Stit Logics, Games, Knowledge, and Freedom

Roberto Ciuni and John Horty

Abstract This paper has two main goals: highlighting the connections between Stit logics and game theory and comparing Stit logics with Matrix Game Logic, a Dynamic Logic introduced by van Benthem order to model some interesting epistemic notions from game theory. Achieving the first goal will prove the flexibility of Stit logics and their applicability in the logical foundations of game theory, and will lay the groundwork for accomplishing the second. A comparison between Stit logics and Matrix Game Logic is already offered in a recent work by van Benthem and Pacuit. Here, we push the comparison further by embedding Matrix Game Logic into a fragment of group Stit logic, and using the embedding to derive some properties of Matrix Game Logic—in particular, undecidability and the lack of finite axiomatizability. In addition, the embedding sheds light on some open issues about the so-called "freedom operator" of Matrix Game Logic.

1 Introduction

Johan van Benthem's career has been about research—often ground breaking, transformative research—but also, and especially in recent years, about building bridges and establishing conversations: across disciplines, between research communities, and among researchers from different nations and cultures. At a stage when so many others of his stature would be content with focusing inward, solidifying results, and protecting their territory, Johan has been looking in fresh directions, breaking down barriers, and seeking to involve others in a common enterprise. In twenty years time,

Roberto Ciuni
Ruhr University Bochum, Department of Philosophy II, Universitätsstraβe 150, 44780, Bochum, Germany, e-mail: ciuniroberto@yahoo.it

John Horty
Philosophy Department, University of Maryland, College Park, MD 20742, USA, e-mail: horty@umiacs.umd.edu

logic will be a stronger, more integrated discipline because of his ambassadorial work; in fifty years, it will be stronger still.

This chapter contributes to one of the many conversations that Johan has begun—in this case, between those working in the tradition of Stit Logics, and those working with the different picture of agency underlying Dynamic Logic and Dynamic Epistemic Logic, a theory to which Johan himself has made seminal contributions.

Stit logics—which we here characterize, collectively, as STIT—were originated by Nuel Belnap and his many collaborators in a series of papers culminating with the monograph [4]; the framework was then connected to issues in decision theory, deontic logic, and cooperative game theory in [24]. STIT takes its name from the phrase "seeing to it that," which the theory interprets as a modal operator, known as the "stit operator," capturing the idea that an agent i sees to it that ϕ just in case ϕ is true in all states, or courses of events, compatible with a particular choice made by i. The main semantic ingredient of the theory is, accordingly, the notion of the *choices* available to an agent, which STIT characterizes—in a purely extensional way—as sets of states, or courses of events. Acting is then interpreted as selecting some such sets and excluding others. Beside this, STIT has two eye-catching features: it does not include labels for actions, which in turn find no expression in the language, and it assumes an independence condition according to which any choice of any agent is compatible with any choice of any other agent.

STIT has its roots in the field of formal philosophy, and has been applied to clarify some crucial conceptual issues in the theory of action and ethics— for example, connections between moral responsibility and the principle of "could have done otherwise" [4], the rigorous formulation of criteria for consequentialist theories of action [24], the judicial notion of *mens rea* [11], and the attribution of individual responsibility in cases of group agency [16]. However, in the twenty some years since its introduction,[1] the applications of STIT have slowly shifted to theoretical computer science and related areas, particularly the logical foundations of multi-agent systems, artificial intelligence, and game theory.

This shift has opened interesting issues. STIT and game theory talk different jargons and have been directed toward different targets. Though there are game-theoretical roots in STIT, a clear display of the connections between STIT and games has not been undertaken.[2] Also, the arena of formalisms for the logical foundations of games and multi-agent systems is very rich, and an analysis of the relations with prominent formalisms in this family constitutes a fascinating area of applications for STIT.

In the present chapter, we touch on both of these issues. First, we try to clarify potential of STIT as a logical foundation for game theory by describing its adequacy for modeling certain game-theoretic notions. And second, we compare STIT with Matrix Game Logic, Dynamic Logic first introduced by van Benthem in [5] and

[1] Belnap and Perloff [3] and von Kutschera [32] are usually regarded as the two papers that, independently, lay the foundations of STIT.

[2] See, however, the important earlier work by Kooi and Tamminga [25], Tamminga [27], and Turrini [28].

developed in his later [6] and [7]; in this comparitive project, we continue, and hope to advance, the conversation initiated by van Benthem and Pacuit [8].[3]

The chapter proceeds as follows. Section 2 introduces structures and semantics for a particular stit logic together with a Hilbert style axiomatization, and reviews some interesting validities and formal properties. Section 3 focuses on a comparison between STIT and strategic games, and tries to fill the gap between these two areas; some interesting points of comparison are the possibility of reading STIT game-theoretically, and of isolating a "STIT component" within games. Sections 4 and 5 compare STIT with Matrix Game Logic. Section 5 contains the most novel result of the paper: a mutual embedding between Matrix Game Logic and a particular STIT for group agency, with a consequent property transfer. Section 6 presents some conclusions.

2 STIT

STIT logics and stit operators abound, and the choice among them largely depends on one's purposes. We will rely here on the so-called "Chellas stit"[4] and we interpret our logic on *Choice Kripke frames*, with no temporal ordering on states of evaluation: more complex operators and the temporal dimension of agency are not needed for the comparisons we draw in later sections of this paper.[5]

Choice Kripke Frames. Formally, a *Choice Kripke frame*—a CKF, for short—is a triple $\langle W, Ags, \{\sim_i^C \mid i \in Ags\}\rangle$ where:

- W is a non-empty set $\{w, w', w'' \dots\}$ of *states*
- Ags is a finite, non-empty set $\{1, \dots, n\}$ of individual *agents*
- For each agent $i \in Ags$, the relation $\sim_i^C \subseteq W \times W$ is a *choice-equivalence relation* satisfying the following conditions:

(R1) $\sim_i^C$ is an equivalence relation

[3] Others have also tried to developed unified perspectives encompassing STIT and dymanic logics. See for instance Herzig and Lorini [20], which presents a dynamic logic of agency in the tradition of Propositional Dynamic Logics. In this framework, a basic stit operator can be defined as an existential quantifier over the actions of a given agent.

[4] This particular operator was first isolated, and given this name, by Horty and Belnap [23]; the name reflects the fact that the operator captures, in the different framework of stit semantics, ideas introduced much earlier by Chellas [14]. A comparison between Chellas's early work on agency and the later STIT can be found in Chellas [15], and also in [23].

[5] STIT is traditionally interpreted on branching-time structures (see [4] and [24]) where moments are linearly ordered toward the past but are not linearly ordered toward the future. In such structures, *choices* are sets of histories, and *histories* are in turn sets of *moments* which are (1) maximal with respect to inclusion and linearly ordered toward the future. However, the most widespread stit operators do not express any temporal dimension, and thus the indeterministic framework can be replaced by Kripke frames where no temporal order is imposed. Such frames are used by Balbiani, Herzig, and Troquard [1], and by Herzig and Schwarzentruber [21]; and we follow them in the present paper.

(R2) $W \times W \ \subseteq \sim_1^C \circ \sim_2^C \cdots \circ \sim_n^C$

If $w \sim_i^C w'$, we say that w is choice-equivalent with w' for agent i, and we let $[w]_i$ be the equivalence class including the states which are choice-equivalent with w—that is, $[w]_i = \{w' \mid w \sim_i^C w'\}$. The condition R2—known as *strong confluence*—is equivalent to the requirement that $[w]_1 \ \cap \ [w']_2 \ \cap \cdots \cap \ [w'']_n \ \neq \ \emptyset$; in other words, this condition guarantees the *independence condition* mentioned earlier, that all the choices of all the different agents are compatible. This is a very demanding condition, but it turns to be a key element for the game-theoretical reading of STIT, which in turn forms the bridge between STIT and Matrix Game Logic; we return to this issue in subsection 3.2.

We define a *restricted Choice Kripke frame*—a CKF$^+$, for short—as a CKF satisfying the further conditions:

(R3) For every state w, $[w]_1 \cap \cdots \cap [w]_n = \{w\}$
(R4) For every state w, $[w]_{\bar{i}} \ = \ \bigcap_{j \in \bar{i}} [w]_j$, where $\bar{i} = Ags/i$

R3 states that the combination of the choices of all the agents at w consists in w itself: the combined choices of all the agents are enough to determine a unique state of the world.[6] We refer to $\bar{i}$ as the *anti-group* of i—the group including all agents except i. R4 then states that each choice of i's anti-group at w is equal to the intersection of the choices of its members at w.[7] In what follows, we will confine ourselves to CKF$^+$'s, which will make the comparison with strategic games easier.[8] To illustrate, Figure 1 exemplifies a CKF$^+$:

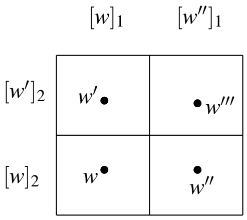


Figure 1: a CKF$^+$ with two agents, 1 and 2 and two choices per agent.

[6] In case this condition seems to strong, it is helpful to think of one the agents as "nature," which removes any remaining indeterminacy once all the more ordinary agents have made their choices; this tactic was mentioned in [24, p. 91].

[7] R4 is just an instance of the game-theoretical principle of *additivity*, which characterizes the construction of all groups in group STIT; see section 5 below.

[8] Actually, the correspondence between games and consequentialist CKF$^+$ (see below for a definition) can also be established without imposing condition R3; see, for example, van Benthem and Pacuit [8] and Tamminga [27]. However, the condition makes the proof of such a correspondence much more straightforward and general. In addition the proofs which do not use R3 essentially rely on consideration about language, while the correspondence result which uses R3, established by Turrini in [28], relies only on the structures in question.

The two columns represent the choices $[w']_1$ and $[w''']_1$ of agent 1, while the rows represent the choices $[w']_2$ and $[w]_2$ represent the choices of agent 2. A moment's consideration is enough to see that R1–R4 are satisfied by the structure represented by the figure.

Language and Semantics. In addition to the set *Ags* of agents, assume a denumerable set of atomic formulas. Our language $\mathscr{L}_{\mathsf{CSTIT}}$ then has the Backus-Naur form

$$\phi ::= p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid [i]\phi,$$

where p is atomic, where ϕ_1 and ϕ_2 are arbitrary formulas, and where $i \in Ags$; the other Boolean connectives are defined as usual on the basis of $\neg$ and $\wedge$. The symbol $[i]$ is the Chellas stit operator, so that $[i]\phi$ should be taken to mean that the agent i sees to it that ϕ—that is, that the truth of ϕ is guaranteed by a choice due to i. Its dual is the symbol $\langle i \rangle$, so that $\langle i \rangle\phi$ should be understood as meaning that ϕ is consistent with the choice exerted by i.

The formulas in $\mathscr{L}_{\mathsf{CSTIT}}$ are evaluated on Choice Kripke Models—CKM$^+$, for short—where a CKM$^+$ $\mathscr{M}$ is a pair $\langle \mathscr{K}, V \rangle$, with $\mathscr{K}$ a CKF$^+$ and V a function from atomic formulas into sets of states at which they are true. The satisfaction relation $\models$ for formulas in $\mathscr{L}_{\mathsf{CSTIT}}$ can then be defined as follows:

(TC1) $\mathscr{M}, w \models p$ iff $w \in V(p)$
(TC2) $\mathscr{M}, w \models \neg\phi$ iff $\mathscr{M}, w \not\models \phi$
(TC3) $\mathscr{M}, w \models \phi \vee \psi$ iff $\mathscr{M}, w \models \phi$ or $\mathscr{M}, w \models \psi$
(TC4) $\mathscr{M}, w \models [i]\phi$ iff for all w', if $w' \in [w]_i$ then $\mathscr{M}, w' \models \phi$

Truth in a CKM$^+$, in a CKF$^+$, and in all CKF$^+$'s is defined by the usual universal quantifications; and as usual, we will take $\mid \phi \mid^{\mathscr{M}} = \{w \mid \mathscr{M}, w \models \phi\}$ as the set of states from the model $\mathscr{M}$ satisfying ϕ. Figure 2 represents a CKM$^+$ built from the CKF$^+$ depicted in Figure 1. It is easy to see that $[1]\phi$ is true at w' and w but false at w''' and w'', while $[2]\phi$ is true at w and w'' but false at w' and w'''.

Figure 2: a CKM based on Figure 1

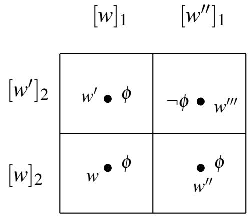
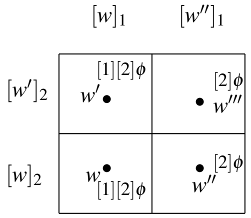


Figure 3: $[1][2]\phi \rightarrow \Box\phi$



An interesting feature of the semantics for $[i]$ is that the necessity operator $\Box$ can be defined as any combination of distinct agency operators, such as $[j][i]$. This follows at once from R2—strong confluence—and the usual reading of $\Box$ in terms of a *total* equivalence relation on W. Figure 3 provides an intuitive illustration. Here, $[1][2]\phi$ holds at w, from which it follows by TC4 that $[2]\phi$ holds at both w and w', and then, once again by TC4, that ϕ holds both at w and w'' and at w' and w'''. The fact that $[1][2]\phi$ holds somewhere thus implies that ϕ is true at all states in the model. In what follows, we will often use $\Box\phi$ as an abbreviation for $[j][i]\phi$, or for any other combination of stit operators for different agents.

2.1 Axiomatics and Interesting Validities

Axiomatics. The axiomatization of CSTIT consisting of

(S5) S5's axioms for $\Box$ and each $[i]$
(MA) $\Box\phi \ \rightarrow\ [i]\phi$
(IA) $\bigwedge_{i\in Ags}\Diamond[i]\phi_i \ \rightarrow\ \Diamond\bigwedge_{i\in Ags}[i]\phi_i$

together with Modus Ponens and the Rule of Necessitation for $\Box$ and every stit operator $[i]$ is sound and complete relative to CKF$^+$.[9] The S5 properties follow from that fact the fact that $[w]_i$ is an equivalence class, for every agent i; this tells us also that CSTIT is a multi-modal S5—a very nice one, in fact, as we shall see below. Given our definition of $\Box$, the mixing axiom MA is short for $[j][i]\phi \ \rightarrow\ [i]\phi$; thus it is an instance of axiom T, included in S5, but we display it separately for convenience. IA is the well-known axiom of independence of agents, which states that any combination of independently possible actions, one for each agent, is jointly possible. This is, as we mentioned earlier , a very strong principle and a key feature of STIT; we discuss it further in subsection 3.2, after highlighting the connections between STIT and game theory. [[IA does not correspond to strong confluence, but rather to a connected property: for every agent i and choice $[w]_i$, there is at least one state w^* such that for all the agents $j \in \bar{i}$, there is one choice $[w']_j$ such that $w^* \in [w]_i \cap [w]_j$.

Interesting Valid Formulas. In fact, the property of strong confluence corresponds to the principle of *Triviality of Coercion*,

(TC) $[i][j]\phi \rightarrow \Box\phi$ (for $i \neq j$),

which is justified by the definition of $\Box$ as $[i][j]$, and which states that one agent can guarantee that another agent guarantees a certain proposition only if that proposition is itself trivial. It was shown by Balbiani, Herzig, and Troquard [1] that the two-agent version of IA ($\Diamond[i]\phi \wedge \Diamond[j]\phi \ \rightarrow \Diamond([i]\phi \wedge [j]\phi)$) can be derived by TC and the definition of $\Box$ as $[j][i]$, while the n-agent version can be proved by induction on the $n-1$ case and an additional principle $\Diamond\phi \ \rightarrow\ \langle k\rangle \bigwedge_{1\leq i<k}\langle i\rangle\phi$, for $k > 1$. Also worth mentioning is the *Permutation Principle*,

[9] See Balbiani, Herzig, and Troquard [1] for discussion. This axiomatization is due to Xu [30], where, however, the Chellas stit was replaced by the deliberative stit, and completeness is proved relative to trees endowed with choices and agents.

(PP) $[i][j]\phi \leftrightarrow [j][i]\phi$

which follows by TC and the definition of $\Box$.

Formal Properties of STIT and group CSTIT. CSTIT has very convenient formal properties, some of which are surprising in a multi-modal S5. First, the axiomatization above is complete with respect to CKF. Since R3 is not expressible in $\mathscr{L}_{\mathsf{CSTIT}}$, the system does not distinguish the frames satisfying the condition, and thus it proves complete also relative to CKF$^+$.

Also, CSTIT is decidable and finitely axiomatizable.[10] This is trivial if we confine our attention to a two-agent CSTIT; in this case, the STIT property of strong confluence reduces to the confluence property of S5^2 ("squared S5"), which is decidable and finitely axiomatizable.[11] The interesting virtue of STIT is that it keeps these properties even when more than two operators are at stake. The key issue is indeed strong confluence, which in turn reduces *all* the confluences of arbitrary length to the confluence $\sim_{[i]} \circ \sim_{[j]}$, with $[i]$ and $[j]$ arbitrary. This has two interesting consequences: (1) It implies that CSTIT with $n > 2$ agents does not yield the product logic S5^n, so that, for example, CSTIT with 3 agents is not S5^3 ("cube S5"); the implicit virtue here is clear, since any product S5^n with $n > 2$ is undecidable and not finitely axiomatizable.[12] (2) It implies that, in spite of the independence of agents, CSTIT does not encode grid structure enough to become undecidable.

STIT's view on agency and the many fashions of STIT. Our presentation has emphasized the distinctive marks of STIT mentioned in section 1—the extensional view on choices, acting as state selection, the absence of action types in PDL or DEL style, and the independence of agents. CSTIT is not the only STIT logic available, but most STIT logics share these features.[13]

None of this tells us exactly how the stit operator is to be read. Here we just briefly mention two prominent readings. The first is the original reading due to Belnap and Perloff, and is today the most widespread among philosophers; according to this reading, "seeing to it that" captures the contribution of the agents to a change in the causal structure of the world.[14] The second points at a game-theoretical interpretation of STIT, and is suggested by [24] and other research in the computer science side of STIT; here the idea is to focus on what an agent guarantees, or "sees to," by following a winning strategy.

Comparing such readings here would go beyond the scope of the chapter; we note only that the former focus on agency as a factor of change in spatial regions of

[10] These results were first established by Xu [29]; see also Balbiani, Herzig, and Troquard [1].

[11] Balbiani, Herzig, and Troquard [1, p. 395] prove that the logic of two-agent Chellas stit is nothing but S5^2.

[12] See Hirsch, Hodkinson, and Kurucz [19].

[13] A noticeable exception is the combinations of STIT and actions explored by Xu [31].

[14] If we follow Belnap's reading, deliberative stit may prove more suitable than the Chellas stit, since the former does not allow for trivial truths to be seen to by any agent; the Chellas stit allows for this and does not seem to fit equally well the idea of a causal contribution to a change in the world.

the universe, while the second stresses the choice-making dimension of agency. The two readings thus suggest different applications and questions: Belnap's reading naturally calls for a specification of what "causal contribution" to a change is, and consequently requires also a picture of what the causal structure of the world is; by contrast, the game-theoretical reading highlights the logical dynamics of choice-making and multi-agent interaction, and can be adapted to the different structures provided by game theory, such as, for example, extended game forms or strategic games.

An interesting point of Belnap's reading is that it naturally calls for representing agency in time, since causation presupposes a temporal dimension.[15] The CKF's are not enough to this purpose, and thus it is no surprise that Belnap and colleagues use *branching-time* structures endowed with a set of agents and a choice function (the so-called "BT + AC structures"), [16] which are indeed intended to capture the notion of an indeterministic causal change in the world.[17]

Finally, STIT traditionally does not define choices in relational terms, but rather in functional ones: the standard option is to introduce a choice function $Ch(i)$: $Ags \longmapsto \wp(W)$ which partitions W and such that $Ch(i,w) \; \cap \; Ch(j,w') \; \neq \; \emptyset$ for $j \neq i$. It is the easy to see that our relation $\sim_i^C$ can be defined in terms of $Ch(i)$ and vice versa. We use this interchangeability in section 3, and we also impose $Ch(\bar{i},w) = \bigcap_{j \in \bar{i}} Ch(j,w)$ for every $w \in W$ and $\bigcap_{i \in Ags} Ch(i) = \mid W \mid$, in analogy with R3 and R4.

3 STIT and Strategic Games

Although STIT was first presented as a theory of the contribution of agents to changes in the causal structure of the world, it was soon applied to problems concerning choice-making, and much of the current research is in keeping this direction; this applies especially the application of STIT in the logical foundations of multi-agent systems. The reason for such a shift is that STIT presents some inter-

[15] Such a component proves relevant also if we wish to represent the sequential aspect of choice-making in extended game forms, since a sequence of choice-making acts presuppose a temporal dimension.

[16] To be more precise, the temporal component of BT + AC structures are *trees*. Along the years, other temporal components of indeterministic time for choices and agents have been introduced; see, for instance, the XSTIT *frames* due to Broersen [11], the *bundled choice trees* of Ciuni and Zanardo [18], and the *Kripke STIT frames* of Lorini [26].

[17] Display of a temporal order is necessary to define the so-called 'stit operators for *non-instantaneous* agency', that is operators that express a temporal hiatus between choice and result. Examples of such operators are the fused xstit in Broersen [11]; similar operators are introduced in Ciuni and Mastop [16] and Ciuni and Zanardo [18]. A hiatus between choices and results can be also expressed by *combining* autonomous operators for agency and for temporal distinctions. Broersen follows this line in [9] and [10], as does Lorini [26]. A very complex stit operator is the original "achievement stit" due to Belnap and Perloff [3] which captures the cross-temporal dimension of agency by expressing the notion that a result holds at *m* due to a previous choice of *i*; variants of the achievement stit are proposed by Ciuni [17] and Zanardo [33].

esting connections with game theory, which is the most widespread framework of reference in studies on multi-agent systems.[18] Here, we consider some of these connections, focusing on *strategic games*. This is an indispensable move for the formal comparison between STIT and Matrix Game Logic in section 4, and also helps clarify some features of STIT which have been debated, such as the independence of agents.

3.1 Bridging Two Worlds

A strategic game $\mathscr{G}$ is a five-tuple $\langle W, Ags, \{a_i \mid i \in Ags\}, o, \{\succeq_i \mid i \in Ags\}\rangle$, where W and Ags are as in CKF's; for each $i \in Ags$, a_i is an action available to agent i. We call A_i the set of actions $\{a_i, a'_i, \dots\}$ available to i.[19] Each sequence $\{a_1, \dots, a_n\} \in \Pi_{i \in Ags} A_i$ is a *action profile*; such an action profile can also be denoted as $(a_i, a_{\bar{i}})$, which separates the action performed by the particular agent i from the actions performed by all the other agents from $\bar{i}$.

The function $o : \Pi_{i \in Ags} A_i \longmapsto W$ maps each action profile into a resulting state in from W, according to the standards in the definition of a strategic game. Notice that each set in $\Pi_{i \in Ags} A_i$ represents a possible combination of actions in the game; as a consequence, all actions of any agent i are compatible with all the actions of any other agent j from $\bar{i}$; also, o is total, and thus this compatibility is represented at the level of the outcomes. We generalize the signature of o in two different ways. First, we let it take two arguments, so that $o(a_i, a_{\bar{i}})$ defines the outcome as resulting from a pair consisting of the action of the particular agent i, in the first place, taken together with the actions of all the other agents, from $\bar{i}$, in the second. Second, we denote the outcomes of a_i as $o(a_i)$, where $o(a_i) = \{w \mid w = o(a_i, a_{\bar{i}})$ for some $a_{\bar{i}} \in A_{\bar{i}}\}$. It is then clear that while the outcomes of action profiles are single states, the outcomes of the actions of individual agents are sets of states.

Finally, for each $i \in Ags$, the relation $\succeq_i$ is a reflexive preference ordering between outcomes of action profiles. The reading of $o(a_i, a_{\bar{i}}) \succeq_i o(a'_i, a_{\bar{i}})$ is the standard one: agent i weakly prefers the state resulting from action profile $(a_i, a_{\bar{i}})$ to that resulting from action profile $o(a'_i, a_{\bar{i}})$. The preference relation is easily extended to (outcomes of) actions of a given agent.

In order to draw our comparisons, we first extend CKF$^+$'s with preference relations $\{\succeq_i \mid i \in Ags\}$, thus obtaining *consequentialist* CKF$^+$'s, or CCKF$^+$'s for short. More exactly, a CCKF$^+$ $\mathscr{C}$ is a pair $\langle \mathscr{K}, \{\succeq_i \mid i \in Ags\}\rangle$, where $\mathscr{K}$ is a CKF$^+$. A model built from a CCKF$^+$—that is, a consequentialist CKM$^+$, or a CCKM$^+$, for

[18] One should not forget that game-theoretical ideas were very important in STIT since its very beginning. This is clear from [4, pp. 283, 343–344], where the matrix representation of games is mentioned and independence of agents is explained with it, and where a comparison between extended game forms and BT + AC is briefly drawn.

[19] Since we do not deal with the sequential aspect of choice-making here, we prefer to use the term 'action' rather than 'strategy'.

short—is obtained by supplementing the CCKF$^+$ with an evaluation function V in the standard way.

Turrini [28] has proved an interesting correspondence between strategic games and CCKF$^+$'s in their functional version (see end of section 2 above). Relying of the definition of the choice function $Ch(i)$, he first introduces the notion of *a choice structure $Ch^{\mathscr{G}}$ in a game $\mathscr{G}$* as follows: $X \in Ch^{\mathscr{G}}(i)$ if and only if there is an action a_i such that $\{o(a_i, a_{\bar{i}}) \mid a_{\bar{i}} \in \Pi_{j \in \bar{i}} A_{\bar{j}}\} = X$. He then proves:

Proposition 1 (Representation Theorem). For every (functional) CCKF$^+$ $\mathscr{C} = \langle W, Ags, Ch, \{\succeq_i \mid i \in Ags\}\rangle$, there is a strategic game $\mathscr{G}$ such that $Ch(i) = Ch^{\mathscr{G}}(i)$, and vice versa.[20]

The result can be adapted to our *relational* CCKF$^+$'s with no risk of loss, due to the definition of the function of choice at the end of section 2.[21]

Turrini's observation has three interesting consequences. First, it implies that for every CCKF$^+$ $\mathscr{C} = \langle W, Ags, \{\sim_i^C \mid i \in Ags\}, \{\succeq_i \mid i \in Ags\}\rangle$ we can construct the corresponding *choice structure* of a game (call it $\mathscr{G}^{\mathscr{C}}$ for short); the converse also holds: the CCKF$^+$ $\mathscr{C}^{\mathscr{G}^{\mathscr{C}}}$ built on the choice structure $\mathscr{G}^{\mathscr{C}}$ is in turn nothing but $\mathscr{C}$. There is then a correspondence between CCKF$^+$'s and choice structures of games. Second, the choices of i in a CCKF$^+$ $\mathscr{C}$ are actually the outcomes of some action of i in the game with the corresponding choice structure $\mathscr{G}^{\mathscr{C}}$, or more exactly:

> For every CCKF$^+$ $\mathscr{C}$, $w \sim_i^C w'$ iff $w, w' \in o(a_i)$ for some action $a_i \in A_i$ in the strategic game whose choice structure corresponds to CKF$^+$.

We can express this also by saying that $Ch(i) = Ch^{\mathscr{G}}(i) = \{o(a_i) \mid a_i \in A_i\}$ in $\mathscr{G}^{\mathscr{C}}$. By R3 and R4, this allows to express (outcomes of) action profiles as intersections $[w]_i \cap [w']_{\bar{i}}$ of a choice of i and one of her anti-group: $w = o(a_i, a_{\bar{i}})$ iff $w = [w]_i \cap [w']_{\bar{i}}$ (for some $w' \in W$ and every CCKF$^+$ $\mathscr{C}$). Finally, proposition 1 also guarantees that CCKF$^+$ can represent a number of game-theoretical notions; the most paradigmatic examples is that of *weak dominance*:

Definition 1 (Weak Dominance in a CCKF$^+$).
$[w]_i \geq_i [w']_i$ iff $[w]_i \cap [w'']_{\bar{i}} \succeq_i [w']_i \cap [w'']_{\bar{i}}$ for each $w'' \in W$

This idea, which corresponds to the standard game-theoretical definition of weak dominance,[22] was first introduced into STIT by [24], and has been the main focus of consequentialist work in the STIT tradition.[23] However, CCKF$^+$'s can also model other interesting notions of action preference, such as:

[20] This result, established as Theorem 1 in [28], is actually stated there for *full* groups of agents ("coalitions" in the standard game-theoretical terminology) and their anti-groups. Notice that the result in [28] naturally extends to CKF$^+$ without preference relation and strategic game forms, which obtain from games by dropping the preference relation.

[21] Thus, from every game $\mathscr{G}$ we can construct the corresponding CCKF$^+$ $\mathscr{C}^{\mathscr{G}} = \langle W^{\mathscr{G}}, Ags^{\mathscr{G}}, \{\sim_i^{C\mathscr{G}} \mid i \in Ags\}, \{\succeq_i^{C\mathscr{G}} \mid i \in Ags\}\rangle$, where $w \sim_i^{C\mathscr{G}} w'$ iff $w' \in [w']_i$ for $[w]_i \in Ch^{\mathscr{G}}$.

[22] $o(a_i)$ is a *weakly dominant action* iff $o(a_i, a_{\bar{i}}) \succeq_i o(a'_i, a_{\bar{i}})$ for all $a'_i \in A_i$ and all $a_{\bar{i}} \in A_{\bar{i}}$.

[23] See, for example, Kooi and Tamminga [25], Turrini [28], and Tamminga [27].

Definition 2 (Best Choices in a CCKF$^+$).
$[w]_i$ is a best choice for i given $[w']_{\bar{i}}$ iff $[w]_i \cap [w']_{\bar{i}} \succeq_i [w'']_i \cap [w']_{\bar{i}}$ for each $w'' \in W$.

which displays a clear correspondence with the game-theoretical notion of a *best action*.[24]

3.2 Some Conceptual Insights

Proposition 1 and the related facts are revealing in many different respects. Though philosophers know STIT mainly under the causative reading, the theory may be naturally used for modeling game-theoretical notions: its strong ties with game theory allow us to trade notions defined in STIT with notions defined in games, and vice versa. In a nutshell, we can give STIT a game-theoretical reading without loosing any relevant feature of the framework, leading to some interesting consequences.

Game-theoretical Reading of stit. First, if we trade the notion of 'choice' with that of 'outcome of some action', it is clear that 'seeing to it that' equates with 'displaying a winning action'. For take a CCKF$^+$ $\mathscr{C}^{\mathscr{G}^{\mathscr{C}}}$ built on a game structure $\mathscr{G}^{\mathscr{C}}$. Due to proposition 1, for every choice $[w]_i$ defined in $\mathscr{C}^{\mathscr{G}^{\mathscr{C}}}$, there is some action $a_i \in A_i$ such that for every $w' \in W$, $w' \in o(a_i)$ iff $w' \in [w]_i$. Thus, $[i]\phi$ is true at w iff ϕ is true in all the states w' which are in the same outcome $o(a_i)$ as w. But any such states will also be in (the outcome of) some action of the rest of the agents; as a consequence, the fact that $[i]\phi$holds at w implies that, for every $a_{\bar{i}} \in A_{\bar{i}}$, there is some state $w'' \in o(a_{\bar{i}})$ where ϕ holds. As a consequence, $\bar{i}$ cannot see to it that $\neg\phi$. In other words, if i sees to it that ϕ, i is performing a winning action to the effect that ϕ.

STIT, Games and Independence. The independence condition—R2 from section 2—may sound strong and even surprising if we cast STIT against the background of the physical world and our role in its changes: from an intuitive standpoint, it is very infrequent that we are beyond any possibility of being deprived of our choices by others. However, the principle makes good sense if we read STIT game-theoretically. If we trade, once again, choices for outcomes of actions, R2 will amount to the assumption highlighted at the beginning of section 3.1: each action profile (1) includes *one* action per agent,[25] and (2) correspond to *one* state. Independence is nothing but this, and thus proves a very game-theoretically oriented feature of STIT. The logical principles IA and TC, likewise in subsection 2.1, follow from R2 and the definition of $[i]$ (namely, the truth-clause TC4).

Independence and non-winning actions. There is an interesting feature of STIT which is not usually highlighted: in principle, you can have condition R2 *without* having an operator for agency which coincides with 'displaying a *winning action*'.

[24] $o(a_i)$ is a *best action* of i iff $o(a_i, a_{\bar{i}}) \succeq_i o(a'_i, a_{\bar{i}})$ for all $a'_i \in A_i$.

[25] This also explains why the function $Ch(i)$ is defined as a *partition* (see end of section 2).

This may become clear by analogy with *strictly competitive games*. In such games, the outcome function o may be defined according to the points (1) and (2) above, exactly as we just did for strategic games in general. At the same time, these games are characterized by the fact that no agent has a winning action: all actions of all the different agents are compatible, but no action of a single agent can ensure a given result: any relevant result in such games depend on the interaction of the different agents. One can have quite the same in a STIT setting, if the truth-clause of the stit operator is weakened; for instance, the probabilistic STIT presented by Broersen in [12] and [13] retain the independence of agents at the level of the frames, but defines an operator which equates with displaying the choice that comes with the highest chance of success in getting the given result (and the latter is clearly compatible with failure in ensuring the result).

Independence, continued. Game theory goes much further than giving formal expression to the notion of "winning action," as strictly competitive games prove. Recent work in STIT logic shows that its potential is not confined to that notion. At the same time, it suggests that adapting STIT to a broader set of game-theoretical notions is compatible with retaining the independence of agents. Approaching some phenomena of game theory without imposing independence is clearly possible and equally sound. For instance, van Benthem and Pacuit [8] suggests dropping the totality of the outcome function o in order to model the game-theoretical notion of a *correlation*, where there is some form of dependence between some agent's choices. This suggestion is in line with a general tradition of Dynamic Logics in modeling game-theoretical notions. The suggestion is very reasonable, but if the temporal aspect of choice-making is acknowledged, STIT provides a natural alternative: agents are independent when it comes to their simultaneous choices, but the present choice of one agent may limit the choices available to others at subsequent moments.[26]

Preferences and Ought. The *consequentialist* CKF$^+$ are a further proof of the entwinement of STIT and game theory. Indeed, CCKF$^+$ have their origins in the application of STIT to a consequentialist perspective on action, which was first carried by Horty in [22] and [24]. Roughly speaking, a consequentialist perspective evaluates what an agent *ought to do* on the ground of the value that can be attached to the consequences of the agent's choices; if we read such consequences as the sets of states extending each choices, we will define our framework by assigning values to states and—indirectly—to choices, exactly as we did with the preference relations.

There is one point, however, where the analysis of [24] significantly differs from the present CCKF$^+$: in [24] the values (or preferences) are *agent-independent*, that is the values it imposes does not vary with the agent in question.[27] Here, we re-

[26] This is no proof that STIT can deal with correlation as intended by [8], but is a general sign of the adaptability of STIT relative to the issue of independence.

[27] Also, notice that the "utilitarian STIT frames" introduced by [24] are grounded on branching-time structures. In such frames, the value attached to a history is not only agent-independent, but also *moment-independent*, that is it does not vary with time. This reminds the definition of preferences and priorities in standard rational-choice theory.

laxed this condition and allowed for preferences to be agent-relative. This relaxation shows the match between the consequentialist perspective implicit in game theory and the perspective encoded in the STIT analyses based on [24]. Also, we need to go to agent-relative preferences if we wish to model other game-theoretical notions which are "intrinsically multi-agent." Think of the notion of a Nash Equilibrium: it implies a consideration of the preferences of *all* different agents, and keeping preferences agent-independent would make such a consideration trivial. The same applies to other phenomena, like the removal of strictly dominated strategies.

The main modal operator in [24] is the so-called "dominance ought operator" $\bigodot[i]$—where $\bigodot[i]\phi$ should be taken to mean that, if i exerts any of her weakly dominant choices, then ϕ is the case. The operator clearly models a notion of weak dominance. An interesting point is that allowing for agent-relative preferences, as we do here, also allows for the definition of alternative operators meaning, for example, that, if i exerts one of her *best choices*, then ϕ is the case, or if i exerts one choice of her in the *Nash Equilibrium*, then ϕ is the case. Finally, as in the work of Kooi and Tamminga [25] and Tamminga [27], we can consider the choices of i relative to the utility they have for j, and define an operator meaning, intuitively, that, if i exerts any choice of her *that is weakly dominant for* j, then ϕ is the case. This work, which opens the interesting issue of modeling the notion of "acting in the interest of someone else," was elaborated by Turrini [28] to apply also to notions of dominance taking into account the interests of other groups, or coalitions.

4 STIT and Matrix Game Logic: Ex Interim Knowledge

Matrix Game Logic—or MGL, for short—made its first appearance in van Benthem [5], in order to model the notion of iterated removal of strictly dominated strategies. The logic was later enriched with a notion of "freedom" that deserves attention, since it is thought to capture the margin of action that a combined choice of all the other agents leaves to a particular agent i. Van Benthem and Pacuit [8] contains a very interesting comparison between STIT and MGL, and proves that there is an embedding of the former into the latter. Here, we continue the comparison with a mutual embedding between MGL's operator for *ex interim knowledge*—e.i.-knowledge, for short—and the Chellas stit. More important, we push the comparison further and show that a mutual embedding holds between MGL's operator for freedom and a Chellas stit for the agency of anti-groups. This determines interesting property transfers, sheds some light on the freedom operator, and allows some interesting considerations on the applications of STIT.

4.1 Matrix Game Logic for Epistemic Notions and STIT: a Formal Comparison

A matrix game frame, or MGF, is a structure $\mathscr{MG} = \langle \mathscr{G}, \{\sim_i \mid i \in Ags\}, \{\approx_i \mid i \in Ags\}\rangle$, specified as follows. $\mathscr{G}$ is a game, as defined in subsection 3.1.[28] For each agent i, $\sim_i$ is an *equivalence relation* which represents the *ex interim* (*e.i.*-) *uncertainty* of i: if $w \sim_i w'$, then i is uncertain whether w or w' is the actual state *after i performs her action* (see more details below). Finally, $\approx_i$ is the "freedom relation for i": $w \approx_i w'$ just in case w' is in the outcome $o(a_{\bar{i}})$ which also includes w; in other words, the relation models the 'margin of freedom" of i—what i is free to achieve by changing her action, while the given action of its anti-group $\bar{i}$ is kept fixed. A matrix game model, or MGM, obtains by extending a MGF with an evaluation function V in the standard way.

Figure 4:

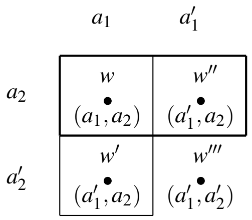
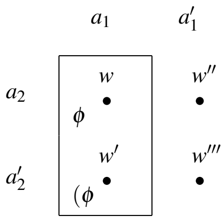


Figure 5: $K_1\phi$ is true at w (and w')



(we omit the freedom relation and do not specify action profiles here)

Figure 4 exemplifies a MGF with two agents. Here, a_1 is the action 1 performs at w or w'. The thin box represents the class of states which are e.i.-equivalent with w; i is uncertain where state w or w' are the actual state of the world, since the latter crucially depends on what action 2 performs, and—coherently with the characterization of e.i.-knowledge in games, see below—i has only *ex post* knowledge of this. The thick box represents the states which are "freedom equivalent" with w: w itself, which results from (a_1, a_2) and w'', which results from switching from a_1 to a'_1 while keeping the same action of 2—that is, a_2. Let us set aside the freedom relation for now and focus on the e.i.-uncertainty relation $\sim_i$.

Games, Ex interim Knowledge, and Perfect Information. Typically, the game-theoretical literature distinguishes three kinds of knowledge: *ex ante* knowledge is knowledge of the rules of the game, *ex interim knowledge* is the knowledge one has

[28] In its original version, MGL sees *action profiles* themselves as *states*. Thus, W and o are not included in the original definition of a MGF. Here we consider situations where there is no action-profile gap, which can be in turn seen as situations where the function o is total and no restriction is imposed on the construction of action profiles. In this case, MGF's can be defined as in the text.

after performing an action, *e post* knowledge is knowledge of what actions others play and what state results from the game. The games we consider here presuppose *perfect information*. In such games, agents are not uncertain about the state of the world, they know their own preferences and those of other agents, and they have common knowledge of rationality. Since what results from *i*'s action is determined by the rules of the game, and since agents have no uncertainty about the states they are in before the game-round, it is easy to see that in games with perfect information e.i.-knowledge is essentially knowledge of what one does.

The epistemic fragment EGML. For the time being, let us drop the freedom relations $\approx_i$ from MGM's, thus obtaining *epistemic* MGM's, or EMGM's EMGM. We will likewise refer to the epistemic fragment of MGL, or EMGL, as that logic whose only operators are the agent-relative e.i.-knowledge operators K_i, with $i \in Ags$, and defined as follows:

$$(\mathrm{TC}K_i)\ \mathscr{MG}^{\mathscr{E}}, w \models K_i\phi \ \text{ iff } \text{ for all } w' \sim_i w,\ \mathscr{MG}^{\mathscr{E}}, w' \models \phi$$

where $\mathscr{MG}^{\mathscr{E}}$ is an EMGM. Figure 5 gives an intuitive grasp of TCK_i. As is clear from van Benthem [5] and [7], and van Benthem and Pacuit [8], the $\sim_i$ relation satisfies strong confluence—we have, that is, $W \times W \ \subseteq \sim_1 \circ \sim_2 \circ \ldots \circ \sim n$. This is a plausible principle for e.i.-knowledge. In game theory, for each agent *i* e.i.-knowledge is basically knowledge of what is due to the action *i* performs. Suppose a_i is the action in question and consider any states $w, w' \in o(a_i)$. Of course, we have $w \sim_i w'$: *i* knows that, due to her action, some state in $o(a_i)$ results from the game round, but she does not know which does. This crucially depends on the action $\bar{i}$ plays, which is something *i* knows only *ex post*, that is, when the result is settled. Thus, if $w, w' \in o(a_i)$ then $w \sim_i w'$. But also the converse holds. Suppose $w \in o(a_i)$ and $w' \notin o(a_i)$. In performing a_i, *i* also gets e.i.-knowledge of what states she is selecting away. Hence, $w \nsim_i w'$ As a consequence, we have $w \sim_i w'$ iff $w, w' \in o(a_i)$. Strong confluence and equivalence are the only conditions defining $\sim_i$, and thus, by Proposition 1, we know that

For every EMGM $\mathscr{MG}^{\mathscr{E}}$, $w \sim_i w'$ iff $w \sim_i^C w'$ in $\mathscr{C}^{\mathscr{MG}^{\mathscr{E}}}$

where $\mathscr{C}^{\mathscr{MG}^{\mathscr{E}}}$ is the CCKF$^+$ corresponding to the given EMGM $\mathscr{MG}^{\mathscr{E}}$ (with $Ch(i) = \{o(a_i) \mid a_i \in A_i\}$ for every *i*). We can therefore define the following truth-preserving translation τ:

$$\tau([i]\phi) \ = \ K_i\phi$$

which in turn guarantees a mutual embedding between CSTIT and EMGL.[29] This comes with very convenient properties: indeed, it now follows from the similar properties of CSTIT that EMGL is decidable and finitely axiomatizable, no matter the number of agents.

[29] The translation τ above is already defined in and [7] and [8]. However, it does not define a mutual embedding there, since the full MGL is considered, and as we shall see, CSTIT is a proper fragment of it.

Strong confluence also allows us to introduce □ as short for K_jK_i, exactly as we did with $[j][i]$, and guarantees an interesting principle: $K_iK_j\phi \to \Box\phi$. The analysis of e.i.-knowledge above suffices to explain why: if i has e.i.-knowledge of what follows from the action of another agent j, this means that what result from j's action was trivial.

Ex Interim Knowledge and Best Actions between Matrix Game Logic and STIT. A strategy, or action, a_i of agent i is *strictly dominated* if there is another strategy—or action—a'_i such that $o(a_i) \prec_i o(a'_i)$—that is, a'_i is strictly preferred to a_i by i, no matter what action $a_{\bar{i}}$ is performed by $\bar{i}$. Since no agent would play a strictly dominated strategy, in foreseeing the moves of the other players, we may remove their strictly dominated strategies. This will create a sub-game and change the range of the agents' preferences; new strictly dominated strategies will emerge and will once again be removed, and so on, step-by-step. This is the procedure of iterated removal of strictly dominated strategies.

Figure 6: a game model with two agents, each with three actions available.

	a_1	a'_1	a''_1
a_2	w_1 ● 2,3	w_2 ● 2,2	w_3 ● 1,1
a'_2	w_4 ● 0,2	w_5 ● 4,0	w_6 ● 1,0
a''_2	w_7 ● 0,1	w_8 ● 1,4	w_9 ● 2,0

Figure 7: the sub-game obtaining by removing a''_1.

	a_1	a'_1
a_2	w_1 ● 2,3	w_2 ● 2,2
a'_2	w_4 ● 0,2	w_5 ● 4,0
a''_2	w_7 ● 0,1	w_8 ● 1,4

An example of iterated removal of strictly dominated strategies is given by Figures 6 and 7. The pairs of numbers denotes utilities of the two agents. It is clear that agent 1 will not play a''_2, since it is a strictly dominated strategy for her. This action will then be removed, thus generating the sub-game in Figure 7. The iterated removal of strictly dominated strategies would then continue by removing a''_2, then a'_1, and then a'_2, thus generating three further sub-games. The last one is constituted by w_1 alone, which is the Nash Equilibrium of the initial game. Following a tradition in game theory, MGL explains iterated removal of strictly dominated strategies in *epistemic terms*: considerations about rationality, interests and strategies of others lead us to remove some strategies from an initial EMGM, and thus transform it in a sub-EMGM where other strategies become dominated. Becoming, model transformation, rationality: all this naturally calls for Dynamic Logic, and EMGL has been an answer to the call. An interesting consequence of the embedding above is that STIT can also capture these dynamics.

We cannot describe this in detail here, but we give the basic ingredients. First, we confine ourselves to finite EMGM's and CCKF$^+$'s—that is frames where each agent has a finite number of actions available. By the definition of action profiles and the outcome function, this suffices to guarantee a finite number of states. The notions involved are those of *best action* and the *epistemic notions* of *e.i.-knowledge*, *strong rationality* and *weak rationality*. Let us consider these in order.

Best actions are represented by van Benthem [5] as atoms b_1, b_2, b_3, ... They are agent-indexed and defined by the truth-clause: $\mathscr{C}^{\mathscr{M}\mathscr{G}^{\mathscr{E}}}, w \models b_i$ iff there is an action $a_i \in A_i$ such that $w \in o(a_i, a_{\bar{i}})$ such that $o(a_i, a_{\bar{i}}) \succeq_i o(a'_i, a_{\bar{i}})$ for all $a'_i \in A_i$. In other words, the atom b_i is to be interpreted as meaning that i is performing her best action, and it is true at all and only those states which are in the outcome of some best action of i.[30] An extension of $\mathscr{L}_{\mathsf{CSTIT}}$ with the same propositional constants is straightforward, and thus also this notion can be expressed by STIT. The notion of a best action is in turn indispensable to define the notions of *strong* and *weak* rationality.

Strong Rationality is expressed by the sentence $\neg K_i \neg b_i$ ('*i* does not know that she is not performing one of her best actions'). Basically, then, i is strongly rational if she knows she has at least one best action over the whole game; and note that K_i satisfies *negative introspection*, and thus we have $\neg K_i \neg b_i \leftrightarrow K_i \neg K_i \neg b_i$. Van Benthem proves that for every agent i, sentences of strong rationality hold in at least some state of a *finite* EMGM, though the same may fail for infinite EMGM or even for sub-EMGM.

Given our translation τ and the extensions with atoms, STIT can express strong rationality by $\neg[i]\neg b_i$, now taken to mean that i does not prevent herself from performing a best action. This reading points out at the purely agentive side of strong rationality: in a game-theoretical context, rational agents do not play actions different from their best ones.[31]

Weak rationality actually leads to the "dynamic" part of EMGL. While strong rationality consists in not choosing a strategy that is strictly dominated in the whole given EMGM, weak rationality consists in not choosing a strategy that is not strictly dominated in the sub-EMGM in question. The difference can be appreciated by considering those cases where agents do not know that their current action is best relative to the whole game, but where they do know that such an action has no better alternative, where alternatives are now limited to the sub-EMGM considered. This also requires "relative best actions"—that is atoms b^*_1, b^*_2, b^*_3, ... Where W^* is the set of states of the sub-EMGM into account, the truth-clause for these new atoms is: $w^* \models b^*_i$ iff there is an action $a_i \in A_i$ with $w \in o(a_i, a_{\bar{i}})$ and such that $o(a_i, a_{\bar{i}}) \succeq_i o(a'_i, a_{\bar{i}})$ for all $a'_i \in A_i$ and $o(a_i), o(a'_i), o(a_{\bar{i}}) \subseteq W^*$. In other words, b_i

[30] It may seem that introducing linguistic atoms $b_1, b_2, b_3, \ldots$ to express the notion of a best action is a kind of trick. However, the move makes sense if the goal is not providing an analysis or of the notion, but simply to give us linguistic means to express the fact that such an action is being performed.

[31] Theorem 6 in [5] is easily adapted to finite CCKF$^+$.

states that i is performing her best action relative to the actions which have not been removed.

Weak rationality is expressed by $\neg K_i \neg b_i^*$. The sentence in turn proves interesting since it is false at any state which extends the outcome of a strictly dominated strategy of i. Remarkably, given a state w in the solution zone for strictly dominated strategy removal, repeating assertions of weak rationality stabilizes at a sub-game which include w and whose domain is in that solution zone, an observation due to van Benthem [5, theorem 7].

If we extend STIT with atoms for relative best actions, we get that weak rationality of i is expressed by $\neg[i]\neg b_i^*$: if i is rational, then she does not prevent herself from performing a relative best action. Also, it is easy to see that van Benthem's result, mentioned just above, can be straightforwardly adapted to STIT. The interesting point is that, where w is in the solution zone for strictly dominated strategies in a given game, the iterated removal of strictly dominated strategies stabilizes at a sub-game which includes w and solves the game if we iterate the choice of not preventing ourselves from playing our best action—where "iterating the choice" here means that we apply it at any sub-game resulting in the removal process.

A very brief conceptual insight. The mutual embedding has shown a surprising virtue of STIT: though designed to express *purely* agentive notions, in some situations it can also express interesting epistemic notions, such as *e.i.-knowledge*, *strong* and *weak rationality*. Thus, STIT also shows potential relative to certain notions which are crucial in the epistemic foundations of game theory. In particular, STIT can capture some crucial notions in iterated removal of strictly dominated strategies. The crucial issue here is what exact notions can be expressed by STIT. The ability to express strong rationality does not prove a striking result, once the mutual embedding between CSTIT and EMGL is considered. The interesting point is rather the expression of weak rationality. Indeed, such a notion seems to witness the dynamic character of the iterated removal process, and the surprising point is that STIT can frame some of this character, though it has not been designed to capture those dynamics of model-transformation which are captured by Dynamic Logics. At the same time, we need to be aware of the limits of such connections. The possibility of connecting e.i.-knowledge and seeing to it is confined to strategic games with perfect information: if an agent i could be wrong about the current state of the world, she could also be confused about the results of her action, and thus she could see to it that ϕ without having e.i.-knowledge of ϕ. In this situation, the gap between the agentive dimension encoded by STIT and the epistemic dimension encoded by K_i resurfaces, and the latter must be explicitly introduced in the logic. Also, the ability to capture some dynamic features of the iterated removal process does not guarantee the possibility of capturing *any* dynamic aspect of action and game-theoretical interaction.

5 Matrix Game Logic, Freedom and STIT

Let us now return to 'full' MGL and the 'freedom relations', which, as we recall, are defined as follows: $w \approx_i w'$ iff $w, w' \in o(a_{\bar{i}})$ for some action $a_{\bar{i}}$ of $\bar{i}$. The definition of the "freedom operator" is where the new relation comes in:

$$(\mathrm{TC}[\approx_i])\ \mathscr{MG}, w \models [\approx_i]\phi \ \text{ iff } \ \text{for all } w' \approx_i w, \text{ then } \mathscr{MG}, w' \models \phi$$

Here, $[\approx_i]\phi$ should be taken to mean that i is left free to achieve ϕ, by $\bar{i}$'s current action. Figure 8 provides an example:

Figure 8: $[\approx_1]\phi$ is true at w (and w'')

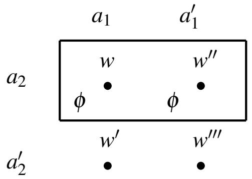
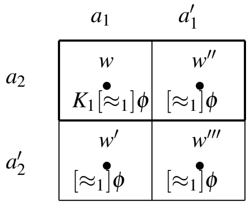


Figure 9: What follows from having



$K_1[\approx_1]\phi$ true at w (we include the e.i.-uncertainty relation here)

Figure 8 is based on Figure 4, but it omits labels for action profiles and the column representing the e.i.-knowledge of 1 at w and w'. What is left from Figure 4 is then the margin of freedom of i at w and w''. $[\approx_1]\phi$ is true at w since it is true in all the states which are freedom-equivalent to w—namely, w itself and w''. The same holds if we evaluate the formula at w''. Figure 9 extends Figure 8 with the e.i.-knowledge of i at w (represented by the left column) and w'' (represented by the right column). The figure provides an easy way to check the validity of $K_i[\approx_i]\phi \ \rightarrow \ \Box\phi$ below.

The new freedom operator carries a lot of information: it calls upon the action of $\bar{i}$, as is clear from the definition of $\approx_i$, and tells us what options the current action of $\bar{i}$'s leaves to i. This involves much more than talking about the individual agent i and what action she performs, of course: it also implies reference to a more complex concept, which hints at the way some of i's actions interact with those of $\bar{i}$. As a consequence $[\approx_i]$ is a complex operator, and a systematic investigation of its property may prove difficult.

Two clear features are that $[\approx_i]$ is an S5 operator and that it satisfies a strong confluence axiom in combination with K_i—that is, $W \times W \ \subseteq \sim_i \circ \approx_i$. These observations are established by van Benthem and Pacuit in [7] and [8]), and illustrated in Figure 9. As a result, we have the principle:

$$K_i[\approx_i]\phi \ \rightarrow \Box\phi$$

This is an interesting principle: it states that only in case ϕ is trivial can the agent i e.i.-know whether she is free to achieve ϕ; knowledge of what is not excluded by the choice of an agent's own anti-group is trivial, and so not interesting. The validity of the formula is easily checked with the help of Figure 9: here, $K_i[\approx_i]\phi$ is true at w. By this and TCK_i, we have that $[\approx_i]\phi$ is true at w and w'. Since $[\approx_i]\phi$ is true at w, we have that $[\approx_i]\phi$ is true at w'', and since $[\approx_i]\phi$ is true at w', we have that $[\approx_i]\phi$ is true at w'''. $[\approx_i]\phi$ is then true at all the states of the model represented by Figure 9. By this and TC$[\approx_i]$, we have that ϕ is true at every state of the model. As a consequence, we have $\Box\phi$ true at all of them.

Other questions remain open. Can $[\approx_i]$, $[\approx_j]$ be turned into stit operators? What about a confluence property involving freedom operators only? Is MGL decidable and finitely axiomatizable? These questions have not yet been settled in the literature. We do this below, but in order for us to show the result, we need to extend CSTIT to *group* CSTIT first.

A STIT logic for Groups. MGL is a proper extension of EMGL. Since there is a mutual embedding between the latter and CSTIT, we have that CSTIT is a proper fragment of MGL. However, the relation may change (and indeed changes) if we consider *group* CSTIT, which allows us to talk, among the other things, of groups, singletons, and anti-groups.[32] Let us start from a full group STIT and then isolate the anti-group fragment.

A (relational) *group* CKF$^+$—henceforth, a GCKF$^+$—is a triple $\langle W, Ags, \{\sim^C_I \mid I \subseteq Ags\}, \{\succeq_i \mid I \subseteq Ags\}\rangle$, where W and Ags are as in CCKF$^+$, and $\sim^C_I$ is a group choice-equivalence relation between states in W, such that:

(R1′) $\sim^C_I$ is an equivalence relation
(R2′) $W \times W \ \subseteq \sim^C_{J_1} \circ \ldots \sim^C_{J_n}$ if $J_i \cap J_k \ = \ \emptyset$ for all $i,k \in \{1,\ldots,n\}$
(R3′) $\sim_I = \sim_J \cap \sim_{I/J}$ where $J \ \subseteq \ I$
(R4′) $\sim^C_{Ags} = \mid W \mid$
(R5′) $\sim^C_I \subseteq \sim^C_J$ if $J \subseteq \ I$

Here, R2′ is the group version of the strong confluence (notice the restriction to *disjoint* groups), while R3′–R4′ are additivity (the choice of a group is the intersection of the choices of its disjoint subgroups)[33] and grand group determinism (the

[32] *Group* STIT was already present at the beginning of STIT; see Belnap [4] and Horty [24]. Both Belnap and Horty assume additivity; see [4, definition 10-3], and [24, definition 2.10]. Neither Belnap nor Horty assume condition R4′ below—that the joint agency of all the agents may determine a unique outcome—although, as mentioned earlier, Horty [24, p. 91] considers models that satisfy this condition. Basically, the conditions we present here build on those presented in Horty [24] by adding the standard game-theoretical condition of coalitional monotonicity. The latter can be actually derived by R2′, but we present it here as a basic condition in order to conform with the standard presentation of group STIT.

[33] The standard condition in game theory is actually *superadditivity*, which allows for the choice of I to be a subset of the choices of I's members; the condition is actually a consequence of coalition

grand group can determine a unique state of the world).[34] The two principles are clearly the group versions' of R3–R4. R5 is the so-called coalition monotonicity which holds that, if agents (and groups) join their efforts, they improve their result; the condition mirrors the standard assumptions about coalition effectivity functions in game theory. An individual agents i is here taken as a special case of groups, namely the singleton $\{i\}$, we keep the individual notation for the sake of readability. An analog to Proposition 1, above, also applies to GCKF$^+$'s—indeed the actual statement of Theorem 1 in Turrini [28] is about coalitions. Finally, the new operator $[I]$ is just the group version of $[i]$, so that: $[I]\phi$ is true at w in a model for CGKF$^+$ iff ϕ is true at every state $w' \sim_I^C w$.

The principles R1, R2, R3 and R5 correspond to the following principles:

(P1) S5 axioms for $[I]$;
(P2) $[I][J]\phi \ \rightarrow \ \Box\phi$, where $I \ \cap \ J \ = \emptyset$
(P4) $\phi \leftrightarrow [Ags]\phi$
(P5) $[J]\phi \ \rightarrow [I]\phi$ if $J \subseteq \ I$

And, as the reader will notice, many distinctive features of individual STIT are transferred to the group level, though with restrictions: for instance, we have $[I][J]\phi \rightarrow \Box\phi$ if $I \ \cap \ J \ = \emptyset$, but otherwise the principle may fail. Also, it is easy to show that $\Diamond[I]\phi \wedge \Diamond[J]\psi \ \rightarrow \ \Diamond([I]\phi \wedge [J]\psi)$ holds *if* $I \neq J$. Herzig and Schwarzentruber have proved that group CSTIT with more than two agents is undecidable and is not finitely axiomatizable; see [21, theorems 22 and 23]. This is primarily due to the fact that strong confluence fails if $I \ \cap \ J \ \neq \emptyset$; if—additionally—the groups I and J overlap without being subsets one of another, then the logic can be mapped into $S5^n$, with n the number of agents in the group STIT in question. Since all extensions of S5^3 are undecidable and not finitely axiomatizable so is group CSTIT with three or more agents.[35]

STIT, Anti-groups and Freedom. Let us call *anti-group* CSTIT that fragment of *group* CSTIT where only agents in Ags and their anti-groups are included. Some interesting connections between individual agents and anti-groups are easily captured:

(R2$''$) $W \times W \ \subseteq \sim_i^C \circ \sim_{\bar{i}}^C$ for every $i \in Ags$
(R5$''$) $\sim_i^C \subseteq \sim_{\bar{j}}^C$ for all $j \in \bar{i}$

R2$''$ holds since an agent and her anti-group satisfy by definition the disjointness proviso in R2$'$; R5$''$ holds because, by definition, any agent will be a subgroup of the anti-group of any other agent. As a consequence,

(P2$''$) $[i][\bar{i}]\phi \rightarrow \Box\phi$
(P5$''$) $[i]\phi \rightarrow \bigwedge_{j\in\bar{i}Ags}[\bar{j}]\phi$

monotonicity. However, we prefer to include R4 in order to comply with the standard choice in group STIT, and also because it makes the construction of groups conceptually easier.

[34] The principle is also called "Rectangularity" in Turrini [28].

[35] Again, see Hirsch, Hodkinson, and Kurucz [19] for these results concerning $S5^n$.

hold. The most interesting connection, however, is with MGL's $\approx_i$. Indeed, since the correspondence result in section 3 extend to group STIT and coalitional games (see above), we have that: $w \sim_{\bar{i}}^{C} w'$ iff $w, w' \in o(a_{\bar{i}})$ for some action $a_{\bar{i}}$ of $\bar{i}$. From this, it follows that the relation $\sim_{\bar{i}}^{C}$ in GCKF^{+} is nothing but the relation $\approx_i$ in MGM. We thus have:

for every MGM $\mathscr{MG}$, $w \approx_i w'$ iff $w \sim_{\bar{i}}^{C} w'$ in $\mathscr{C}^{\mathscr{MG}}$

and we can therefore define a truth-preserving translation τ' such that

$$\begin{aligned} \tau'([i]\phi) &= K_i\phi \\ \tau'([\bar{i}]\phi) &= [\approx_i]\phi \end{aligned}$$

where $\mathscr{C}^{\mathscr{MG}}$ is the GCKF^{+} corresponding to the given MGM $\mathscr{MG}$ (with $Ch(i) = \{o(a_i) \mid a_i \in A_i\}$ for every i). There is therefore a mutual embedding between *anti-group* CSTIT and MGL. This answers one question we asked earlier: $[\approx_i]$ can indeed be interpreted as a stit operator. And with this answer comes both bad news and good news.

Bad news. The bad news is that we can now conclude that the "full" MGL[36] is undecidable and not finitely axiomatizable. This follows from the mutual embedding, together with the fact that anti-group CSTIT with more than two agents has these properties, which transmit to MGL.[37]

Good news. The good news is that we can gain insight into the properties of $[\approx_i]$ via established results about group stit operators—particularly those concerning $\bar{i}$. For instance, we can now see that $K_i\phi \;\rightarrow\; \bigwedge_{j\in\bar{i}}[\approx_j]\phi$ holds, from P5$''$ and τ'. This is an interesting principle: it states that, if i has e.i.-knowledge that ϕ, then she is not excluding that any other agent j achieves ϕ. If we dig into the conditions that define the e.i.-knowledge and freedom relations, it is evident that this principle is sensible: the agent i can have e.i.-knowledge that ϕ because her current action removes the possibility of achieving "non-ϕ" states. Thus, j also has a margin to achieve ϕ with her current action, while it is excluded that she achieves $\neg\phi$ with any of her available actions.

The mutual embedding also helps us understand the issue of strong confluence. Contrary to what happens with K_i, the freedom operator $[\approx_i]$ does not satisfy the strong confluence property: $[\approx_i][\approx_j]\phi \;\rightarrow\; \Box\phi$ does not hold, since $[\bar{i}][\bar{j}]\phi \rightarrow \Box\phi$ does not hold in group CSTIT with more than two agents, since, in that case, $\bar{i} \cap \bar{j} \neq \emptyset$. This failure implies that there are cases where ϕ is not trivial and yet i has a

[36] Here we mean MGL as defined in this paper, , not the full logic defined in van Benthem [7], which also includes an operator for preferences.

[37] See our observation above on the conditions for undecidability and failure of finite axiomatizability in group CSTIT. Of course, decidability and finite axiomatizability are restored if we confine to MGL with only two agents, so that $Ags = \{1,2\}$. In that case, $\bar{1} = 2$ and $\bar{2} = 1$. The anti-groups thus collapse into different agents, and MGL with two agents actually collapse into EMGL with two agents—which is indeed decidable and finitely axiomatizable, since EMGL is, no matter the cardinality of Ags. Thus, the case with two agents does hold much interest.

margin of freedom to let j have a margin of freedom to achieve ϕ—or equivalently: the current choice of i's anti-group does not imply that j's anti-group achieves $\neg\phi$. Transmission of freedom, it turns out, is not trivial, after all!

For analogous reasons, $\bigwedge_{i\in Ags}\Diamond[\approx i]\phi_i \rightarrow \Diamond\bigwedge_{i\in Ags}[\approx i]\phi_i$ also fails in MGL: even though two different agents 1 and 2 are left free to achieve ϕ_1 and ϕ_2 respectively, their results may be incompatible. Thus, agents are not independent in their margins of freedom. This sounds plausible: after all, the margins of freedom that one agent has *depend* on what the current choice of the other agents.

6 Conclusions

In this paper we have accomplished two main tasks. First, we have highlighted the ties between STIT and the basic settings of game theory. This has involved demonstrating the possibility of reading STIT game-theoretically and expressing game-theoretical notions in STIT's terms. The connection thus established proves a very good hint at the flexibility and richness of STIT theory.

Second, we have considered the MGL logic of games, a form of Dynamic Logic, and have furthered the comparison begun by van Benthem and Pacuit [8] between STIT and MGL. This comparison has led, we believe, to some interesting results. First, as noted in this earlier work, the "epistemic fragment" of MGL has a mutual embedding with the logic CSTIT for individual agency; thus, STIT has the potential to capture the notions of ex interim knowledge and the assertions of weak and strong rationality. Also, decidability and finite axiomatizability transmit from CSTIT and the fragment of MGL.

It was established here, however, that full MGL, including the "freedom operator," has a mutual embedding with a group version of CSTIT which includes arbitrarily many individual agents and their anti-groups. This suffices to secure that full MGL is undecidable and not finitely axiomatizable.

However, the embedding also allows us to explore issues about the freedom operator—which is conceptually very rich—"through the mirror" of STIT. This helps us to notice that the freedom operator does not obey independence (for reasons which are explained by the very setting of group STIT), and shows an interesting relation between the ex interim knowledge of an agent and the margin of freedom left to all other agents.

This proves STIT to be illuminating, not only for its own sake, but also as a tool for developing formal and conceptual perspectives on other frameworks for agency. A thorough comparison with Dynamic Epistemic Logic in the style of Baltag, Moss, and Solecki [2] could be a further interesting step in bridging STIT and the dynamic framework. The ground for this has been provided in [8]. The merging of the two methodologies could prove extremely fruitful in the modeling of multi-agent situations where it is crucial to express whether the information update in the doxastic state of agent i has been brought about i herself or passively received by other agents.

Acknowledgements. This work was carried out while Roberto Ciuni was a Humboldt Postdoctoral Fellow with the project 'A Tempo-Modal Logic for Responsibility Attribution in Many-Step Actions' (2011–2013).

References

1. Balbiani Philippe, Herzig Andreas, Troquard Nicolas (2008) Alternative Axiomatics and Complexity of Deliberative STIT Theories. Journal of Philosophical Logic, 37(4): 387–406.
2. Baltag Alexandru, Moss Lawrence S., Solecki Slawomir (1998) The Logic of Public Announcements, Common Knowledge and Private Suspicions, Proceedings TARK, Los Altos, Morgan Kaufmann Publishers (Updated versions through 2004), pp. 43–56.
3. Belnap Nuel, Perloff Michael (1988) Seeing to it that: a canonical form for agentives, Theoria, 54: 175–199.
4. Belnap Nuel, Perloff Michael, Xu Ming (2001) Facing the Future: agents and choices in our Indeterminist World, Oxford, Oxford University Press.
5. van Bethem Johan (2007) Rational Dynamics and Epistemic Logic in Games. International Game Theory Review, 9(1): 13–45.
6. van Bethem Johan (2011) Logical Dynamics of Information and Interaction, Cambridge, Cambridge University Press.
7. van Bethem Johan (forthcoming) Logic in Games, Cambridge (MA), The MIT Press.
8. van Benthem Johan, Pacuit Eric (forthcoming) Connecting Logics for Choice and Change, in Müller Thomas (ed.) Volume in Honour of Nuel Belnap. Berlin, Springer, Outstanding Logicians Series.
9. Broersen Jan, Herzig Andreas, Troquard Nicholas (2006) From Coalition Logic to STIT. Electronic Notes in Theoretical Computer Science, 157: 23–35.
10. Broersen Jan, Herzig Andreas, Troquard Nicholas (2006) Embedding ATL in Strategic STIT Logic of Agency. Journal of Computation, 16(5): 559–578.
11. Broersen Jan (2011a) A Deontic Epistemic Stit Logic Distinguishing Modes of 'Mens Rea', Journal of Applied Logic, 9(2): 137–152.
12. Broersen Jan (2011) Probabilistic Stit Logic. Proceedings 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011), Lecture Notes in Artificial Intelligence, 6717, Berlin, Springer pp. 521–531.
13. Broersen Jan (2011) Modeling Attempt and Action Failure in Probabilistic Stit Logic. Proceedings of Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011), pp. 792–797.
14. Chellas Brian (1969) The Logical Form of Imperatives. PhD thesis, Philosophy Department, Stanford University.
15. Chellas Brian (1992) Time and Modality in the Logic of Agency. Studia Logica, 51: 485–517.
16. Ciuni Roberto, Mastop Rosja (2009) Attributing Distributed Responsibility in Stit Logic, in Xiangdong He, Horty John, Pacuit Eric (eds.) Logic, Rationality, and Interaction (Lecture Notes in Computer Science, Vol. 5834), Berlin, Springer, pp. 66–75.
17. Ciuni Roberto (2010) From Achievement Stit to Metric Possible Choices, Logica 2009 Yearbook, London, College Publications, pp. 33–46.
18. Ciuni Roberto, Zanardo Alberto (2010) Completeness of a Branching-Time Logic with Possible Choices, Studia Logica, 96(3): 393–420.
19. Hirsch Robin, Hodkinson Ian, Kurucz Agi (2002) On Modal Logics between K×K×K and S5×S5×S5. The Journal of Symbolic Logic, 67(1): 221–234.
20. Herzig Andreas, Lorini Emiliano (2010) A Dynamic Logic of Agency I: STIT, Abilities and Powers. Journal of Logic, Language and Information, 19(1): 89–121.
21. Herzig Andreas, Schwarzentruber Francois (2008) Properties of Logics for Individual and Group Agency, in Areces Carlos and Goldblatt Robert (eds.) Advances in Modal Logic vol. VII, London, College Publications, pp. 133–149.

22. Horty John (1996) Agency and Obligation. Synthese, 108(2): 269–307.
23. Horty John, Belnap Nuel (1995) The deliberative stit: A study of action, omission, and obligation. Journal of Philosophical Logic, 24(6): 583–644.
24. Horty John (2001) Agency and Deontic Logic, Oxford, Oxford University Press.
25. Kooi Barteld and Allard Tamminga (2008) Moral Conflicts between Groups of Agents. Journal of Philosophical Logic, 37: 1–21.
26. Lorini Emiliano (forthcoming) A STIT logic analysis of commitment and its dynamics. Forthcoming in the Journal of Applied Non-Classical Logic.
27. Tamminga Allard (2013) Deontic Logic for Strategic Games. Erkenntnis, 78(1): 183–200.
28. Turrini Paolo (2012) Agreements as Norms, in Agotnes Thomas, Broersen Jan and Elgesem Dag (eds.) DEON 2012, LNAI 7393, Berlin, Springer, pp. 31–45.
29. Xu Ming (1994) Decidability of Deliberative Stit Theories with Multiple Agents. In Gabbay Dov and Ohlbach Hans (eds.) Proceedings of the First International Conference in Temporal Logic, Berlin, Springer, pp. 332-348.
30. Xu Ming (1998) Axioms for Deliberative Stit. Journal of Philosophical Logic, 27: 505–552.
31. Xu Ming (2010) Combinations of STIT and Actions. Journal of Logic, Language and Information, 19(4): 485–503.
32. Von Kutschera Franz (1986) Bewirken, Erkenntnis, 24(3): 253–281.
33. Zanardo Alberto (forthcoming) Indistinguishability, Choices and Logics of Agency. Studia Logica Special Issue “Advances in Philosophical Logic”.