Journal of Applied Research in Memory and Cognition 4 (2015) 197–211

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Original Article

Modeling trust dynamics in strategic interaction

Ion Juvina[a,*], Christian Lebiere[b], Cleotilde Gonzalez[c]

[a] Department of Psychology, Wright State University, 3640 Colonel Glenn Hwy, Dayton, OH 45435, USA
[b] Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
[c] Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 17 October 2013
Accepted 30 September 2014
Available online 22 October 2014

Keywords:
Cognitive modeling
Trust dynamics
Social Dilemmas
Strategic interaction
Transfer of learning

ABSTRACT

We present a computational cognitive model that explains transfer of learning across two games of strategic interaction – Prisoner's Dilemma and Chicken. We summarize prior research showing that, when these games are played in sequence, the experience acquired in the first game influences the players' behavior in the second game. The same model accounts for human data in both games. The model explains transfer effects with the aid of a trust mechanism that determines how rewards change depending on the dynamics of the interaction between players. We conclude that factors pertaining to the game or the individual are insufficient to explain the whole range of transfer effects and factors pertaining to the interaction between players should be considered as well.

1. Introduction

Humans have a strong tendency toward cognitive parsimony: they tend to develop cognitive strategies that make only minimal use of the potentially relevant information in the environment (Gigerenzer, Todd, & the ABC Research Group, 1999). In doing so, they do not compromise their ability to adapt and thrive, quite the contrary. For example, a trust heuristic assists us in dealing with the complexities of interpersonal interaction (e.g., Wegwarth & Gigerenzer, 2013). Once we have identified a trustworthy person, we tend to suspend the meticulous analysis of the benefits and risks of cooperating with that person; we just assume (i.e., trust) that he or she will reciprocate in kind. Applying a heuristic (i.e., simple rule) can speed up decision-making and reduce cognitive load, releasing cognitive resources that allow us to adapt to complex and dynamic environments. Forgoing meticulous analysis and relying on simple rules derived from experience is one of the characteristics of intuition (Gigerenzer, 2007). Intuitive decision making can be very effective in handling the complexity and uncertainty of social environments by exploiting evolved capacities and environmental regularities (Hertwig & Hoffrage, 2013). Here we use computational cognitive modeling to investigate how the coupling between simple heuristics, cognitive capacities, and social environments might work in strategic interpersonal interaction. We build on previous research suggesting that cognitive architectures and particularly instance-based learning (IBL) approaches may provide a general explanation of intuitive decision-making (Gonzalez, Ben-Asher, Martin, & Dutt, 2015; Gonzalez, Lerch, & Lebiere, 2003; Thomson, Lebiere, Anderson & Staszewski, 2015).

Games of strategic interaction have successfully been used to model various real-world phenomena. For example, the game Prisoner's Dilemma has extensively been used as a model for real-world conflict and cooperation (Rapoport, Guyer, & Gordon, 1976). These games are often called *social dilemmas* to emphasize their relevance for the real world. There has been a recent tendency toward studying ensembles of games, as most social dilemmas rarely occur in isolation; more often they take place either concurrently or in sequence (Bednar, Chen, Xiao Liu, & Page, 2012). This is particularly true in organizations with complex structures, roles, and processes. For instance, when games are played in sequence (i.e., one after another), an effect known as "spillover of precedent" may occur: a precedent of efficient play in a game can be transferred to the next game (e.g., Knez & Camerer, 2000). We refer here to games that are repeated multiple times; the players acquire extensive experience with one game before they switch to another game. We determine the effect that the first game has on the second one and we refer to this effect as *transfer of learning in games of strategic interaction*. This effect has important practical implications. For example, most organizations employ training exercises to develop cooperation and trust among their employees. The assumption is that what is learned in a very specific, ad-hoc exercise transfers to organizational life once the training is over. Much of expertise is generally of an intuitive nature (Gigerenzer, 2007), which makes it

* Corresponding author. Tel.: +1 9377753519.
E-mail addresses: ion.juvina@wright.edu (I. Juvina), cl@cmu.edu (C. Lebiere), coty@cmu.edu (C. Gonzalez).

http://dx.doi.org/10.1016/j.jarmac.2014.09.004
2211-3681/

inaccessible to conscious thought and thus hard to study with traditional methods like self-reporting. Here, we employ a cognitive architectural approach to analyze the interplay of cognitive processes and interaction dynamics that underlie what appears as gut feelings or intuition.

Research in behavioral game theory attempting to explain what causes transfer of learning in games of strategic interaction can be summarized as follows: (1) Bednar et al. (2012) use the concept of *entropy* or strategic uncertainty to explain when learned behavior is likely to spillover from one game to another. They suggest that prevalent strategies in games with low entropy are more likely to be used in games with high entropy, but not vice versa (Bednar et al., 2012). In other words, individuals develop strategies for easier games and apply them to more complex games. (2) Another explanation says that expecting others to do what they did in the past (and expecting that they will think you will do what you did in the past, etc.) can coordinate expectations about which of many equilibria will happen (Devetag, 2005). In other words, players transfer what they did in the past to the subsequent game. (3) Finally, Knez and Camerer (2000) found that transfer of learning across games strongly depended on the presence of superficial, surface similarity (what they call 'descriptive' similarity) between the two games. When the games differed in (what we call) surface characteristics (e.g., actions were numbered differently in the two games) transfer of learning from one game to another did not occur (see a more detailed discussion in Juvina, Saleem, Martin, Gonzalez, & Lebiere, 2013).

These approaches emphasize factors that pertain to the games (entropy, similarity) or the individuals (expectations). We focus here on factors pertaining to *the interaction* between individuals while not excluding factors related to the game and the individual. We demonstrate that the dynamics of a relational construct – *reciprocal trust* – are key to explaining transfer of learning across games of strategic interaction. Generally, we attempt to bring cognitive-computational and socio-cognitive perspectives into the field of experimental economics, aiming to contribute to theory building and unification.

In the remainder of this paper, we summarize an empirical study on transfer of learning in strategic interaction and present a computational cognitive model as an aid in our attempt to explain the empirical results. We also discuss some of the challenges and opportunities that modeling transfer of learning in strategic interaction brings to the computational cognitive modeling field.

2. Experiment

Only a summary of the experiment is given here; a more detailed description was presented elsewhere (Juvina et al., 2013). We selected two of the most representative games of strategic interaction: Prisoner's Dilemma (PD) and the Chicken Game (CG). They are both mixed-motive non-zero-sum games that are played repeatedly. The individually optimal and the collectively optimal solutions may be different. Players can choose to maximize short- or long-term payoffs by engaging in defection or cooperation and coordinating their choices with each other. These features give these games the strategic dimension that makes them so relevant to real-world situations (Camerer, 2003). What makes PD and CG particularly suitable for this experiment is the presence of theoretically interesting similarities and differences, providing an ideal material for studying transfer of learning. Table 1 presents the payoff matrices of PD and CG that were used in this experiment.

Both PD and CG have two symmetric (win-win and lose-lose) and two asymmetric (win-lose and lose-win) outcomes. Besides these similarities there are significant differences between the two games. The Nash equilibria are [−1,−1] or [1,1][1] in PD and [10,−1] or [−1,10] in CG. The number of rounds was not known in advance, so the participants could not apply backward induction. In CG, either of the asymmetric outcomes is more lucrative in terms of joint payoffs than the [1,1] outcome. This is not the case in PD where an asymmetric outcome [10,−10] is inferior in terms of joint payoffs to the [1,1] outcome. Mutual cooperation in CG can be reached by a strongly optimal strategy (i.e., alternation of [−1,10] and [10,−1]) or a weakly optimal strategy [1,1]. The optimal strategy in PD corresponds to the weakly optimal strategy in CG numerically, while the strongly optimal strategy of alternation in CG shares no surface-level similarities with the optimal strategy in PD. Thus, although mutual cooperation corresponds to different choices in the two games (i.e., surface-level dissimilarity), they share a deep similarity in the sense that mutual cooperation is, in the long run, superior to competition in both games.

Table 1
Payoff matrices of prisoner's dilemma (PD) and chicken game (CG).

PD	A	B	CG	A	B
A	−1,−1	10,−10	A	−10,−10	10,−1
B	−10,10	1,1	B	−1,10	1,1

Studying these two games in a sequential ensemble provides a great opportunity to test the theoretical accounts summarized above. Based on the concept of entropy (Bednar et al., 2012), one would expect transfer of learning to only occur in one direction, that is from PD to CG, because CG has relatively higher entropy (i.e., outcome uncertainty) than PD. According to the "expectation account" (Devetag, 2005), one would predict that the prevalent strategy from the first game would transfer to the second game. For example, if the two players settle in the [1,1] outcome in PD, they will be more likely to settle in the [1,1] outcome in CG as well; if they alternate between the two asymmetric outcomes in CG, they will be more likely to alternate in PD as well. If surface similarities were essential for transfer (Knez & Camerer, 2000), one would only expect the [1,1] outcome to drive transfer, because it is identical in the two games.

In contrast, an account focused on interaction would predict that players learn about each other and transfer that learning across games, regardless of surface dissimilarities between games or the order in which games are played.

In both Prisoner's Dilemma and Chicken, learning must occur not only at an individual level but also at a dyad level. If learning occurs only in one of the players in a dyad, the outcomes may be disastrous for that player, because the best solution also bears the highest risk. For example, if only one player understands that alternating between the two moves is the optimal solution in CG, the outcome for that player can be a sequence of −1 and −10 payoffs. Only if both players understand the value of alternation and are willing to alternate, the result will be a sequence of 10 and −1 payoffs for each player, which in average gives each player a payoff of 4.5 points per round. Thus, the context of interdependence makes unilateral individual learning not only useless but also detrimental. The two players must jointly learn that only a solution that maximizes joint payoff is viable in the long term. However, this solution carries the most risk and thus it is potentially unstable in the long term. To ensure that the optimal solution is maintained from one round to another, there must exist a mechanism that mitigates the risk associated with this solution.[2] It has been suggested that

[1] According to the folk theorem (Friedman, 1971), the [1,1] outcome can be a Nash equilibrium if the game is infinitely repeated against the same opponent.

[2] We do not claim that learning occurs in the two players in the same way or at the same time. It is possible that only one player understands the value of alternating