Common Counterfactual Belief of Rationality
Subsumes Superrationality on Symmetric Games

Ghislain Fourny

October 24, 2017

Abstract This paper shows that, for symmetric games in normal form, strategy profiles that satisfy Hofstadter’s Superrationality criterion also satisfy both of Halpern’s and Pass’s criteria under Common Counterfactual Belief of Rationality: minimax-rationalizability and individual rationality.

Keywords Counterfactual dependency, Non-Cooperative Game Theory, Non-Nashian Game Theory, Minimax-Rationalizability, Symmetric Games, Superrationality

1 Introduction

Most of the equilibrium concepts in game theory are Nash equilibria [Nash 1951]. This is because most of them are defined under the assumption that an agent considers the opponent’s strategy to be counterfactually independent of their own strategy. In the case of two players playing a game in normal form, this means that, for example, the row player will hold a column for fixed, and pick the row that provides the best payoff in that column.

Over the last few decades though, there has been a few results that investigate what happens if this assumption does not hold, in other words, for agents that consider that, if they picked a different strategy, the opponent would also pick a different strategy. This is similar to the idea behind the Stackelberg equilibrium [von Stackelberg 1934] where the reaction of the opponent is anticipated and taken into account in the reasoning, but without the time dimension.

The idea of counterfactuals is not new and dates back to 1968 and 1973 with Robert Stalnaker [1968] and David Lewis [1973]. In the same period,
Robert Nozick (1969), and then Martin Gardner (1973) published Newcomb’s problem, which materializes the concept of predicting the actions of a free, rational agent to the actual contents of an opaque box, and points out that a careful distinction between causal relationships and counterfactual dependences is paramount. The two-boxer resolution of Newcomb’s problem is in agreement with the Nash paradigm—the choice of an agent is independent from the past—while the one-boxer resolution would correspond to the alternate, non-Nashian paradigm on which this paper focuses.

Douglas Hofstadter (1983) defined the concept of superrational players in a symmetric game such as the prisoner’s dilemma. The idea is that a superrational player playing against another superrational player will consider that both will pick the same strategy, include this knowledge in her reasoning, and only reason on the diagonal of the normal-form matrix. In other words, a superrational player considers that her actions and the opponent’s actions are counterfactually interdependent. As a consequence, the equilibrium reached will not (always) be a Nash equilibrium. Interestingly, after Douglas Hofstadter submitted his 20-way prisoner’s dilemma challenge to his colleagues, Martin Gardner compared it to Newcomb’s paradox in his explanations, but eventually chose to defect rather than cooperate, “with a sign of regret.”

Jean-Pierre Dupuy (1992) showed a strong relationship between Newcomb’s paradox and the prisoner’s dilemma, confirming Martin Gardner’s intuition, and, conjectured (Dupuy, 2000) that it should be possible to define an equilibrium concept for extensive form games that would implement the idea of counterfactual dependence between an agent’s action and its prediction. Such an equilibrium would have an underlying notion of preemption to solve Grandfather’s paradoxes that necessarily arise in the presence of a timeline, leading players to follow a Kantian imperative to avoid such paradoxes. It was proved in 2004 that this equilibrium exists, is unique and is Pareto optimal. The equilibrium and proofs were made public in 2014 as the Perfect Prediction Equilibrium (PPE) (Fourny et al, 2014).

An alternate course of research for counterfactual reasoning for games in extensive form is pursued by Richard Shiffrin et al (2009). Shiffrin’s approach mainly differs from the PPE in that it disagrees on the relevance of preemption, and as such does not permanently discard outcomes deemed irrational in the middle of the reasoning.

Wolpert and Benford (2013) formally established that both accounts of Newcomb’s problem (one-boxer reasoning, two-boxer reasoning) can be explained with two different modeling of conditional probability distributions (incompatible Bayes nets), and as such, that both resolutions are acceptable.

Based on the work of David Lewis and Robert Stalnaker, Halpern and Pass (2013) defined the counterpart of Rationalizability for games in normal form in which players assume that their strategies are counterfactually interdependent, which they call Common Counterfactual Belief of Rationality (CCBR). Halpern and Pass define two criteria that strategy profiles can satisfy under the CCBR assumption: minimax rationalizability and individual rationality.
In this paper, we establish that, in the case of symmetric strategic games, superrational strategy profiles—we call them Hofstadter equilibria—satisfy both minimax rationalizability and individual rationality. In other words, the work of Joseph Halpern subsumes (is a superset of) Douglas Hofstadter’s Superrationality. This formally establishes a link between CCBR and Superrationality.

2 Background on non-Nashian game theory

Before we come to this result, we give some background on games in normal form, Nash equilibria, Superrationality and CCBR.

2.1 Games in normal form

In game theory, games are typically expressed in two forms: normal form and extensive form. In extensive form, the game is expressed as a tree. At each node, a player picks a child node, and the leaves describe possible outcomes and are labelled with payoffs. On the other hand, in normal form, time plays no role and the payoffs are organized in a matrix—or a tensor of higher dimensionality than two if there are more than two players. Figure 1 shows a two-player game in normal form. One player plays on the row, the other plays on the columns.

**Definition 1 (Game in normal form)**

A game in normal form is defined with:

- a finite set of players $P$;
- a set of strategies $\Sigma_i$ for each player $i \in P$;
- a specification of payoffs $u_i(\sigma)$ for each player $i \in P$ and strategy profile $\sigma = (\sigma_j)_{j \in P}$.

The payoff space only needs to be totally ordered (preference relation). In particular, when numbers are used, they are only meant as ordinals. In other words, comparing 1 to 1000 is no different than comparing 1 to 2.

Very often, given a player $i$, we denote as $\Sigma_{-i}$ the cartesian product of the remaining strategy spaces, and given a strategy profile $\sigma$, we denote as $\sigma_{-i}$ the projection of the profile on $\Sigma_{-i}$. This very conveniently allows writing $u_i(\sigma)$ as $u_i(\sigma_i, \sigma_{-i})$ with a slight abuse of notation that is always clear from the context.

|   | 1             | 2             |
|---|---------------|---------------|
| A | $u_1(A,1), u_2(A,1)$ | $u_1(A,2), u_2(A,2)$ |
| B | $u_1(B,1), u_2(B,1)$ | $u_1(B,2), u_2(B,2)$ |

Fig. 1 A game in normal form, with two players that each can pick two strategies (A and B for the row player, 1 and 2 for the column player)
Fig. 2 The prisoner’s dilemma. A player can either cooperate or defect. If both cooperate, they get more than if both defect. However, a player who unilaterally defects will get even more payoff than with mutual cooperation.

|       | Defect | Cooperate |
|-------|--------|-----------|
| Defect| 1, 1   | 3, 0      |
| Cooperate | 0, 3   | 2, 2      |

Fig. 3 The chicken game. A player can either stay straight or swerve. If both swerve, they get more (aka lose less) than if they both stay straight, and a player who unilaterally goes straight gets more payoff than if both swerve. The difference with the prisoner’s dilemma, however, is that the “betrayed” player has interest in not reciprocating the betrayal (0 and 1 are swapped).

|       | Straight | Swerve |
|-------|----------|--------|
| Straight | 0, 0    | 3, 1   |
| Swerve  | 1, 3     | 2, 2   |

Fig. 4 The coordination game. In this game, the players have a mutual interest to pick the same strategy, even though one of the two strategies is better for both of them (aligned interest).

|       | Sushi | Pizza |
|-------|-------|-------|
| Sushi | 1, 1  | 0, 0  |
| Pizza | 0, 0  | 2, 2  |

Figures 2, 3 and 4 show three famous examples of symmetric games. The payoffs have been normalized to the first natural numbers as cardinality is irrelevant.

2.2 Nash equilibria

Nash equilibria are defined having in mind that players hold their opponent’s choices of strategies as fixed. The definition of a Nash equilibrium naturally arises as a strategy profile $\bar{\sigma}$ for which, for each player, the picked strategy $\sigma_i$ is the best response to the other players’ strategies $\sigma_{-i}$.

Formally:

**Definition 2 (Nash equilibrium)** Given a game $(P, \Sigma, u)$, a strategy profile $\bar{\sigma}$ is a Nash equilibrium if, for any player $i \in P$:

$$\forall \tau_i \in \Sigma_i, u_i(\sigma_i, \sigma_{-i}) \geq u_i(\tau_i, \sigma_{-i})$$

Figures 5, 6 and 7 show the Nash equilibria for our example games.

In contrast to the above Nash paradigm, the essence of Non-Nashian Game Theory, common to much of the work mentioned in this paper (Hofstadter, Fourny, Dupuy, Reiche, Shiffrin, Halpern, Pass), is that strategies are not
optimized given a fixed opponent strategy, but the opponent strategy is rather assumed to be correlated to the strategy of the player at hand.

2.3 Symmetric games

There is a subclass of games that is of particular relevance, because players are interchangeable: symmetric games.

In a symmetric game, the strategy spaces are identical for all players, and the payoffs are defined in such a way that the game is invariant through a permutation of players.

**Definition 3 (symmetric game)** A game is symmetric if

- the strategy spaces are identical

\[ \forall i, j \in P, \Sigma_i = \Sigma_j = \mathcal{Y} \]

- the payoffs are symmetric

\[ \forall \pi \in Sym(P), \forall \sigma \in \Sigma, u_i(\sigma) = u_{\pi(i)}(\sigma_{\pi(.)}) \]

where \( Sym(P) \) is the permutation group on \( P \) and \( \sigma_{\pi(.)} \) denotes \( (\sigma_{\pi(1)}, \sigma_{\pi(2)}, ..., \sigma_{\pi(|P|)}) \).

All games shown on Figures 5, 6, and 7 are symmetric.
2.4 Superrationality

Superrationality was introduced by Douglas Hofstadter in 1983 in a Scientific American column. In his own words:

“If reasoning dictates an answer, then everyone should independently come to that answer. Seeing this fact is itself the critical step in the reasoning toward the correct answer [...]. Once you realize this fact, then it dawns on you that either all rational players will choose D or all rational players will choose C. This is the crux.

Any number of ideal rational thinkers faced with the same situation and undergoing similar throes of reasoning agony will necessarily come up with the identical answer eventually, so long as reasoning alone is the ultimate justification for their conclusion. Otherwise reasoning would be subjective, not objective as arithmetics is. A conclusion reached by reasoning would be a matter of preference, not of necessity.”

It appears straightaway that the above consideration directly puts in question the fundamental assumption behind Nash equilibria that players consider their decisions to be independent from other players’ decisions. Superrational players consider that their reasonings are correlated, not because of any causal effect or any kind of retrocausality, but because their reasonings and conclusions are identical.

The games described in the column are all symmetric, which is a requirement for the reasonings to be identical. Identical reasonings and conclusions mean that only outcomes on the diagonal of the normal form are considered. An equilibrium is reached if among all outcomes of the diagonal, it leads to the highest payoffs (which does not depend on the player as the game is symmetric).

Formally, this is expressed as follows.

**Definition 4 (Hofstadter equilibrium)** Given a symmetric game in normal form, a strategy profile $\sigma$ is an equilibrium reached by Superrational players (a Hofstadter equilibrium) if:

- the strategy profile is on the diagonal:

  $$\exists v \in \mathcal{R}, \sigma = (v, v, ..., v)$$

  which we can also express as

  $$\sigma \in diag(\Sigma)$$

- it maximizes the payoff on the diagonal

  $$\forall \sigma \in diag(\Sigma), \forall i \in P, u_i(\sigma) \geq u_i(\overrightarrow{\sigma})$$

Figures 8, 9 and 10 show the Hofstadter equilibria for our example games.
2.5 CCBR and minimax rationalizability

A classical assumption made in game theory is Common Belief of Rationality: all players are rational, and believe that they all are, and believe that they all believe that they all are, and so on. In other words, rationality is *profactually* assumed.

A framework in which players consider that their opponents’ strategies are correlated to theirs, and who want to anticipate their reactions, needs to extend the assumption of rationality to other possible worlds: belief of rationality must also be *counterfactually* assumed, recursively. This line of thought is common to the work by Hofstadter, Dupuy/Fourny/Reiche, Shiffrin, Halpern/Pass, etc\(^1\). Not only do players believe that they are rational, but they *would also* believe so if they switched strategies. This is key to modelling the reaction of an opponent when considering alternate strategies.

Joseph Halpern and Rafael Pass explicitly name—and formally define—this assumption as Common Counterfactual Belief of Rationality (CCBR): the beliefs of a player would also hold if this player would play a different strategy—they call this a counterfactual belief: all players are rational, and believe that they all are, and each one of them believes that all players would also be rational if they played otherwise, and so on. Building on this terminology, all players are rational, and counterfactually believe that they all are, and counterfactually believe that they all counterfactually believe that they all are, and so on.

---

\(^1\) This is also to be put in relation with the terminology “(common knowledge of) substantive rationality” as discussed between Stalnaker, Aumann, Binmore and Halpern regarding the Backward Induction Paradox.
Fig. 11 A symmetric 3x3 game in normal form, for which strategy C is minimax-dominated by both A and B for both players, and (after eliminating C) B is minimax-dominated by A for both players. The iteration goes from left to right, and eliminated profiles are marked in gray. C and B are not minimax-rationalizable, in other words, are not rational according to CCBR, even if each player considers the strategy of the other player not to be fixed.

Furthermore, CCBR does not assume that the opponents’ strategies would be unchanged if a player unilaterally changed his strategy. Dropping this assumption leaves room for the above assumption on the counterfactuals. This is a bit similar to Stackelberg’s idea, in that a player who changes his strategy would anticipate in his reasoning how other players would adapt. CCBR is thus a major change, away from the Nash paradigm.

Halpern and Pass further define minimax rationalizability to select strategies that make sense under CCBR. One way of characterizing minimax rationalizability is by iterated deletion of strategies that are minimax dominated.

Informally, in a game in normal form, a strategy of player \( i \) is minimax dominated if there exists another strategy that guarantees him a higher payoff no matter what the opposite player does or would do. In other words, there is another strategy for which the minimum payoff is greater than the maximum payoff of the dominated strategy.

The intuition behind this definition is that, even if the player considers that the opponent’s strategy is correlated with his choice, a minimax-dominated strategy will never be a good choice as the payoffs will nevertheless always be less, no matter what the assumed correlation is. If a player P picked a minimax-dominated strategy, even if the opponent’s strategy is the best possible case for P, there is another strategy that would give him a higher payoff even if the opponent’s strategy were the worst possible case.

This is to be put in contrast with “classical” rationalizability, in which strategies which are not best responses are eliminated, which is a weaker requirement for elimination. In other words, a strategy that is minimax-dominated would also be eliminated according to classical rationalizability.

Figure 11 shows an example of game in which some strategies are minimax-dominated and can thus be eliminated under CCBR. The games shown on Figures 2, 3 and 4 have no minimax-dominated strategies.

**Definition 5 (minimax rationalizability):** Given a game \((P, \Sigma, u)\), given a player \( i \), a strategy \( \sigma_i \) is minimax-dominated\(^2\) if

\[^2\text{In the original paper [Halpern and Pass, 2013], the opponent’s strategy is taken from a subset of the opponent’s strategies to account for successive eliminations. We are leaving out this aspect here for pedagogical reasons. Indeed, one can also mentally update \( \Sigma \) in place as strategies get eliminated.}\]
Joseph Halpern and Rafael Pass give alternate characterizations of minimax-rationalizability, but iterated deletion is the most intuitive one, and the one that we will use for our proof.

The concept of minimax-rationalizability, defined on strategies, extends to strategy profiles, i.e., a strategy profile can be considered to be minimax-rationalizable if all the strategies it is made of are minimax-rationalizable.

2.6 CCBR and individual rationality

Halpern and Pass define another concept under the assumption of CCBR: individual rationality of strategy profiles. The concept of individual rationality, unlike that of minimax-rationalizability, is not meant to be standalone, but rather to be a useful tool in some proofs related to games with translucent players\footnote{Offline discussion with Joseph Halpern at the ASIC conference in Interlaken, Switzerland, July 2017.}. Nevertheless, it is worth mentioning because superrationality, as we will see, also relates to individual rationality.

Informally, a strategy profile is individually rational if it Pareto-dominates a virtual strategy profile made of all “best worst payoffs”, that is, each player gets at least what he has the power to guarantee himself by picking the strategy with the highest worst payoff, regardless of what the opponents do. Here again, one recognizes CCBR, because the opponent’s strategy is not fixed, and instead, all possibilities are considered.

**Definition 6 (individually rational strategy profile)** a strategy profile $\sigma$ is individually rational if

$$\forall i \in P, u_i(\sigma) \geq \max_{\tau_i \in \Sigma_i} \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\tau)$$

Minimax-rationalizability and individual rationality, for strategy profiles, are not subsuming each other in any way: an individually rational strategy profile may not survive iterated minimax-deletion, and not all strategy profiles that survive it are individually rational. As Joseph Halpern and Rafael Pass point out, an individually rational strategy profile will always survive the first round of minimax elimination, but may get eliminated in the second. The intuitive reason is that, after a round of elimination, this strategy profile may “lose” its individual rationality because the elimination of some strategies can increase the threshold required for individual rationality.

Figures 12 and 13 show the individually rational strategies for the prisoner’s dilemma and the chicken game. In these examples, some strategy profiles that are minimax rationalizable (they all are) are not individually rational (DC and CD are not, as well as Stay Straight-Stay Straight).
Fig. 12 The prisoner’s dilemma with individually rational outcomes shown in gray. The virtual strategy profile with the best worst payoffs (maximins) is (1, 1). It would not be individually rational to have opposite strategies.

|                  | Defect | Cooperate |
|------------------|--------|-----------|
| Defect           | 1, 1   | 3, 0      |
| Cooperate        | 0, 3   | 2, 2      |

Fig. 13 The chicken game with individually rational outcomes shown in gray. The virtual strategy profile with the best worst payoffs (maximins) is (1, 1). It would not be individually rational for both players to stay straight.

|                  | Straight | Swerve |
|------------------|----------|--------|
| Straight         | 0, 0     | 3, 1   |
| Swerve           | 1, 3     | 2, 2   |

Fig. 14 The coordination game with individually rational outcomes shown in gray. The virtual strategy profile with the best worst payoffs (maximins) is (5, 5). All strategy profiles are individually rational.

|      | Sushi | Pizza |
|------|-------|-------|
| Sushi| 1, 1  | 0, 0  |
| Pizza| 0, 0  | 2, 2  |

Fig. 15 The 3x3 game with individually rational outcomes shown in gray. The virtual strategy profile with the best worst payoffs (maximins) is (5, 5). Indeed, strategy A guarantees for both players a minimum payoff of 5 regardless of what the other does.

|     | A     | B     | C     |
|-----|-------|-------|-------|
| A   | 9, 9  | 8, 6  | 5, 1  |
| B   | 6, 8  | 7, 7  | 4, 2  |
| C   | 1, 5  | 2, 4  | 3, 3  |

Figure 14 shows the coordination game, where minimax rationalizability and individual rationality coincide (all strategy profiles qualify).

Figure 15 shows the individually rational strategy profiles on the example we used for minimax rationalizability. Here, it can be seen that the strategy profiles AB, BA and BB are individually rational, but do not survive two rounds of minimax elimination.

3 CCBR subsumes superrationality

We now come to the main results of this paper. The theorems are straightforward to formulate given our introductory explanations.

**Theorem 1** Given a symmetric game in normal form, a Hofstadter equilibrium is always minimax-rationalizable.

**Theorem 2** Given a symmetric game in normal form, a Hofstadter equilibrium is always individually rational.
We now give the proofs of these two theorems.

*Proof (Theorem 1)*

In minimax rationalizability, the order of elimination is not relevant. Because of symmetry, if a strategy gets eliminated for a player, then it will be eliminated for all players. We reorder eliminations in such a way that strategies get eliminated for all players in batches, so that after each elimination, the game remains symmetric. We can now show that, for a symmetric game, a Hofstadter equilibrium cannot get minimax-eliminated.

Let $\bar{\sigma}$ be a Hofstadter equilibrium. We can write it as $\bar{\sigma} = (\sigma, \sigma, ..., \sigma)$ for some $\sigma \in \mathcal{Y}$. By definition of the maximum:

$$\max_{\tau_{-i} \in \Sigma_{-i}} u_i(\sigma, \tau_{-i}) \geq u_i(\sigma, \sigma, ..., \sigma)$$

because $(\sigma, ..., \sigma)$ is in the set over which the maximum is taken (opponents’ strategies).

Let now $i$ denote a player, and $\upsilon \in \mathcal{Y}$ now be any of its strategies. By definition of a Hofstadter equilibrium, the payoffs are maximal on the diagonal, so that:

$$u_i(\sigma, \sigma, ..., \sigma) \geq u_i(\upsilon, \upsilon, ..., \upsilon)$$

Finally, by definition of the minimum:

$$u_i(\upsilon, \upsilon, ..., \upsilon) \geq \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\upsilon, \tau_{-i})$$

because $(\sigma, ..., \sigma)$ is in the set over which the maximum is taken (opponents’ strategies).

By transitivity, we get:

$$\max_{\tau_{-i} \in \Sigma_{-i}} u_i(\sigma, \tau_{-i}) \geq \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\upsilon, \tau_{-i})$$

which directly contradicts the existence of a strategy that allows minimax-domination, and this holds for any player.

There is one more thing to say for the proof to be complete. After an iteration of deletion of minimax-dominated strategies as described above, a Hofstadter equilibrium remains a Hofstadter equilibrium. This is because eliminating other rows or columns than that of the maximum diagonal payoff does not affect this maximum diagonal payoff. Hence, a Hofstadter equilibrium will recursively survive all iterations and, in the end, satisfy minimax rationalizability. □

*Proof (Theorem 2)*

Let $\bar{\sigma}$ be a Hofstadter equilibrium. We can write it as $\bar{\sigma} = (\sigma, \sigma, ..., \sigma)$ for some $\sigma \in \mathcal{Y}$. Let $i$ denote a player.

By definition of a Hofstadter equilibrium, the payoffs are maximal on the diagonal, so that:
Fig. 16 A diagram depicting the relationship between minimax rationalizability, individual rationality, and Superrationality for strategy profiles.

\[ u_i(\sigma, \sigma, ..., \sigma) \geq \max_{\upsilon \in \Upsilon} u_i(\upsilon, \upsilon, ..., \upsilon) \]

Furthermore, for any strategy \( \upsilon \),

\[ u_i(\upsilon, \upsilon, ..., \upsilon) \geq \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\upsilon, \tau_{-i}) \]

(the minimum payoff on its line can only be smaller than the payoff on the diagonal).

Combining the above inequalities:

\[ u_i(\sigma, \sigma, ..., \sigma) \geq \max_{\upsilon \in \Upsilon} u_i(\upsilon, \upsilon, ..., \upsilon) \geq \max_{\upsilon \in \Upsilon} \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\upsilon, \tau_{-i}) \]

Considering that \( \upsilon \) is a mute variable and that \( \Upsilon = \Sigma_i \) (symmetric game),

\[ u_i(\sigma, \sigma, ..., \sigma) \geq \max_{\tau_i \in \Sigma_i} \min_{\tau_{-i} \in \Sigma_{-i}} u_i(\tau_i, \tau_{-i}) \]

this fulfills the definition of an individually rational outcome. □

Note that both inclusions are strict: there are minimax-rationalizable strategy profiles that are not Hofstadter equilibria (e.g., DC, CD, DD in the prisoner’s dilemma), as well as individually rational strategy profiles that are not Hofstadter equilibria (e.g., Swerve-Stay Straight in the chicken game, which is not even on the diagonal, or Sushi-Sushi in the coordination game, which is not optimal).
4 Conclusion and future work

We have shown that Superrational players also behave in a way that is consistent with Joseph Halpern’s and Rafael Pass’s Common Counterfactual Belief of Rationality. This shows that Halpern’s and Pass’s work manages to capture the behavior described by Douglas Hofstadter. This validates the relevance of his rationalizability criteria for settings that deviate from Nashian game theory by assuming that a switch of strategy would be correlated with a switch of strategy by the opponents. The inclusions between the three concepts are depicted on Figure [16].

Further investigations include finding out whether the Perfect Prediction Equilibrium, as well as Shiffrin’s equilibrium, are also subsumed by CCBR. Interestingly, it is straightforward to show with a counterexample that the PPE (with the game converted to its normal form equivalent) is not always individually rational, because CCBR expressed on the normal form cannot capture preemption. However, we conjecture that the PPE may qualify as minimax rationalizable.

5 Acknowledgement

I am grateful to Joseph Halpern for providing feedback on a first draft of this paper, as well as to Jean-Pierre Dupuy, Stéphane Reiche, Alexei Grinbaum, Bernard Walliser, Richard Shiffrin, Robert French for motivating and challenging discussions on the topic of Non-Nashian Game Theory.

References

Dupuy JP (1992) Two Temporalities, Two Rationalities: A New Look At Newcomb’s Paradox. Economics and Cognitive Science, Elsevier pp 191–220

Dupuy JP (2000) Philosophical Foundations of a New Concept of Equilibrium in the Social Sciences: Projected Equilibrium. Philosophical Studies 100:323–356

Fourny G, Reiche S, Dupuy JP (2014) Perfect Prediction Equilibrium. arXiv

Gardner M (1973) Free Will Revisited, With a Mind-Bending Prediction Paradox by William Newcomb. Scientific American 229

Halpern JY, Pass R (2013) Game theory with translucent players. arXiv URL https://arxiv.org/abs/1308.3778 [1308.3778]

Hofstadter D (1983) Dilemmas for Superrational Thinkers, Leading Up to a Luring Lottery. Scientific American

Lewis D (1973) Counterfactuals. Harvard University Press

Nash J (1951) Non-cooperative Games. Annals of Mathematics 54:286 – 295

Nozick R (1969) Newcomb’s Problem and Two Principles of Choice. Essays in Honor of Carl G Hempel
Shiffrin R, Lee M, Zhang S (2009) Rational Games - ‘Rational’ stable and unique solutions for multiplayer sequential games. Tech. rep., Indiana University

von Stackelberg HF (1934) Marktform und Gleichgewicht. Springer

Stalnaker R (1968) A theory of conditionals. Studies in Logical Theory

Wolpert D, Benford G (2013) The lesson of newcomb’s paradox. Synthese 190(9):1637–1646, DOI 10.1007/s11229-011-9899-3, URL http://dx.doi.org/10.1007/s11229-011-9899-3