Intentions and transformations of strategic games*

Olivier Roy
University of Groningen
O.Roy@rug.nl

Abstract

In this paper we take a game-theoretic perspective to study the effects of previously adopted intentions in rational decision making. We investigate the question of how agents transform the decision problems they face in the light of what they intend, and provide conditions under which such transformations, when iterated, leave room for deliberation, i.e. do not exclude all the options of the decision maker.

1 Introduction

In this paper we take a game-theoretic perspective to study the effects of previously adopted intentions in rational decision making. We investigate the question of how agents transform the decision problems they face in the light of what they intend, and provide conditions under which such transformations, when iterated, leave room for deliberation, i.e. do not exclude all the options of the decision maker.

There is a broad consensus among philosophers of action, e.g. [Bratman, 1987] and Velleman [2003], that previously adopted intentions, alongside beliefs and desires, shape decision problems. This role, however, has until now attracted little if no attention in game theory. The present work attempts at (partially) filling this gap. This not only results in a richer game-theoretic framework, but also sheds new lights on the philosophical theory of intentions, especially concerning the interactive character of intention-based transformations of decision problems.

The approach in this paper differs in important respects from the one in "BDI" architectures, e.g. Georgeff et al. [1998] and van der Hoek et al. [2007]. Studies in that paradigm have mainly focused on the relation that intentions can or should have with beliefs and desires, and on different policies of intention revision. Furthermore, these approaches do not directly use game-theoretic formalisms, but rather frameworks tailored for the analysis of multi-agent systems. Here we use strategic form games, and focus on how intentions transform them.

We consider two ways of transforming decision problems on the basis of the agents' intentions. For each of them we characterize the conditions under which they do not remove all possible choices for the agents. Proofs of the technical results can be found in the Appendix.

2 Strategic games with intentions

We use standard strategic form games, as in e.g. [Osborne and Rubinstein, 1994], except that preferences are represented qualitatively. A *decision problem* or *strategic game* $\mathbb{G}$ is a tuple $\langle I, S_i, X, \pi, \leq_i \rangle$ such that :

- I is a finite set of agents.
- S_i is a finite set of *actions* or *strategies* for i. A *strategy profile* $\sigma \in \Pi_{i \in I} S_i$ is a vector of strategies, one for each agent in I. The strategy s_i which i plays in the profile σ is noted $\sigma(i)$.
- X is a finite set of *outcomes*.
- $\pi : \Pi_{i \in I} S_i \to X$ is an *outcome function* that assigns to every strategy profile $\sigma \in \Pi_{i \in I} S_i$ an outcome $x \in X$. We use $\pi(s_i)$ to denote the set of outcomes that can result from the choice of s_i. Formally: $\pi(s_i) = \{x : x = \pi(s_i, \sigma_{j \neq i})$ for some $\sigma_{j \neq i} \in \Pi_{j \neq i} S_j\}$.
- $\leq_i$ is a reflexive, transitive and total preference relation on X.

We study the effect of *previously adopted* intentions on such decision problems, rather than the process by which the agents form these intentions. Furthermore, we restrict our attention to intentions to *realize certain outcomes* in the game, in contrast with intentions to play certain strategy—although there is an obvious connection between the two. We thus assign to each agent $i \in I$ an *intention set* $\iota_i \subseteq X$. The intention set ι_i of agent i should be thought as the intentions that i has formed some time before entering the

*This is an extended abstract of [Roy, 2008, chap.4]. Some results presented here have been obtained in collaboration with Martin van Hees (Rijksuniversiteit Groningen).

game, and on the basis of which he now has to make his decision. Following common assumption by philosophers of action we suppose that $\iota_i \neq \emptyset$, which amounts to agents not having inconsistent intentions. An *intention profile* ι is a vector of intention sets, one for each agent.

Many philosophers of action have stressed that previously adopted intentions *transform* decision problems, a phenomenon which is called the *reasoning-centered commitment* of intentions. They imposes a "standard for *relevance* for options considered in deliberation. And they constrain solutions to these problems, providing a *filter of admissibility* for options." [Bratman, 1987, p.33, emphasis in the original]. These are the two effects of intentions on deliberation that we study in the next sections.

3 Filter of admissibility

We take "providing a filter of admissibility" to mean ruling out options that are incompatible with the agents achieving their intentions. Agents, in that sense, discard some of their strategies because they are incompatible with what they intend. We will study two different admissibility/compatibility criteria, depending on whether the agents take each others' intentions into account.

We start with a generic definition of the discarding process, which we call *cleaning*, and in which the two notions of admissibility will be plugged in. The *cleaned version* $cl(S_i)$ of a strategy set S_i is defined as:

$$cl(S_i) = \{s_i \mid s_i \text{ is admissible for deliberation for } i\}$$

The *cleaned* version of a game $\mathbb{G}$ with intention profile ι is the tuple $cl(\mathbb{G}) = \langle I, X^{cl}, \{cl(S_i), \leq_i^{cl}\}_{i\in I}, \pi^{cl}\rangle$ such that:

- $X^{cl} = \pi(\Pi_{i\in I} cl(S_i)) = \{x \mid x = \pi(\sigma) \text{ for some } \sigma \in \Pi_{i\in I} cl(S_i)\}$.
- $\leq_i^{cl}$ is the restriction of $\leq_i$ to X^{cl}.
- π^{cl} is π with the restricted domain $\Pi_{i\in I} cl(S_i)$.

We do not study intention revision, and so we assume that the agents adapt their intentions to the decision problem they face after cleaning by giving up on achieving the outcomes that are no longer achievable. We thus take the cleaned version ι_i^{cl} of the intention set ι_i to be $\iota_i \cap X^{cl}$, reminding plain belief expansion in e.g. Rott [2001] and Gärdenfors [2003].

The first criterion for admissibility we consider is individualistic: we say that a strategy s_i of agent i is *individualistically admissible* for him when choosing it can yield an outcome he intends. Formally, a strategy s_i of agent i is individualistically admissible with respect to his intention set ι_i when $\pi(s_i) \cap \iota_i \neq \emptyset$. Conversely, a strategy is not admissible for i when choosing it would not realize any of his intentions.

It can be that no strategy survive cleaning with individualistic admissibility, simply because some outcome x can be unrealizable, i.e. it can happen that there is no profile σ such that $\pi(\sigma) = x$. In such case we say that cleaning *empties* a decision problem for the agent. Intuitively agents should avoid intentions which, once used for cleaning, empty the decision problem. This this leaves them no strategy to choose. It is thus important to characterize the intention sets that do not lead to empty cleaned games.

When there is only one agent, cleaning empties a decision problem if and only if ι_i contains no realizable outcomes. In interactive situations, however, agents who clean individualistically can make intentions of others unrealizable. Table 1 is an example, with the numbers in the cells representing which outcomes are in ι_i for the corresponding agent, 1 being the row and 2 being the column player. When

$\mathbb{G}$	t_1	t_2
s_1	1	
s_2	2	

$cl(\mathbb{G})$	t_1
s_1	1

Table 1. A game which an empty cleaning.

more than one agent is involved, to have realizable intentions is thus not enough to avoid ending up with empty strategy sets after cleaning. To pinpoint the conditions which ensure such non-emptiness in the general case, we look at iteration of cleaning, in a way that draws from van Benthem [2003] and Apt [2007].

Given a strategic game $\mathbb{G}$, let $cl^k(\mathbb{G}) = \langle I, X^{cl^k}, \{cl^k(S_i), \leq_i^{cl^k}\}_{i\in I}, \pi^{cl^k}\rangle$ be the strategic game that results after k iterations of the cleaning of $\mathbb{G}$. That is, $cl^1(\mathbb{G}) = cl(\mathbb{G})$ and $cl^{k+1}(\mathbb{G}) = cl(cl^k(\mathbb{G}))$. The smallest cleaning *fixed-point* $cl^{\#}(\mathbb{G})$ of $\mathbb{G}$ is defined as $cl^k(\mathbb{G})$ for the smallest k such that $cl^k(\mathbb{G}) = cl^{k+1}(\mathbb{G})$. In what follows we ignore the "smallest" and only write about the fixed point.

Every game has a unique cleaning fixed point with individualistic cleaning but, as just noted, it may be an empty one. This is avoided only if the intentions of the agents are sufficiently entangled with one another.

Let us call the *cleaning core* of a strategic game $\mathbb{G}$ is the set of strategy profile S^* inductively defined as follows, with $\pi^{S^n}(s_i) = \pi(s_i) \cap \{\pi(\sigma') : \sigma' \in S^n\}$.

- $S^0 = \Pi_{i\in I} S_i$.
- $S^{n+1} = S^n - \{\sigma : \text{ there is an } i \text{ such that } \pi^{S^n}(\sigma(i)) \cap \iota_i = \emptyset\}$.

- $S^* = \bigcap_{n<\omega} S^n$.

For each strategy s_i and profile σ in the cleaning core such that $\sigma(i) = s_i$, there is at least one agent j for whom strategy $\sigma(j)$ is admissible, by looking only at what can result from the profiles in the core.

Fact 3.1 *For any strategic game $\mathbb{G}$ and intention profile ι, $S^* \neq \emptyset$ iff $cl^{\#}(\mathbb{G})$ is not empty.*

From this we learn that the individualistic character of admissibility must be compensated by an interlocking web of intentions and strategies if cleaning is not to make the game empty. Intentions which yield a non-empty cleaning core closely fit the admissible strategies of *all* agents. By intending outcomes that are realizable in the cleaning core, an agent somehow acknowledges that he interacts with other agents who, like him, clean inadmissible options from their strategy set.

The following alternative form of admissibility emphasizes this interactive character. A strategy s_i of agent i is *altruistically admissible* with respect to his intention set ι_i when there is a $j \in I$ such that $\pi(s_i) \cap \iota_j \neq \emptyset$. Following this second criterion, a strategy of agent i is admissible whenever it can yield an outcome that some agent, *not necessarily* i, intends. When agents discard option on the basis of this criterion, there is no risk of emptying the game, and the process does not need to be iterated.

Fact 3.2 *For $\mathbb{G}$ an arbitrary strategic game, $cl^{\#}(\mathbb{G}) = cl(\mathbb{G})$ for cleaning with altruistic admissibility.*

Fact 3.3 *For any strategic game $\mathbb{G}$, intention profile ι and cleaning with altruistic admissibility, there is, for all i, a realizable $x \in \iota_i$ iff $cl^{\#}(\mathbb{G})$ is not empty.*

It is thus crucial for agents to take the others' intentions into account when ruling out options in strategic games. If, on the one hand, agents rule out options without taking care of what the others intend, they run the risk of ending up with no strategy at all, unless their intentions are already attuned to those of their co-players. If, on the other hand, their intentions do not fit so well with those of others, then they should at least take heed of what the others intend when ruling out options. This aspect of the reason-centered commitment of intentions has, up to now, been overlooked in philosophical theories of intentions.

4 Standard of relevance

We now turn to the second aspect of the reason-centered commitment of intentions: transformations of decision problem based on the "standard of relevance".

Here we take this idea to mean discarding options which differences are not relevant in terms of what one intends. We say that such options are *redundant*. Formally, two strategies s_1 and s_2 in S_i are *redundant*, noted $s_1 \approx s_2$, whenever $\pi(s_1, \sigma_{j\neq i}) \in \iota_i$ iff $\pi(s_2, \sigma_{j\neq i}) \in \iota_i$ for all combinations of actions of other agents $\sigma_{j\neq i} \in \Pi_{j\neq i} S_j$. Strategies s_1 and s_2 in Table 2 are redundant for the row player in that sense.

	t_1	t_2	t_3
s_1	$1, 2$	2	1
s_2	1	2	1
s_3		1	2

Table 2. A game with redundant strategies for the row player.

The relation $\approx$ clearly induces a partition of the set of strategies S_i into subsets $[s_i]^{\mathbb{G}}_{\approx} = \{s'_i \in S_i | s'_i \approx s_i\}$, each of which represents a distinct "means" for agent i to achieve what he intends. We take the standard of relevance imposed by intentions to induce such a means-oriented perspective on decision problems.

To make a decision from that perspective agents have to sort out these means according to some preference ordering. Here we assume that they "pick" a representative strategy for each means, and collect them to form their new strategy set. This allows to define preferences in the game that result from this transformation from those in the original game. Regarding the picking process itself, we take an abstract point of view and leave implicit the criterion which underlies it.

Given a strategic game $\mathbb{G}$, a function $\theta_i : \mathcal{P}(S_i) \to S_i$ such that $\theta_i(S) \in S$ for all $S \subseteq S_i$ is called i's *picking function*. A *profile* of picking functions Θ is a combination of such θ_i, one for each agent $i \in I$. These functions return, for each set of strategies—and in particular each equivalence class $[s_i]_{\approx}$—the strategy that the agents picks in that set. We define them over the whole power set of strategies to facilitate the technical analysis.

The *pruned version* $pr(S_i)$ of a strategy set S_i, with respect to an intention set ι_i and a picking function θ_i is defined as:

$$pr(S_i) = \{\theta([s_i]^{\mathbb{G}}_{\approx}) : s_i \in S_i\}$$

Pruned version of a strategic game $\mathbb{G}$ are defined similarly as cleaned ones: given an intention profile ι and a profile of picking function Θ, the pruned version of $\mathbb{G}$ is the tuple $pr(\mathbb{G}) = \langle I, X^{pr}, \{pr(S_i), \leq_i^{pr}\}_{i\in I}, \pi^{pr}\rangle$ such that:

- $X^{pr} = \pi(\Pi_{i\in I} pr(S_i))$.
- $\leq_i^{pr}$ is the restriction of $\leq_i$ to X^{pr}.

- π^{pr} is π with the restricted domain $\Pi_{i \in I} pr(S_i)$.

The pruned version ι_i^{pr} of an intention set ι_i is $\iota_i \cap X^{pr}$. Agents, again, adapt their intentions in the process of pruning.

We once again take a general point of view and analyze iterations of pruning. Given a strategic game $\mathbb{G}$, let $pr^k(\mathbb{G})$ be the strategic game that results after k iterations of the pruning of $\mathbb{G}$. That is, $pr^0(\mathbb{G}) = \mathbb{G}$ and $pr^{k+1}(\mathbb{G}) = pr(pr^k(\mathbb{G}))$. The *pruning fixed point* $pr^{\#}(\mathbb{G})$ of $\mathbb{G}$ is defined as $pr^k(\mathbb{G})$ for the smallest k such that $pr^k(\mathbb{G}) = pr^{k+1}(\mathbb{G})$.

As for cleaning, it can happen that agents end up with empty intentions after a few rounds of pruning, but no pruning makes a game empty.

Fact 4.1 *For all strategic game $\mathbb{G}$ and agent $i \in I$, $pr^{\#}(S_i) \neq \emptyset$.*

Furthermore, the existence of pruning fixed points where all agents have non-empty intentions depends on whether they intend "safe" outcomes. Given a strategic game $\mathbb{G}$, an intention profile ι and a profile of picking functions Θ, the outcome $x = \pi(\sigma)$ is:

- *Safe for pruning at stage 1* iff for all agents i, $\theta_i([\sigma(i)]) = \sigma(i)$.
- *Safe for pruning at stage* $n + 1$ whenever it is safe for pruning at stage n and for all agents i, $\theta_i([\sigma(i)]^{pk^n(\mathbb{G})}) = \sigma(i)$.
- *Safe for pruning* when it is safe for pruning at all stages n.

Safe outcomes are those which the picking functions retain, whatever happens in the process of pruning. Intending safe outcomes is necessary and sufficient for an agent to keep his intention set non-empty in the process of pruning.

Fact 4.2 *For any strategic game $\mathbb{G}$, intention profile ι, profile of picking function Θ and for all $i \in I$, $\iota_i^{pr^{\#}} \neq \emptyset$ iff there is a $\pi(\sigma) \in \iota_i$ safe for pruning in $\mathbb{G}$.*

Agents are thus required to take the others' intentions *and* picking criteria into account if they wish to avoid ending up with empty intentions after pruning. In single-agent cases pruning never makes the intention set of the agent empty, as long as the agent has realizable intentions. This shows, once again, that reasoning-centered commitment really gains an interactive character in situations of strategic interaction.

5 Putting the two transformations together

We now look at how the pruning and cleaning interact with one another, in order to get a more general picture of the reasoning-centered commitment of intentions. We investigate sequential applications of these operations, and consider individualistic admissibility only.

Given a strategic game $\mathbb{G}$, let $t(\mathbb{G})$ be either $pr(\mathbb{G})$ or $cl(\mathbb{G})$. A *sequence of transformation of length k* is any $t^k(\mathbb{G})$ for $k \geq 0$, where $t^1(\mathbb{G}) = t(\mathbb{G})$ and $t^{k+1}(\mathbb{G}) = t(t^k(\mathbb{G}))$. A sequence of transformation $t^k(\mathbb{G})$ is a *transformation fixed point* whenever both $cl(t^k(\mathbb{G})) = t^k(\mathbb{G})$ and $pr(t^k(\mathbb{G})) = t^k(\mathbb{G})$.

The first notable fact about cleaning and pruning sequences is that these operations do not in general commute. Table 3 is a counterexample, with $\theta_2([t_1]) = t_1$. They do commute, however, in the single-agent case.

$\mathbb{G}$	t_1	t_2
s_1		1
s_2	1, 2	1, 2

$pr(\mathbb{G})$	t_1
s_1	
s_2	1, 2

$cl(pr(\mathbb{G}))$	t_1
s_2	1, 2

Table 3. Counter-example to commutativity.

Fact 5.1 $pr(cl(\mathbb{G})) = cl(pr(\mathbb{G}))$ *for any strategic game $\mathbb{G}$ with only one agent, intention set ι_i and picking function θ_i.*

Sequential cleaning and pruning creates new possibilities for empty fixed points. Neither the existence of a cleaning core nor of safe outcomes, and not even a combination of the two criteria are sufficient to ensure non-emptiness. Furthermore, there might not be a unique fixed point, as revealed in Tables 4, 5 and 6, with $\theta_1(\{s_1, s_2\}) = s_2$, $\theta_1(\{s_1, s_2, s_3\}) = s_1$ and $\theta_1(\{s_2, s_3\}) = s_2$.

$\mathbb{G}$	t_1	t_2	t_3
s_1	1	2	
s_2	1, 2		
s_3	1		1

Table 4. A game with two different fixed-points.

Ignoring redundant transformations, all sequences of cleaning and pruning reach a fixed point in a finite number of steps, for every finite strategic games. Non-emptiness of this fixed point is ensured by the following strengthening of safety for pruning and cleaning core. The outcome x of profile $\sigma \in \Pi_{i \in I} S_i$ is:

$cl(\mathbb{G})$	t_1	t_2
s_1	1	2
s_2	1, 2	
s_3	1	

$pr(cl(\mathbb{G}))$	t_1	t_2
s_1	1	2

$cl(pr(cl(\mathbb{G})))$	t_2
s_1	2

Table 5. The route to the first (empty) fixed point of the game in Table 4.

$pr(\mathbb{G})$	t_1	t_2	t_3
s_2	1, 2		
s_3	1		1

$cl(pr(\mathbb{G}))$	t_1
s_2	1, 2
s_3	1

$pr(cl(pr(\mathbb{G})))$	t_2
s_2	1, 2

Table 6. The second fixed point of the game in Table 4.

- *Safe for iterated transformations at stage 1* whenever, for all $i \in I$:
  1. $\pi(\sigma(i)) \cap \iota_i \neq \emptyset$.
  2. $\theta_i[\sigma(i)]^{\mathbb{G}}_{\approx} = \sigma(i)$.
- *Safe for iterated transformations at stage n + 1* whenever it is safe for iterated transformation at stage n and for all $i \in I$:
  1. $\pi^{t^n(\mathbb{G})}(\sigma(i)) \cap \iota_i^{t^n(\mathbb{G})} \neq \emptyset$.
  2. $\theta_i[\sigma(i)]^{t^n(\mathbb{G})}_{\approx} = \sigma(i)$.
- *Safe for iterated transformations* whenever it is safe for transformation at all n.

Fact 5.2 *For any strategic game $\mathbb{G}$, intention profile ι and profile of consistent picking function Θ, if $\pi(\sigma)$ is safe for transformation in $\mathbb{G}$ then for all fixed points $t^{\#}(\mathbb{G})$, $\sigma \in \Pi_i t^{\#}(S_i)$.*

The presence of safe outcomes is thus sufficient not only to ensure that a game has no empty fixed point, but also that all fixed points have a non-empty intersection. Precisely because of that, this does not entail that any game which has no empty fixed point contains safe outcomes. If it can be shown that whenever a game has a non-empty fixed-point then this fixed-point is unique, we would know that safety for transformation exactly captures non-emptiness. Whether this is the case is still open to us at the moment. We do know, however, that the converse of Fact 5.2 holds if we constraint the picking functions.

In the spirit of Sen's [1970] "property α" , let a picking function θ_i be called *consistent* if $\theta_i(X) = s_i$ whenever $\theta_i(Y) = s_i$, $X \subseteq Y$ and $s_i \in X$.

Fact 5.3 *For any strategic game $\mathbb{G}$, intention profile ι and profile of consistent picking function Θ, if $\sigma \in \Pi_i t^{\#}(S_i)$ for all fixed points $t^{\#}(\mathbb{G})$, then $\pi(\sigma)$ is safe for transformation in $\mathbb{G}$.*

If all players intend safe outcomes we thus know that all fixed-point are non-empty, and we can "track" safe outcomes in the agents' original intentions by looking at those they keep intending in all fixed-points.

The existence of empty transformation fixed points and the definition of safety for transformation once again highlight the importance of taking each others' intention into account while simplifying decision problems. The fact that the pruning and cleaning do commute when there is only one agent is in that respect illuminating.

6 Conclusion

We have studied two aspects of the reason-centered commitment of intentions, by extending game-theoretic formalisms with two new operations on strategic form games. We characterized conditions under which these operations keep the games or the intentions of the agents non-empty. This has revealed an important interactive character to the reason-centering commitment, one which went up to now unnoticed in philosophical theories of intentions. This work thus extends game-theoretic models and shew new lights on the theory of intentions.

Taking a epistemic perspective, in the line of Aumann [1999], van Benthem [2003], Brandenburger [2007] and Bonanno [2007], would surely enhance the present work. Mutual knowledge of each others' intentions seems crucial in the process of cleaning and pruning. It would also be interesting to relate the current proposal with game-theoretic work on intention formation and reconsideration, e.g. McClennen [1990] and Gul and Pesendorfer [2001], and with the BDI architectures cited in the introduction.

References

K.R. Apt. The many faces of rationalizability. *The B.E. Journal of Theoretical Economics*, 7(1), 2007. Article 18.

R.J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

G. Bonanno. Two lectures on the epistemic foundations of game theory. URL `http://www.econ.ucdavis.`

edu/faculty/bonanno/wpapers.htm. Delivered at the Royal Netherlands Academy of Arts and Sciences (KNAW), February 8, 2007.

A. Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465–492, 2007.

M. Bratman. *Intention, Plans and Practical Reason*. Harvard University Press, London, 1987.

P. Gärdenfors, editor. *Belief Revision*. Cambridge UP, 2003.

M. Georgeff, B. Pell, M.E. Pollack, M. Tambe, and M. Wooldridge. The belief-desire-intention model of agency. In J. Muller, M. Singh, and A. Rao, editors, *Intelligent Agents V*. Springer, 1998.

F. Gul and W. Pesendorfer. Temptation and self-control. *Econometrica*, 69(6):1403–1435, November 2001.

E.F. McClennen. *Rationality and Dynamic Choice : Foundational Explorations*. Cambridge University Press, 1990.

M.J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, 1994.

H. Rott. *Change, Choice and Inference: A Study of Belief Revision and Nonmonotonic Reasoning*. Oxford Logic Guides. ford University Press, Oxford, 2001.

O. Roy. *Thinking before Acting: Intentions, logic, rational choice*. PhD thesis, Universiteit van Amsterdam, Februry 2008.

A. Sen. *Collective Choice and Social Welfare*. Holden-Day, 1970.

J. van Benthem. Rational dynamic and epistemic logic in games. In S. Vannucci, editor, *Logic, Game Theory and Social Choice III*, pages 19–23. University of Siena, department of political economy, 2003.

W. van der Hoek, W. Jamroga, and M. Wooldrige. Towards a theory of intention revision. *Synthese*, 155, March 2007. Knowledge, Rationality & Action 103-128.

J.D. Velleman. What good is a will? Downloaded from the author's website on April 5th 2006, April 2003.

7 Appendix

7.1 Proof of Fact 3.1

For any strategic game $\mathbb{G}$ and intention profile ι, $S^* \neq \emptyset$ iff $cl^{\#}(\mathbb{G})$ is not empty.

By Definition, $S^* \neq \emptyset$ is the same as saying that we can find a $\sigma \in S^*$ such that for all i, $\pi^{S^*}(\sigma(i)) \cap \iota_i \neq \emptyset$. We show by induction that $\pi(S^k) = X^{cl^k}$, for all k. This is enough to show the equivalence, for then we know that $X^{cl^{\#}} \cap \iota_i \neq \emptyset$, which we know is the same as $cl^{\#}(\mathbb{G})$ being non-empty. The basic case of the induction, $k = 0$, is trivial. For the induction step, assume the claim is proved for k. We have that $x \in \pi(S^{k+1})$ iff there is a $\sigma \in S^{k+1}$ such that $\pi(\sigma) = x$. This in turns happens iff $\pi^{S^k}(\sigma(i)) \cap \iota_i \neq \emptyset$, for all i. But by the inductive hypothesis this is just to say that $\pi(\sigma(i)) \cap X^{cl^k} \cap \iota_i \neq \emptyset$, which is just the definition of x being in X^{x+1}.

7.2 Proof of Fact 3.2

For $\mathbb{G}$ an arbitrary strategic game, $cl^{\#}(\mathbb{G}) = cl(\mathbb{G})$ for cleaning with altruistic admissibility.
We show that $cl(cl(\mathbb{G})) = cl(\mathbb{G})$. Given the definition of the cleaning operation, it is enough to show that $cl(cl(S_i)) = cl(S_i)$ for all i. It should be clear that $cl(cl(S_i)) \subseteq cl(S_i)$. It remains to show the converse. So assume that $s_i \in cl(S_i)$. Since cleaning is done with altruistic admissibility, this means that there is a σ such that $\sigma(i) = s_i$ and a $j \in I$ such that $\pi(\sigma) \in \iota_j$. But then $\sigma(i') \in cl(S_{i'})$ for all $i' \in I$, and so $\sigma \in \Pi_{i \in I} cl(S_i)$. This means that $\pi(\sigma) \in X^{cl}$, which in turns implies that $\pi^{cl}(\sigma) \in \iota_j^{cl}$. We thus know that there is a $\sigma \in \Pi_{i \in I} cl(S_i)$ such that $\sigma(i) = s_i$ and a j such that $\pi^{cl}(\sigma) \in \iota_j^{cl}$, which means that $s_i \in cl(cl(S_i))$.

7.3 Proof of Fact 3.3

For any strategic game $\mathbb{G}$, intention profile ι and cleaning with altruistic admissibility, there is, for all i, a realizable $x \in \iota_i$ iff $cl^{\#}(\mathbb{G})$ is not empty.
There is a realizable $x \in \iota_i$ for all i iff for all i there is a σ such that $\pi(\sigma) \in \iota_i$. But this is this same as to say that for all j there is a strategy s_j such that $\sigma(j) = s_j$ and an i such that $\pi(\sigma) \in \iota_i$ which, by Facts 3.1 and 3.2, means that $cl^{\#}(\mathbb{G})$ is not empty.

7.4 Proof of Fact 4.1

For all strategic game $\mathbb{G}$ and agent $i \in I$, $pr^{\#}(S_i) \neq \emptyset$.
This is shown by induction on $pr^k(\mathbb{G})$. The basic case is trivial. For the induction step, observe that the picking function θ_i is defined for the whole power set of S_i. This means, given the inductive hypothesis, that $\theta_i([s_i]_{\approx}^{pr^k(\mathbb{G})})$ is well-defined and in $[s_i]^{pr^k(\mathbb{G})}$ for any $s_i \in pr^k(S_i)$, which is enough to show that $pr^{k+1}(S_i)$ is also not empty.

7.5 Proof of Fact 4.2

For any strategic game $\mathbb{G}$, intention profile ι, profile of picking function Θ and for all $i \in I$, $\iota_i^{pr^{\#}} \neq \emptyset$ iff there is a $\pi(\sigma) \in \iota_i$ safe for pruning in $\mathbb{G}$.

From right to left. Take any $x \in \iota_i^{pr^{\#}}$. By definition we know that there is a $\sigma \in \Pi_{i \in I} pr^{\#}(S_i)$ such that $\pi(\sigma) = x$. But this happens iff $\sigma \in \Pi_{i \in I} pr^k(S_i)$ for all k, and so that $\theta_i([\sigma(i)]_{\approx}^{pr^k(\mathbb{G})}) = \sigma(i)$ also for all k, which in turns means that x is safe for pruning in $\mathbb{G}$. Left to right, take any such $\pi(\sigma) \in \iota_i$. We show that $\pi(\sigma) \in X^{pr^k}$ for all k. The basic case is trivial, so assume that $\pi(\sigma) \in X^{pr^k}$. We know by definition that $\pi(\sigma)$ is safe for pruning at k, which gives automatically that $\pi(\sigma) \in X^{pr^{k+1}}$.

7.6 Proof of Fact 5.2

For any strategic game $\mathbb{G}$, intention profile ι and profile of consistent picking function Θ, if $\pi(\sigma)$ is safe for transformation in $\mathbb{G}$ then for all fixed points $t^{\#}(\mathbb{G})$, $\sigma \in \Pi_i t^{\#}(S_i)$.

This is shown by induction on k for an arbitrary fixed point $t^k(S_i)$. The proof is a direct application of the definition of safety for transformation.

7.7 Proof of Fact 5.3

For any strategic game $\mathbb{G}$, intention profile ι and profile of consistent picking function Θ, if $\sigma \in \Pi_i t^{\#}(S_i)$ for all fixed points $t^{\#}(\mathbb{G})$, then $\pi(\sigma)$ is safe for transformation in $\mathbb{G}$.

We show by "backward" induction that $\pi(\sigma)$ is safe for transformation at any k for all sequences $t^k(\mathbb{G})$. For the basic case, take k to be the length of the longest, non-redundant fixed point of $\mathbb{G}$. I show that $\pi(\sigma)$ is safe for transformation at stage k for all sequences of that length. Observe that by the choice of k all $t^k(\mathbb{G})$ are fixed points. We thus know by assumption that $\sigma \in \Pi_{i \in I} t^k(S_i)$. But then it must be safe for transformation at stage k. If clause (1) was violated at one of these, say $t'^k(\mathbb{G})$, then we would have $cl(t'^k(\mathbb{G})) \neq t'^k(\mathbb{G})$, against the fact that $t'^k(\mathbb{G})$ is a fixed point. By the same reasoning we know that clause (2) cannot be violated either. Furthermore, by the fact that $t'^{k+1}(\mathbb{G}) = t'^k(\mathbb{G})$, we know that it is safe for transformation at all stages $l > k$.

For the induction step, take any $0 \leq n < k$ and assume that for all sequences $t^{n+1}(\mathbb{G})$ of length $n+1$, $\pi(\sigma)$ is safe for transformation at stage $n+1$. Take any $t^n(\mathbb{G})$. By our induction hypothesis, that $\pi(\sigma)$ is safe for transformation at both $cl(t^n(\mathbb{G}))$ and $pr(t^n(\mathbb{G}))$. This secures clause (2) of the definition of safety for transformation, and also gives us that $\sigma \in \Pi_{i \in I} t^n(S_i)$. Now, because it is safe for transformation in $cl(t^n(\mathbb{G}))$, we know that $\pi^{cl(t^n(\mathbb{G}))}(\sigma(i)) \cap \iota_i^{cl(t^n(\mathbb{G}))} \neq \emptyset$ for all i. But since $\pi^{cl(t^n(\mathbb{G}))}(\sigma(i)) \subseteq \pi^{t^n(\mathbb{G})}(\sigma(i))$, and the same for the intention set, we know that $\pi^{t^n(\mathbb{G})}(\sigma(i)) \cap \iota_i^{t^n(\mathbb{G})} \neq \emptyset$ for all i. For condition (2), we also know that $\theta_i[\sigma(i)]_{\approx}^{cl(t^n(\mathbb{G}))} = \sigma(i)$ for all i from the fact that $\pi(\sigma)$ is safe for transformation at stage $n+1$. By Lemma 7.1 (below) and the assumption that θ_i is consistent for all i, we can conclude that $\theta_i[\sigma(i)]_{\approx}^{t^n(\mathbb{G})} = \sigma(i)$, which completes the proof because we took an arbitrary $t^n(\mathbb{G})$.

Lemma 7.1 *For any game strategic game $\mathbb{G}$ and intention set ι_i and strategy $s_i \in cl(S_i)$, $[s_i]_{\approx}^{\mathbb{G}} \subseteq [s_i]_{\approx}^{cl(\mathbb{G})}$.*

Proof. Take any $s'_i \in [s_i]_{\approx}^{\mathbb{G}}$. Since $s_i \in cl(S_i)$, we know that there is a $\sigma_{j \neq i}$ such that $\pi(s_i, \sigma_{j \neq i}) \in \iota_i$. But because $s'_i \approx s_i$, it must also be that $\pi(s'_i, \sigma_{j \neq i}) \in \iota_i$, and so that $s'_i \in cl(S_i)$. Now, observe that $\{\sigma \in \Pi_{i \in I} cl(S_i) : \sigma(i) = s_i\} \subseteq \{\sigma \in S_i : \sigma(i) = s_i\}$, and the same for s'_i. But then, because $s'_i \approx s_i$, it must also be that $s'_i \in [s_i]_{\approx}^{cl(\mathbb{G})}$. QED