Local Reasoning in Dynamic Games

Andrés Perea*and Elias Tsakas†
Maastricht University

Extended abstract
Workshop on "Reasoning about other minds: Logical and cognitive perspectives"

Abstract

In the theory of dynamic games, it is often assumed that players reason about *all* stages in the game. In this paper we relax this assumption, by allowing players to only reason about *some* stages in the game – not all. We take the forward induction concept of *common strong belief in rationality* (Battigalli and Siniscalchi (2002)) and adapt it to this framework. We also present an algorithm that yields precisely those conditional beliefs and strategies that are allowed by the concept we propose.

1 Introduction

Beliefs play a crucial role in the analysis of dynamic games. Indeed, if a player must make a choice at a certain stage in the game, then it is important for him to first form a belief about the opponents' strategy choices. Moreover, if his previous belief has been contradicted by some of the opponents' choices in the past, then it is important to know how this player would *revise* his belief under such circumstances.

In game theory it is often assumed – either implicitly or explicitly – that players in a dynamic game can reason about *all possible stages* in the game when forming their beliefs. Consider, for instance, the forward induction concept of *extensive-form rationalizability* (Pearce (1984)), which has later been given an epistemic characterization in Battigalli and Siniscalchi (2002)) through the notion of *common strong belief in rationality.* The central idea in this concept is that a player, whenever possible, must believe that his opponents are choosing strategies that are optimal for them at *every* stage of the game where they have to move. We say that the player *strongly believes in his opponents' rationality.* By imposing this condition, it is implicitly assumed that a player always reasons about *all* possible stages in the game, as a player must

*Presenter (if accepted) and corresponding author. *Address:* EpiCenter and Dept. of Quantitative Economics *E-mail:* a.perea@maastrichtuniversity.nl
†EpiCenter and Dept. of Economics. *E-mail:* e.tsakas@maastrichtuniversity.nl

always verify whether for every opponent there exists a strategy which (a) is compatible with the history that has been observed so far, and (b) is optimal for that opponent at *every* stage of the game.

The aim of this paper is to *relax* this assumption, by allowing the players in a dynamic game to only reason about *some* stages in the game – not all. Relaxing this assumption seems important for several reasons. First, it may be cognitively too demanding for a player to reason about all stages of the game, especially when the game is large. It may also be that some stages in the game seem more relevant than other stages, and that a player deliberately chooses to only reason about these stages in the game. Finally, there are some reasoning concepts in game theory that require a player to only actively reason about *some* stages of the game – not all. Take, for instance, the backward induction concepts of *subgame perfect equilibrium* (Selten (1965)), *sequential equilibrium* (Kreps and Wilson (1982)) and *common belief in future rationality* (Perea (2014), see also Baltag, Smets and Svesper (2009) and Penta (2009)). These concepts either implicitly or explicitly assume that players are completely *forward looking,* in the sense that they only critically reason about the stages that lie ahead, and not about choices that have been made in the past.

Our formal approach is that we assume, for every player i, and every history h where player i has to move, some collection $F_i(h)$ of stages, representing the stages that player i reasons about when he finds himself at h. We then take Battigalli and Siniscalchi's (2002) notion of *common strong belief in rationality* as a bechmark model, and adapt this concept to the assumption that player i, at h, only reasons about the stages in $F_i(h)$.

More precisely, let F be the "reasoning mapping", which assigns to every player i, and to every history h where player i is active, the collection $F_i(h)$ of histories that player i reasons about when he is at h. We say that player i *strongly believes in the opponents' rationality relative to* F if at every history h where player i is active, player i believes – whenever possible – that his opponents choose rationally at all stages in $F_i(h)$, but not necessarily at stages not in $F_i(h)$. That is, player i always believes – whenever possible – that his opponents choose optimally at those stages he actually reasons about.

If we choose F such that it always includes *all* histories – so if we assume that players always reason about all stages – then this definition reduces to Battigalli and Siniscalchi's (2002) notion of *strong belief in rationality.* If, on the other hand, $F_i(h)$ only includes histories that follow h – that is, if players only reason about future stages at any point in time – then we obtain Perea's (2014) notion of *belief in the opponents' future rationality.* But our definition allows for much more variants, as we are completely flexible in choosing the stages that players reasons about.

Suppose now we do not only require that players strongly believe in the opponents' rationality relative to F, but in addition also impose that players – whenever possible – believe that the opponents themselves also strongly believe in *their* opponents' rationality relative to F. This is called *two-fold strong belief in rationality relative to* F. In a similar vain we can recursively define k-fold strong belief in rationality relative to F, for every k. We say that a player expresses *common strong belief in rationality relative to* F if he expresses k-fold strong belief in rationality relative to F, for every k.

This concept, which is at the heart of our paper, contains forward induction reasoning and backward induction reasoning as special cases. Indeed, if we choose F such that it always contains all histories, then common strong belief in rationality relative to F is equivalent to plain common strong belief in rationality, as defined by Battigalli and Siniscalchi (2002), which is a forward induction concept. If we choose F such that $F_i(h)$ only contains stages that follow h, then the concept reduces to *common belief in future rationality,* as developed in Perea (2014), which is a typical backward induction concept. Here, by "forward induction" we mean a type of reasoning in which a player – whenever possible – tries to find a rational explanation for the choices that his opponents have made in the past. "Backward induction", in contrast, represents a type of reasoning in which a player does not actively reason about the opponents' past choices, but instead focuses completely on the game that lies ahead. But our concept allows for many other interesting cases as well, by choosing different options for the F.

We then proceed by offering an algorithm, which we call the *iterated conditional dominance procedure relative to* F, which delivers as an output precisely those conditional beliefs and those strategy choices that are possible under *common strong belief in rationality relative to F.* The algorithm starts by designing for every history h in the game, and for every player i who is active at h, the *full decision problem* for i at h. This full decision problem consists of (a) the possible strategies that i can choose at h, and (b) the possible strategy combinations that his opponents can choose at h. We refer to (a) as the possible *decisions,* and to (b) as the possible *beliefs* that i can hold at h. In every further step of the algorithm, we keep at every history h only those beliefs that – whenever possible – prescribe non-eliminated decisions for the opponents at all histories in $F_i(h)$. Subsequently, we keep at h only those decisions that are optimal for a non-eliminated belief at h. We keep removing beliefs and decisions in this way until no further eliminations are possible. The algorithm is shown to always terminate within finitely many steps, and is relatively easy to apply.

If we choose F such that it always contain all histories, then the algorithm is equivalent to Shimoji and Watson's (1998) iterated conditional dominance procedure, which characterizes Battigalli and Siniscalchi's concept of common strong belief in rationality. On the other hand, if $F_i(h)$ only includes histories that follow h, then the algorithm reduces to Perea's (2014) *backward dominance procedure,* which characterizes the concept of common belief in future rationality.

References

[1] Baltag, A., Smets, S. and J.A. Zvesper (2009), Keep 'hoping' for rationality: a solution to the backward induction paradox, *Synthese* **169,** 301–333 (*Knowledge, Rationality and Action* 705–737).

[2] Battigalli, P. and M. Siniscalchi (2002), Strong belief and forward induction reasoning, *Journal of Economic Theory* **106,** 356–391.

[3] Kreps, D.M. and R. Wilson (1982), Sequential equilibria, *Econometrica* **50,** 863–94.

[4] Pearce, D.G. (1984), Rationalizable strategic behavior and the problem of perfection, *Econometrica* **52,** 1029–1050.

[5] Penta, A. (2009), Robust dynamic mechanism design, Manuscript, University of Pennsylvania.

[6] Perea, A. (2014), Belief in the opponents' future rationality, *Games and Economic Behavior* **83,** 231–254.

[7] Selten, R. (1965), Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragezeit, *Zeitschrift für die Gesammte Staatswissenschaft* **121**, 301–324, 667–689.

[8] Shimoji, M. and J. Watson (1998), Conditional dominance, rationalizability, and game forms, *Journal of Economic Theory* **83,** 161–195.