

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Generalization of Learning in Games of Strategic Interaction

Permalink

<https://escholarship.org/uc/item/01d9w6x3>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN

1069-7977

Authors

Juvina, Ion
Lebiere, Christian
Gonzalez, Cleotide
et al.

Publication Date

2012

Peer reviewed

Generalization of Learning in Games of Strategic Interaction

Ion Juvina (ijuvina@cmu.edu), Christian Lebiere (cl@cmu.edu)

Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave.
Pittsburgh, PA 15213 USA

Cleotilde Gonzalez (coty@cmu.edu), & Muniba Saleem (msaleem@cmu.edu)

Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Ave.
Pittsburgh, PA 15213 USA

Abstract

We present a laboratory study investigating the generalization of learning across two games of strategic interaction. The participants' performance was higher when a game was played after, as compared to before, a different game. We found that the generalization of learning from one game to another was driven by both surface and deep similarities between the two games. We developed a computational cognitive model to investigate mechanisms of generalization. Model development highlighted some of the challenges of cognitive modeling in general and modeling strategic interaction in particular. We found that development of reciprocal trust was a key factor that explained the observed generalization effect.

Keywords: Cognitive modeling; Game theory; Strategic interaction; Generalization of learning.

Introduction and Background

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). There has been a recent tendency toward studying ensembles of games, as most real-world "games" rarely occur in isolation; more often they take place either concurrently or in sequence (Bednar, Chen, Xiao Liu, & Page, in press). For instance, when games are played in sequence, 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 to this effect as *generalization 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 generalizes to organizational life once the training is over.

Research on what factors cause generalization of learning in games of strategic interaction can be summarized as follows: (1) Bednar and colleagues (in press) 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., in press). 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). (3) Finally, Knez and Camerer (2000) found that generalization 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. This result is at odds with what is known from the literature on individual problem solving: generalization of learning is facilitated by our ability to perceive abstract, deep-level similarities, and it can be impeded by the presence of superficial, surface similarities (Holyoak & Thagard, 1995).

In this paper we present an experiment aimed at studying generalization of learning in games of strategic interaction. We want to understand when, why, in which direction, and under what conditions generalization occurs. We also present a computational cognitive model as an aid in our attempt to explain the empirical results and settle any potential inconsistencies in the literature.

The next section introduces the experiment and discusses its results. Then the cognitive model is described and its correspondence with the human data is discussed. The paper ends with a general conclusion.

Experiment

Due to space limitation, only a brief description of the experiment is given here. A more detailed description was presented elsewhere (Juvina, Saleem, Gonzalez, & Lebiere, submitted). 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. Players can choose to maximize short- or long-term payoffs by engaging in cooperation or defection 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 generalization of learning. Table 1 presents the payoff matrices of PD and CG.

Table 1: Payoff matrices of PD and 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

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. In CG, either of the asymmetric outcomes is more effective 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 win-win outcome [1,1]. 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-level similarity in the sense that mutual cooperation is, in the long run, superior to competition in both games. This provides a perfect test for our first hypothesis stating that individuals who have learned how to find an optimal strategy in one game will be more likely to find an optimal strategy in the next game even if those optimal strategies are different across the two games.

In both PD and CG, 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 are 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 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. It has been suggested that trust relations are self-sustaining once they have been developed (Hardin, 2002). In situations where there are benefits to individuals that can only be generated through mutual trust, each individual has an incentive to

maintain the relation. A trust relation develops through gradual risk-taking and reciprocation (Cook, Yamagishi, Cheshire, Cooper, Matsuda, & Mashima, 2005). In turn, as trust develops, risk is reduced and the trust relation becomes more stable. Our second hypothesis states that participants develop reciprocal trust throughout the first game, which facilitates learning of the optimal solution in the second game.

Participants and Design

One hundred and twenty participants were paired with anonymous partners (leading to 60 pairs) and were asked to play the two games in sequence. The 60 pairs were randomly assigned to two conditions defined by the order in which the games were played: PD-CG and CG-PD. Participants played 200 unnumbered rounds of each game. At the end of each game, participants completed a five-item questionnaire assessing: how trustful they were of the opponent; how trustful of them the opponent was; how fair they thought the opponent's actions were; how fair the participants' actions were towards their opponents; and how satisfied they were with the overall outcome of the game.

Results¹ and Discussion

To study generalization of learning across the two games, we analyzed the outcomes of a game according to when it was played. We also analyzed the round-by-round dynamics of these outcomes. The statistical significance of the observed effects was tested with the aid of Linear Mixed Effects analysis (*lmer* analysis from the LME4 package in R). This analysis was preferred instead of the classical analysis of variance (ANOVA) because the data violated the assumption of normality.

Similarities and differences The frequencies of the most relevant outcomes (i.e., the two symmetric ones and an alternation of the two asymmetric ones) are displayed in Figure 1 on a round-by-round basis. The first thing to notice is how different the two games are from each other from a behavioral perspective: the [1,1] outcome increases in PD but decreases in CG; alternation is prominent in CG but almost nonexistent in PD; and the mutually destructive outcome ([-1,-1] in PD and [-10,-10] in CG) is more frequent in PD than in CG. However, in spite of these apparent differences, the two games are similar in the sense that mutual cooperation emerges as the preferred solution and it becomes more and more stable over time. These patterns are in line with previous findings (e.g., Rapoport et al., 1976). Given this deep-level similarity, we expect players to be able to generalize their learning of the optimal strategy across the two games, although surface similarities might impede this process (Holyoak & Thagard, 1995). Since we ran the games in both orders (i.e., PD-CG and CG-

¹ Only a summary of the results is provided here as a context for understanding the cognitive model. A more detailed presentation of the results was given in Juvina et al., submitted.

PD), we can also test whether generalization occurs only in one direction, from low to high entropy, as suggested by Bednar and colleagues (in press).

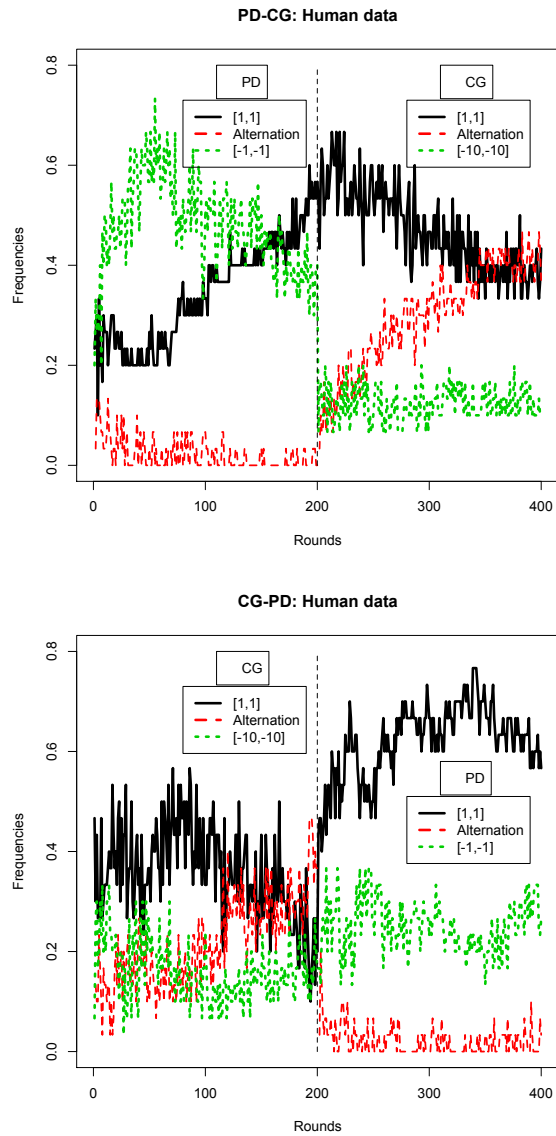


Figure 1: Frequencies of the most relevant outcomes in PD and CG by order (PD-CG on top and CG-PD on bottom) and round averaged across all the human participants.

Generalization Driven by Surface Similarities If learning across games is driven by surface similarities, one would expect the strategy that is learned in the first game to be applied in the second game as well, even though it may not be appropriate for the second game. This is indeed the case with regard to the [1,1] outcome in the PD-CG order: players learn that [1,1] is long-term optimal in PD and they are more likely to achieve it in the subsequent CG, even though it is only weakly optimal in CG. A LME model with occurrence of [1,1] as a dependent variable (binomial distribution), order, round, and their interaction as fixed factors, and participant as a random factor was used to test

the observed effects. There was a main effect of order ($z = 2.21, p = 0.027$) and a main effect of round ($z = -8.171, p < 0.001$); the interaction between order and round was also significant ($z = -7.196, p < 0.001$) indicating that the main effect of order is larger at the beginning of the game and it progressively becomes smaller.

In the CG-PD order, if generalization of learning across games were driven by surface similarities, one would expect the strategy of alternating between the two asymmetrical outcomes to be attempted in the second game as well, at least in the beginning of the game. The main effect of order was non-significant ($z = 1.476, p > 0.10$), suggesting that the strongly optimal strategy in CG (alternation) was not transferred as such (based on surface similarities) to PD. There remains the possibility that the [1,1] outcome was transferred as such from CG to PD. Even though the [1,1] outcome is only weakly optimal in CG, it was selected with relatively high frequency (see Figure 1) and it might have been considered optimal by some participants. We will revisit this point in the section on combined effects of surface and deep-level similarities.

Generalization Driven by Deep-Level Similarities

If learning across games was driven by deep-level similarities, one would expect learning the optimal strategy in the first game to increase the probability of learning the optimal strategy in the second game, even though there is no surface similarity between these strategies. These strategies ([1,1] in PD and alternation in CG) are similar only on an abstract, deep level: they both aim at maximizing joint payoff in a sustainable way, which in these two games is realistically possible only if the two players make (approximately) equal payoffs on a long run. On a surface level, these two strategies are very different. The [1,1] strategy in PD requires that players make the same move at each trial and they do not switch to the opposite move. In contrast, the alternation strategy in CG requires that players make opposite moves at each round and they continuously switch between the two moves. A LME model with occurrence of the alternation outcome in CG as a dependent variable, order, round and their interaction as fixed factors, and participant as a random factor revealed a main effect of order ($z = -2.014, p = 0.044$) indicating a higher level of alternation when CG was played after PD, a main effect of round ($z = 16.205, p < 0.001$) indicating that more and more pairs of participants discovered the alternation strategy as the game unfolded, and a significant interaction between order and round ($z = 8.5, p < 0.001$) indicating that the optimal strategy was learned faster when CG was played second. The same analysis was conducted for the occurrence of the [1,1] outcome in PD and it revealed a main effect of order ($z = -4.340, p < 0.001$) indicating that more pairs of participants discovered the optimal strategy in PD when it was played after CG, a main effect of round ($z = 10.149, p < 0.001$) indicating that more and more pairs of participants found the optimal strategy as the game unfolded, and a significant interaction between order and round ($z = 12.689,$

$p < 0.001$) indicating that the optimal strategy reached a ceiling when PD was played after CG, whereas it increased continuously when PD was played before CG. These results supported our first hypothesis. Specifically, learning the optimal strategy in the first game increased the probability of learning the optimal strategy in the second game, even though the optimal strategies were different in the two games. This generalization effect was significant in both directions (PD-CG and CG-PD) suggesting that entropy (Bednar et al., in press) has little explanatory relevance. If entropy were the causing factor, generalization would have only occurred in one direction.

Combined Effects of Surface and Deep Similarities In the case of deep-level generalization, the main effect of order was smaller in magnitude for CG ($z = -2.014, p = 0.044$) than for PD ($z = -4.340, p < 0.001$). It seems as if CG has a stronger impact on PD than vice versa. A possible explanation for this difference is based on how surface and deep-level similarities combine with each other to drive generalization of learning across games. They may have congruent or incongruent effects. Thus, in the PD-CG order, surface and deep-level similarities act in a divergent, incongruent way: surface similarity makes it more likely that the [1,1] outcome is selected whereas deep-level similarities make it more likely that the alternation outcome is selected. In other words, generalization based on surface similarity interferes with generalization based on deep-level similarity. In contrast, in the CG-PD order, both types of similarities act in a convergent, congruent way: they both increase the probability that the [1,1] outcome is selected. There is no impeding effect of surface similarity on PD because there is no optimal strategy in CG that is similar enough to a non-optimal or sub-optimal strategy in PD. The impeding and/or enabling effects of surface similarities on deep-level generalization are revisited in the modeling section.

Reciprocal Trust In addition to game choices, we analyzed the debriefing questionnaires that were administered at the end of each game. Since the answers to these questions were highly correlated with each other for any one individual participant, we averaged them in one composite variable that we call Reciprocal Trust. Since the debriefing questions were administered twice (at the end of each game) we refer to them as T1 and T2. We calculated correlations between these two trust variables and the variables indicating mutual cooperation in the two games. Spearman's ρ coefficient was used for correlations because the data failed to meet the normality assumption. We found that the more frequent mutual cooperation was in the first game the more likely the players were to rate each other as trustworthy at T1 ($r = 0.75, p < 0.001$ for PD and $r = 0.42, p < 0.001$ for CG). In addition, the more trustworthy players rated each other at T1, the more likely they were to enact mutual cooperation in the second game ($r = 0.28, p = 0.03$ for CG and $r = 0.47, p < 0.001$ for PD). Finally, mutual cooperation in the second

game predicted high levels of trust at T2 ($r = 0.67, p < 0.001$ for CG and $r = 0.88, p < 0.001$ for PD). As expected, the level of reciprocal trust increased from T1 to T2 (mean_{T1} = 11.8, mean_{T2} = 14.1, $t = -3.247, p = 0.001$). These correlations between trust and the frequency of mutual cooperation corroborate our second hypothesis. They suggest that generalization of learning driven by deep-level similarity is facilitated by development and maintenance of reciprocal trust. This finding will be essential for model development.

A cognitive model of generalization of learning

Modeling generalization of learning across games of strategic interaction provides an opportunity to address some of the ongoing challenges of computational cognitive modeling. Three of these challenges are particularly relevant here and are described below as the model is introduced. The model is developed in ACT-R and it will be made freely available to the public on the ACT-R website².

Interdependence

In games of strategic interaction, players are aware of each other and their interdependence. In a previous study we showed that game outcomes were influenced by players' awareness of interdependence. In PD, the more information the two players in a dyad had about each other's options and payoffs the more likely they were to establish and maintain mutual cooperation (Martin, Gonzalez, Juvina, & Lebiere, submitted). Consequently, a cognitive model playing against another cognitive model in a simultaneous choice paradigm needs to develop an adequate representation of the opponent. We use instance-based learning (IBL) and sequence learning (SL) (Gonzalez, Lerch, & Lebiere, 2003) to ensure that the opponent is dynamically represented as the game unfolds. Specifically, at each round in the game an instance (snapshot of the current situation) is saved in memory. The instance contains the previous moves of the two players and the opponent's current move. Saved instances are used to develop contextualized expectations about the opponent's moves based on ACT-R's memory storage and retrieval mechanisms (Anderson, 2007). Expectations can explain some of the spillovers across games (Devetag, 2005).

Generality

Before one attempts to build a model of generalization of learning across two games, one needs to have a model that is able to account for the human data in both games. Although by and large cognitive models are task-specific, there is a growing need to develop more general, task-independent models and there are a few precedents: Lebiere, Wallach, and West (2000) developed a model of PD that was able to account for human behavior in a number of other 2X2 games; and Salvucci (under revision) developed a "supermodel" that accounts for human data in seven

² <http://act-r.psy.cmu.edu/>

different tasks. We build upon these precedents of generality by developing a single model to account for round-by-round human data in both PD and CG. We achieve this generality by using instance-based learning for opponent modeling (as described in the previous section) and reinforcement learning for action selection. Both instance-based learning and reinforcement learning are very general learning mechanisms that can produce different results depending on their input. Specifically, at each round in the game, the model predicts the opponent's move based on the opponent's past behavior and selects its own move based on the utilities of its own past moves in the current context. The input that the model receives as it plays determines the model's behavior. The input is represented by opponent's move, own move, and the payoffs associated with these moves.

An important question is what constitutes the reward from which the model learns the utilities of its actions (moves). Players may try to maximize their own payoff, the opponent's payoff, the sum of the two player's payoffs, the difference, etc. Thus, a large number of reward structures can be imagined. A complicating assumption is that the reward structure might change as the game unfolds depending on the dynamics of the interaction between the two players. This indeed seems to be the case here, as we have realized after a large number of model explorations: no single preset reward structure is sufficient to account for the human data. One could try to computationally explore the space of all possible reward structures and their combinations to find the one that best fit the human data, but the value of this approach is questionable, because it may lead to a theoretically opaque solution. Instead, we chose to employ a theoretically guided exploration that drastically reduces the number of possible reward structures and, more importantly, gives us a principled way to describe the dynamics of players' motives as the game unfolds (see its description in the next section).

Generalization of learning

When the model relies only on the two learning mechanisms described above (i.e., instance-based learning and reinforcement learning) it is able to only account for the generalization driven by surface similarities. Thus, the opponent is expected to make the same move in a given context as in the previous game. Similarly, an action that has been rewarded in the first game tends to be selected more often in the second game. It is impossible in this framework to account for generalization driven by deep-level similarities. For example, if the opponent used to repeat move B when it was reciprocated in PD, there is no reason to switch to alternation between A and B when none of these moves are reciprocated in CG. Moreover, learning within a game may in fact hinder generalization of learning across games if surface similarities are incongruent with the optimal solution in the target game. To find a solution to the deep generalization problem, we need to return to a theoretical and empirical analysis of the two games.

As mentioned in the introduction, in both PD and CG the long-term optimal solution bears the highest risk and, thus, it is unstable in the absence of reciprocal trust. We indeed found that self-reported trust increases after game playing and it positively correlates with the optimal outcome. It may well be that trust explains the deep-level generalization of learning across games. Players learn to trust each other and this affects their reward structure.

Recent literature on trust (e.g., Castelfranchi & Falcone, 2010) suggests that trust is essentially a mechanism that mitigates risk and develops through risk-taking and reciprocity. Inspired by this literature, we added a "trust accumulator" to our model – a variable that increases when the opponent makes a cooperative (risky) move and decreases when the opponent makes a competitive move. In addition, a variable called "willingness to invest in trust" was necessary to overcome situations in which both players strongly distrust each other and persist in a mutually destructive outcome, which further erodes their reciprocal trust, and so on. In such situations, the empirical data shows that players make attempts to develop trust by gradual risk-taking. When these attempts are reciprocated, trust starts to re-develop. In the absence of reciprocation these attempts are discontinued. The willingness to invest in trust increases with each mutually destructive outcome and decreases with each attempt to cooperate that is not reciprocated.

The variables "trust accumulator" and "willingness to invest in trust" are used to determine the dynamics of the reward structure. They both start at zero. When they both are zero or negative, the two players act selfishly by trying to maximize the difference between their own payoff and the opponent's payoff. This quickly leads to the mutually destructive outcome, which decreases trust but increases the willingness to invest in trust. When the latter is positive, a player acts selflessly, trying to maximize the opponent's payoff. This can lead to mutual cooperation and development of trust or players may relapse into mutual destruction. When the trust accumulator is positive, a player tries to maximize joint payoff and avoid exploitation. Thus, the model switches between three reward functions depending on the dynamics of trust between the two players. This mechanism provides a principled solution to the problem of selecting the right reward structure and in the same time solves the generalization problem: due to accumulation of trust in the first game, the model employs a reward structure that is conducive to the optimal solution and thus better performance in the second game.

Modeling results

A cognitive model incorporating the principles described above was developed and fit to the human data presented in the previous section. Fitting the model to the human data was done manually by varying a number of parameters (of which some are standard in the ACT-R architecture and others were introduced as part of the trust mechanism) and trying to increase correlation (r) and decrease root mean square deviation (RMSD) between model and human data.

The results of the current best model ($r = 0.89$, RMSD = 0.09) are presented in Figure 2.

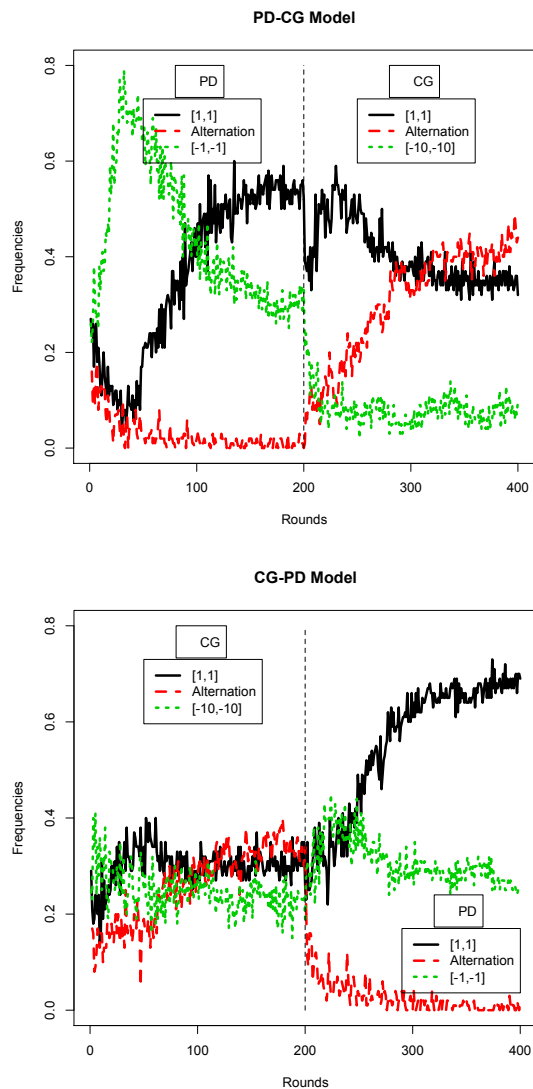


Figure 2: Results of model simulation.

Overall, the model matches the trends in the human data reasonably well (compare to Figure 1). More importantly, the generalization effects are also accounted for.

Discussion and Conclusion

We found that generalization of learning across two games of strategic interaction is driven by deep-level similarities between the two games. Surface similarities may facilitate or hinder generalization depending on whether they are congruent or incongruent with the optimal solution. We used one cognitive model to account for human data in both games. This model helped to explain the observed generalization effect: reciprocal trust was necessary to mitigate the risk associated with the long-term optimal solution. We can conclude that some of the factors suggested in the literature are not necessary (entropy,

cognitive load) or insufficient (expectations, surface similarities), while others are essential (deep-level similarity and reciprocal trust) for generalization of learning in games of strategic interaction.

Acknowledgments

This research was supported by the Defense Threat Reduction Agency (DTRA) grant number: HDTRA1-09-1-0053 to Cleotilde Gonzalez and Christian Lebiere.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Bednar, J., Chen, Y., Xiao Liu, T., & Page, S.E. (in press). Behavioral Spillovers and Cognitive Load in Multiple Games: An Experimental Study. *Games and Economic Behavior*.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, New Jersey: Princeton University Press.

Castelfranchi, C., & Falcone, R. (2010). *Trust Theory: A Socio-Cognitive and Computational Model*. John Wiley and Sons.

Cook, K. S., Yamagishi, T., Cheshire, C., Cooper, R., Matsuda, M., & Mashima, R. (2005). Trust Building via Risk Taking: A Cross-Societal Experiment. *Social Psychology Quarterly*, 68(2), 121-142.

Devetag, G. (2005). Precedent transfer in coordination games: An experiment. *Economics Letters*, 89, 227-232.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, 27, 591-635.

Hardin, R. (2002). *Trust and Trustworthiness*. New York: Russell Sage Foundation.

Holyoak, K. J., & Thagard, P. R. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Juvina, I., Saleem, M., Gonzalez, C., & Lebiere, C. (submitted). Generalization of learning in conflict games: Effects of surface and deep level similarities. *Organizational Behavior and Human Decision Processes*.

Knez, M., & Camerer, C. (2000). Increasing Cooperation in Prisoner's Dilemmas by Establishing a Precedent of Efficiency in Coordination Games. *Organizational Behavior and Human Decision Processes*, 82, 194-216.

Lebiere, C., Wallach, D., & West, R. L. (2000). *A memory-based account of the prisoner's dilemma and other 2x2 games*. Paper presented at the International Conference on Cognitive Modeling.

Martin, J. M., Gonzalez, C., Juvina, I., Lebiere, C. (submitted). Interdependence information and its effects on cooperation.

Rapoport, A., Guyer, M.J., & Gordon, D.G. (1976). *The 2X2 game*. Ann Arbor, MI: The University of Michigan Press.

Salvucci, D.D. (under revision). Integration and reuse in cognitive skill acquisition. *Cognitive Science*.