = 2$, $(\phi_i'', \psi) \in D_2(N)$. This progress can go on and on and we will eventually have $(\phi, \psi) \in D_2(N)$ since the height of ϕ is finite. ⊣

3.3 Simple-minded reusable

In certain situations, it may be appropriate for outputs to be available for recycling as inputs. On the syntactic level, such a principle of reusability is expressed by the rule CT. On the semantic level, we define simple-minded reusable output O_3 to implement reusability.

Definition 5 (Simple-minded reusable output). *Given a set of norms $N \subseteq \mathfrak{L} \times \mathfrak{L}$ and a set of formulas $\Phi \subseteq \mathfrak{L}$, We define a function $f_\Phi^N : 2^{\mathfrak{L}} \rightarrow 2^{\mathfrak{L}}$ such that $f_\Phi^N(X) = Cn(\Phi \cup N(X))$, for all $X \in 2^{\mathfrak{L}}$. It can be proved that f_Φ^N is monotonic with respect to the set theoretical $\subseteq$ relation, and $(2^{\mathfrak{L}}, \subseteq)$ is a complete lattice. Then by Tarski's fixed point theorem there exist a least fixed point of f_Φ^N. Let B_Φ^N be the least fixed point of f_Φ^N,*

$$O_3(N, \Phi) = Cn(N(B_\Phi^N)).$$

We use B^N_ϕ as an abbreviation of $B^N_{\{\phi\}}$. The following theorem shows that the syntactic approach D_3 and the semantics approach O_3 coincide.

Theorem 5. *Given* $N \subseteq \mathfrak{L} \times \mathfrak{L}$, $\psi \in O_3(N, \{\phi\})$ *iff* $(\phi, \psi) \in D_3(N)$.

Proof. The proof mainly uses technics from Sun [20].
($\Leftarrow$) Assume $(\phi, \psi) \in D_3(N)$, then we prove by induction on the length of derivation.

- (Base step) Assume $(\phi, \psi) \in N$, then by Lemma 4 we have $\phi \in B^N_\phi$. Hence $\psi \in N(B^N_\phi) \subseteq Cn(N(B^N_\phi))$.
- Assume $(\phi, \psi) \in D_3(N)$ and it is derived by using SI from $(\chi, \psi) \in D_3(N)$ and $\models \phi \to \chi$. Then by inductive hypothesis we have $\psi \in Cn(N(B^N_\chi))$. By Lemma 6 we know $B^N_\chi \subseteq B^N_\phi$. Therefore we further have $N(B^N_\chi) \subseteq N(B^N_\phi)$, $Cn(N(B^N_\chi)) \subseteq Cn(N(B^N_\phi))$. Hence $\psi \in Cn(N(B^N_\phi))$.
- Assume $(\phi, \psi) \in D_3(N)$, ψ is $\psi_1 \wedge \psi_2$ and it is derived by using AND from (ϕ, ψ_1) and (ϕ, ψ_2). Then by inductive hypothesis we have $\psi_1 \in Cn(N(B^N_\phi))$ and $\psi_2 \in Cn(N(B^N_\phi))$. Therefore $\psi_1 \wedge \psi_2 \in Cn(N(B^N_\phi))$.
- Assume $(\phi, \psi) \in D_3(N)$ and it is derived by using WO from $(\phi, \psi_1) \in D_3(N)$ and $\models \psi_1 \to \psi$. Then by inductive hypothesis we have $\psi_1 \in Cn(N(B^N_\phi))$. Since $\models \psi_1 \to \psi$, we can prove that $\psi \in Cn(N(B^N_\phi))$.
- Assume $(\phi, \psi) \in D_3(N)$ and it is derived by using CT form $(\phi, \psi_1) \in D_3(N)$ and $(\phi \wedge \psi_1, \psi) \in D_3(N)$. Then by inductive hypothesis we have $\psi_1 \in Cn(N(B^N_\phi))$ and $\psi \in Cn(N(B^N_{\phi\wedge\psi_1}))$. Then by Lemma 8 we have $B^N_\phi = B^N_{\phi\wedge\psi_1}$. Therefore $\psi \in Cn(N(B^N_\phi))$.

($\Rightarrow$) Assume $\psi \in Cn(N(B^N_\phi))$, then there exist $\psi^1, \ldots, \psi^n \in N(B^N_\phi)$ such that $\psi^1 \wedge \ldots \wedge \psi^n \models \psi$. For each $i \in \{1, \ldots, n\}$, from $\psi^i \in N(B^N_\phi)$ we know there is $\phi^i \in B^N_\phi$ such that $(\phi^i, \psi^i) \in N$. From $\phi^i \in B^N_\phi$ we know there exist k such that $\phi^i \in B^N_{\phi,k}$. Now by Lemma 9 we know $(\phi, \psi^i) \in D_3(N)$. Then by applying the AND rule we have $(\phi, \psi^1 \wedge \ldots \psi^n) \in D_3(N)$. Then by the WO rule we have $(\phi, \psi) \in D_3(N)$. ⊣

Lemma 3. $B^N_\Phi = \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$, *where* $B^N_{\Phi,0} = Cn(\Phi), B^N_{\Phi,i+1} = Cn(\Phi \cup N(B^N_{\Phi,i}))$.

Proof. We first prove that $\bigcup_{i=0}^{\infty} B^N_{\Phi,i}$ is a fixed point of f^N_Φ. We prove by showing the following:

1. $\Phi \subseteq \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$: this is because $\Phi \subseteq Cn(\Phi) = B^N_{\Phi,0} \subseteq \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$.
2. $N(\bigcup_{i=0}^{\infty} B^N_{\Phi,i}) \subseteq \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$: For every $\phi \in N(\bigcup_{i=0}^{\infty} B^N_{\Phi,i})$, there exist k such that $\phi \in N(B^N_{\Phi,k}) \subseteq B^N_{\Phi,k+1} \subseteq \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$.
3. $Cn(\bigcup_{i=0}^{\infty} B^N_{\Phi,i}) = \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$: the right-to-left direction is obvious; for the other direction: assume $\phi \in Cn(\bigcup_{i=0}^{\infty} B^N_{\Phi,i})$, then there exist $\phi_1, \ldots \phi_n \in \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$ such that $\models \phi_1 \wedge \ldots \wedge \phi_n \to \phi$. Therefore there exist k such that $\phi_1, \ldots \phi_n \in B^N_{\Phi,k}$. Hence $\phi \in B^N_{\Phi,k+1} \subseteq \bigcup_{i=0}^{\infty} B^N_{\Phi,i}$.

With the above items in hand, we can prove that $f^N_\Phi(\bigcup_{i=0}^\infty B^N_{\Phi,i}) \subseteq \bigcup_{i=0}^\infty B^N_{\Phi,i}$. For the other direction, we prove by induction on i that for every i, $B^N_{\Phi,i} \subseteq f^N_\Phi(\bigcup_{i=0}^\infty B^N_{\Phi,i})$. Here we omit the details.
So we have proved that $\bigcup_{i=0}^\infty B^N_{\Phi,i}$ is a fixed point of f^N_Φ. To prove that it is the least fixed point, we can again prove by induction that for every i, $B^N_{\Phi,i} \subseteq f^N_\Phi(B)$, where B is a fixed point of f^N_Φ. Here we omit the details. $\dashv$

Lemma 4. *For every $\Phi \subseteq \mathfrak{L}, N \subseteq \mathfrak{L} \times \mathfrak{L}$, $\Phi \subseteq B^N_\Phi$.*

Proof. By Lemma 3, the proof is trivial. $\dashv$

Lemma 5. *For every $\phi \in \mathfrak{L}, N \subseteq \mathfrak{L} \times \mathfrak{L}$, $B^N_\phi = Cn(B^N_\phi)$.*

Proof. By Lemma 3, the proof is easy. $\dashv$

Lemma 6. *For every $\phi, \psi \in \mathfrak{L}, N \subseteq \mathfrak{L} \times \mathfrak{L}$, if $\models \phi \to \psi$ then $B^N_\psi \subseteq B^N_\phi$.*

Proof. We will prove that for every i, $B^N_{\psi,i} \subseteq B^N_{\phi,i}$. We prove by induction on i.

If $i = 0$, then $B^N_{\psi,0} = Cn(\psi) \subseteq Cn(\phi) \subseteq B^N_{\phi,0}$. Assume $i = k + 1$ and $B^N_{\psi,k} \subseteq B^N_{\phi,k}$. Then $B^N_{\psi,k+1} = Cn(\{\psi\} \cup N(B^N_{\psi,k}))$. From $B^N_{\psi,k} \subseteq B^N_{\phi,k}$ we deduce $N(B^N_{\psi,k}) \subseteq N(B^N_{\phi,k})$. Now by the monotony of $Cn(\bullet)$ we know $Cn(\{\psi\} \cup N(B^N_{\psi,k})) \subseteq Cn(\{\phi\} \cup N(B^N_{\phi,k}))$. Hence $B^N_{\psi,k+1} \subseteq B^N_{\phi,k+1}$.
So we have proved for every i, $B^N_{\psi,i} \subseteq B^N_{\phi,i}$. With this result in hand, we can easily deduce that $B^N_\psi \subseteq B^N_\phi$. $\dashv$

Lemma 7. *If $\psi \in Cn(N(B^N_\phi))$, then $\psi \in B^N_\phi$.*

Proof. By Lemma 3, it is easy to verify that $N(B^N_\phi) \subseteq B^N_\phi$ and $Cn(B^N_\phi) \subseteq B^N_\phi$. The result then follows.

Lemma 8. *If $\psi \in Cn(N(B^N_\phi))$, then $B^N_\phi = B^N_{\phi\wedge\psi}$.*

Proof. It's easy to prove that $B^N_\phi \subseteq B^N_{\phi\wedge\psi}$. For the other direction, we need to prove that for every i, $B^N_{\phi\wedge\psi,i} \subseteq B^N_\phi$. We prove this by induction on i.

- Base step: Let $i = 0$, we then have $B^N_{\phi\wedge\psi,i} = Cn(\phi \wedge \psi)$. By Lemma 4 we have $\phi \in B^N_\phi$. By Lemma 7 we have $\psi \in B^N_\phi$. Then by Lemma 5 we have $\phi \wedge \psi \in B^N_\phi$.
- Inductive step: Assume for $i = k$, $B^N_{\phi\wedge\psi,k} \subseteq B^N_\phi$. Then $B^N_{\phi\wedge\psi,k+1} = Cn(\{\phi \wedge \psi\} \cup N(B^N_{\phi\wedge\psi,k}))$. From $B^N_{\phi\wedge\psi,k} \subseteq B^N_\phi$ we know there exist j such that $B^N_{\phi\wedge\psi,k} \subseteq \bigcup_{i=0}^j B^N_{\phi,i}$. Therefore $N(B^N_{\phi\wedge\psi,k})) \subseteq N(\bigcup_{i=0}^j B^N_{\phi,i}) \subseteq \bigcup_{i=0}^{j+1} B^N_{\phi,i} \subseteq B^N_\phi$. So we have proved $N(B^N_{\phi\wedge\psi,k})) \subseteq B^N_\phi$. By the base step we have $\phi \wedge \psi \in B^N_\phi$. Then by Lemma 5 we know $Cn(\{\phi \wedge \psi\} \cup N(B^N_{\phi\wedge\psi,k})) \subseteq B^N_\phi$. That is, $B^N_{\phi\wedge\psi,k+1} \subseteq B^N_\phi$.

Lemma 9. *For all i, if $\chi \in B^N_{\phi,i}$ and $(\chi, \psi) \in N$, then $(\phi, \psi) \in D_3(N)$*

Proof. We prove by induction on i.

– Base step: Let $i = 0$. Then $\chi \in B^N_{\phi,0} = Cn(\phi)$. Hence $\models \phi \to \chi$. Therefore we can apply SI to $\models \phi \to \chi$ and (χ, ψ) to derive (ϕ, ψ).
– Inductive step: Assume for $i = k$, if $\chi \in B^N_{\phi,k}$ and $(\chi, \psi) \in N$, then $(\phi, \psi) \in D_3(N)$. Now let $\chi \in B^N_{\phi,k+1}$. Then $\chi \in Cn(\{\phi\} \cup N(B^N_{\phi,k}))$, and there exist $\chi_1 \dots \chi_n \in N(B^N_{\phi,k})$ such that $\phi \wedge \chi_1 \wedge \dots \wedge \chi_n \models \chi$. Then apply SI to $(\chi, \psi) \in N$ and $\phi \wedge \chi_1 \wedge \dots \wedge \chi_n \models \chi$ we have $(\phi \wedge \chi_1 \wedge \dots \wedge \chi_n, x) \in D_3(N)$. Note that for each $i \in \{1, \dots, n\}$, from $\chi_i \in N(B^N_{\phi,k})$ we know there is $\phi_i \in B^N_{\phi,k}$ such that $(\phi_i, \chi_i) \in N$. Now by inductive hypothesis we have $(\phi, \chi_i) \in D_3(N)$. Then applying the AND rule we have $(\phi, \chi_1 \wedge \dots \wedge \chi_n) \in D_3(N)$. From $(\phi, \chi_1 \wedge \dots \wedge \chi_n) \in D_3(N)$ and $(\phi \wedge \chi_1 \wedge \dots \wedge \chi_n, \psi) \in D_3(N)$ we can adopt the CT rule to derive $(\phi, \psi) \in D_3(N)$.

4 Decidability

Concerning the decidability of input/output STIT logic, we study on the following problems:

– Compliance problem: given a finite set of norms N, a finite set of formulas Φ and a formula ψ, is $\psi \in O(N, \Phi)$?
– Violation problem: given a finite set of norms N, a finite set of formulas Φ and a formula ψ, is $\neg\psi \in O(N, \Phi)$?
– Compatibility problem: given a finite set of norms N, a finite set of formulas Φ and a formula ψ, is $\neg\psi \notin O(N, \Phi)$?

Intuitively, the compliance problem asks whether certain proposition complies the normative system. The violation problem asks whether certain proposition violates the normative system and the compatibility problem asks whether the normative system is compatible with certain proposition. Both the violation problem and the compatibility problem can be reduced to the compliance problem, therefore we only study the decidability of the compliance problem.

We prove that all the input/output STIT logic introduced in this paper is decidable by showing that the compliance problem is solvable by oracle Turing machines.

Definition 6 (oracle Turing machine [3]). *An oracle for a language L is a device that is capable of reporting whether any string w is a member of L. An oracle Truing machine M^L is a modified Turing machine that has the additional capability of querying an oracle. Whenever M^L writes a string on a special oracle tape it is informed whether that string is a member of L, in a single computation step.*

4.1 Simple-minded

Theorem 6. *The compliance problem of simple-minded input/output STIT logic is decidable.*

Proof: We provide the following algorithm on an oracle Turing machine with oracle STIT-$SAT = \{\phi \in \mathfrak{L} : \phi \text{ is satisfiable}\}$ to solve the compliance problem of simple-minded input/output STIT logic.

Let $N = \{(\phi_1, \psi_1), \dots, (\phi_n, \psi_n)\}$, Φ be a finite set of formulas and ψ be a formula.

1. for each $\phi_i \in \{\phi_1, \ldots, \phi_n\}$, ask the oracle if $\neg(\bigwedge \Phi \to \phi_i)$ is satisfiable.
   (a) If the oracle answer "no", then mark ψ_i
   (b) Otherwise do nothing.
2. Let $\psi_{i_1}, \ldots \psi_{i_k}$ be all those ψ_i which are marked in step 1.
3. Ask the oracle if $\neg(\psi_{i_1} \wedge \ldots \wedge \psi_{i_k} \to \psi)$ is satisfiable.
   (a) If the oracle answer "no", then return "accept"
   (b) Otherwise return "reject".

It can be verified that $\psi \in Cn(N(Cn(\Phi)))$ iff the algorithm returns "accept". Therefore simple-minded input/output STIT logic is decidable. ⊣

Remark 1. Here the decidability of individual STIT logic is crucial for the decidability of input/output STIT logic. If we choose group STIT, of which the satisfiability problem is undecidable [11], as our base logic, then our input/output STIT logic will be undecidable because the satisfiability problem of the base logic can be reduced to the compliance problem by making $N = \emptyset$.

Corollary 1. *The violation problem and compatibility problem of simple-minded input/output STIT logic is decidable.*

4.2 Basic

Theorem 7. *The compliance problem of basic input/output STIT logic is decidable.*

Proof: We provide the following algorithm on an oracle Turing machine with oracle STIT-SAT to solve the compliance problem.

Let $N = \{(\phi_1, \psi_1), \ldots, (\phi_n, \psi_n)\}$, $\Phi = \{\chi_1, \ldots, \chi_m\}$ be a finite set of formulas and ψ be a formula.

1. Let $B_1, \ldots, B_m$ be the sequence of all minimal disjunctive extension of Φ.
2. Let $i = 1$.
3. Let $\Phi = B_i$.
4. for each $\phi_j \in \{\phi_1, \ldots, \phi_n\}$, ask the oracle if $\neg(\bigwedge \Phi \to \phi_i)$ is satisfiable.
   (a) If the oracle answer "no", then mark ψ_j
   (b) Otherwise do nothing.
5. Let $\psi_{j_1}, \ldots \psi_{j_k}$ be all those ψ_j which are marked in step 4.
6. Ask the oracle if $\neg(\psi_{j_1} \wedge \ldots \wedge \psi_{j_k} \to \psi)$ is satisfiable.
   (a) If the oracle answer "no", then let $i = i + 1$.
      i. if $i \leq m$, then goto step 3.
      ii. if $i = m + 1$, then return "accept"
   (b) Otherwise return "reject".

It can be verified that $\psi \in O_2(N, \Phi)$ iff the algorithm returns "accept". Therefore simple-minded input/output STIT logic is decidable. ⊣

Corollary 2. *The violation problem and compatibility problem of basic input/output STIT logic is decidable.*

4.3 Simple-minded reusable

Theorem 8. *The compliance problem of simple-minded reusable input/output STIT logic is decidable.*

Proof: We provide the following algorithm on an oracle Turing machine with oracle STIT-SAT to solve the compliance problem of simple-minded reusable input/output STIT logic. The case for simple-minded input/output STIT logic is easier and left to the readers.

Let $N = \{(\phi_1, \psi_1), \ldots, (\phi_n, \psi_n)\}$, Φ be a finite set of formulas and ψ be a formula.

1. Let $X = \Phi, Y = Z = N, U = \emptyset$.
2. for each $(\phi_i, \psi_i) \in Y$, ask the oracle if $\neg(\bigwedge X \to \phi_i)$ is satisfiable
   (a) if "no", then let $X = X \cup \{\psi_i\}$, $Z = Z - \{(\phi_i, \psi_i)\}$.
   (b) Otherwise do nothing.
3. If Y equals to Z, goto 4. Otherwise let $Y = Z$, goto step 2
4. for each $(\phi_i, \psi_i) \in N$, ask the oracle if $\neg(\bigwedge X \to \phi_i)$ is satisfiable
   (a) If "no", then let $U = U \cup \{\psi_i\}$.
   (b) Otherwise do nothing
5. Ask the oracle if $\neg(\bigwedge U \to \psi)$ is satisfiable.
   (a) If "no", then return "accept".
   (b) Otherwise return "reject".

The correctness of the above algorithm is routine to be proven and we left it to the readers. Therefore simple-minded reusable input/output STIT logic is decidable. ⊣

Corollary 3. *The violation problem and compatibility problem of simple-minded reusable input/output STIT logic is decidable.*

5 On Ross' paradox

Ross' paradox [18] originate from the logic of imperatives, and is a well-known puzzle in deontic logic. Ross' paradox says that the inference rule WO cannot be valid, since if it were, then from

(1) You ought to post the letter

we could conclude that

(2) You ought to post the letter or burn it

and we obviously cannot.

Both Makinson and van der Torre's input/output logic and deontic STIT logic [12, 13, 19] are not free from this paradox.

Ross' paradox relies on the rule $Ought(\phi) \to Ought(\phi \vee \psi)$ of deontic logic. In our input/output STIT logic, we choose deliberative STIT as our base logic. Therefore we don't have $\models [i^d]\phi \to [i^d](\phi \vee \psi)$ because it might be $\models \Box(\phi \vee \psi)$. Therefore $(\top, [i^d](\phi \vee \psi))$ is not derivable from $(\top, [i^d]\phi)$, which means Ross' paradox is solved.

6 Conclusion

In this paper we study input/output STIT logic. We introduce the semantics, proof theory and prove the completeness theorem. Input/output STIT logic has stronger expressive power than Mankinson and van der Torre's input/output logic. We show that input/output STIT logic is decidable and free from Ross' paradox.

Directions of future work are manifold. Two natural directions includes: (1) What is the semantics for basic reusable input/output STIT logic? (2) What is the complexity of input/output STIT logic?

References

1. Thomas Ågotnes, Wiebe van der Hoek, Juan A. Rodríguez-Aguilar, Carles Sierra, and Michael Wooldridge. On the logic of normative systems. In Manuela M. Veloso, editor, *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 1175–1180, 2007.
2. Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leendert W. N. van der Torre, editors. *Normative Multi-Agent Systems*, volume 4 of *Dagstuhl Follow-Ups*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2013.
3. Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, USA, 2009.
4. Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.
5. Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the future: agents anc choice in our inderterminist world*. Oxford, 2001.
6. Guido Boella and Leendert van der Torre. A logical architecture of a normative system. In *Deontic Logic and Artificial Normative Systems*, pages 24–35. Springer, 2006.
7. Guido Boella, Leendert van der Torre, and Harko Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1):1–10, 2008.
8. Silvano Colombo Tosatto, Guido Boella, Leendert van der Torre, and Serena Villata. Abstract normative systems: Semantics and proof theory. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning*, pages 358–368, 2012.
9. Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leendert van der Torre, editors. *Handbook of Deontic Logic and Normative Systems*. College Publications, 2014.
10. Andreas Herzig, Emiliano Lorini, Frédéric Moisan, and Nicolas Troquard. A dynamic logic of normative systems. In Toby Walsh, editor, *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 228–233. IJCAI/AAAI, 2011.
11. Andreas Herzig and François Schwarzentruber. Properties of logics of individual and group agency. In Carlos Areces and Robert Goldblatt, editors, *Advances in Modal Logic*, pages 133–149. College Publications, 2008.
12. John Horty. *Agency and Deontic Logic*. Oxford University Press, New York, 2001.
13. Barteld Kooi and Allard Tamminga. Moral conflicts between groups of agents. *Journal of Philosophical Logic*, 37:1–21, 2008.
14. David Makinson and Leendert van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

15. David Makinson and Leendert van der Torre. Constraints for input/output logics. *Journal of Philosophical Logic*, 30(2):155–185, 2001.
16. David Makinson and Leendert van der Torre. Permission from an input/output perspective. *Journal of Philosophical Logic*, 32:391–416, 2003.
17. Xavier Parent and Leendert van der Torre. I/O logic. In J. Horty, D. Gabbay, X. Parent, R. van der Meyden, and L. van der Torre, editors, *Handbook of Deontic Logic and Normative Systems*. College Publications, 2014.
18. Alf Ross. Imperatives and logic. *Theoria*, 7(5371), 1941.
19. Xin Sun. Conditional ought, a game theoretical perspective. In J. Lang H. van Ditmarsch and S. Ju, editors, *Logic, Rationality, and Interaction: Proceedings of the Thire International Workshop*, pages 356–369, Guangzhou, China, 2011.
20. Xin Sun. How to build input/output logic. In Nils Bulling, Leendert W. N. van der Torre, Serena Villata, Wojtek Jamroga, and Wamberto Vasconcelos, editors, *Computational Logic in Multi-Agent Systems - 15th International Workshop, CLIMA XV, Prague, Czech Republic, August 18-19, 2014. Proceedings*, volume 8624 of *Lecture Notes in Computer Science*, pages 123–137. Springer, 2014.
21. Xin Sun and Leendert van der Torre. Combining constitutive and regulative norms in input/output logic. In Fabrizio Cariani, Davide Grossi, Joke Meheus, and Xavier Parent, editors, *DEON*, volume 8554 of *Lecture Notes in Computer Science*, pages 241–257. Springer, 2014.
22. Michael J. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2009.