Modeling Strategic Dynamics Under Alternative Information Conditions

Alessandro Oltramari (aoltrama@andrew.cmu.edu)[1]
Christian Lebiere (cl@cmu.edu)[1]
Noam Ben-Asher (noamba@cmu.edu)[2]
Ion Juvina (ion.juvina@wright.edu)[3]
Cleotilde Gonzalez (coty@cmu.edu)[2]

[1] Functional Modeling Laboratory, Department of Psychology, Carnegie Mellon University
Pittsburgh, PA 15213 USA

[2] Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University
Pittsburgh, PA 15213 USA

[3]Department of Psychology, Wright State University
Dayton, OH 45435 USA

Abstract

In this paper we present a computational cognitive model of two-player interaction under different information conditions. In particular, the model aims at reproducing human behavior in a laboratory experiment based on repeated "Chicken Game". This research shows that information availability deeply influences the strategies and cognitive mechanisms chosen by players and the resulting outcomes. In general, we claim that alternative information conditions can be modeled in ACT-R by adopting a compositional approach.

Keywords: cognitive modeling; game theory, strategic interaction.

Introduction

Cooperative and conflicting phenomena have been comprehensively studied using game theory (Rapoport, Guyer and Gordon 1976), where complex social dynamics are narrowed down to relatively simplified frameworks of strategic interaction. Valid models of real-world phenomena can provide better understanding of the underlying socio-cognitive variables that influence strategic interaction. However, these models should be consistent with the structural characteristics of games, and with the actual everyday situations in hand.

In general, humans gradually *learn* how to play a game, i.e. exploring different moves and their interdependence with actual payoffs, eventually building a mental representation of the opponent's decisions, and harmonizing their choices to the optimal strategy. Different theories have been proposed in the literature regarding the way learning occurs, depending on the settings of the game (Camerer 2003)[1]. In this study, the key aspects of this learning process are manifold: from basic *reinforcement* approaches, where previous payoffs are used by a player to update her strategy, to *sophistication* methods, where players use information about the opponent's payoffs to predict future outcomes. Despite the intrinsic dissimilarities, theories of learning mainly stem from the information available to players, e.g. payoff matrix, mutual moves, previous decisions, beliefs, etc. (Ben-Asher, Dutt & Gonzalez, 2013; Gonzalez and Martin 2011; Gonzalez, et al, 2013). Thus, in order to account for the practical implications of a game, it is important to focus on discrete information conditions: accordingly, in this paper we present a computational cognitive model of game interaction. In particular, our model aims at reproducing the human behavior observed in a laboratory experiment on repeated "Chicken Game" (CG), as introduced in (Russell 1959): the behavioral data indicate that information availability deeply influences which strategies players embrace during the game and the payoffs they gain accordingly. We show that variability in the availability of information can be suitably modeled in ACT-R (Anderson 2007) by marshaling different components of the architecture. In general, cognitive architectures attempt to capture at the computational level the invariant *mechanisms* of human cognition, including those underlying functions of control, learning, memory, adaptivity, perception and action: our analysis of strategic dynamics in CG isolates the contribution of each mechanism to the overall performance. It also emphasizes the need to integrate further factors in the ACT-R architecture, such as emotion-driven mechanisms for decision-making, similarly to what Dancy and Ritter (2012) have recently proposed.

After an overview of the behavioral aspects of repeated CG, we introduce the architectural principles and the design methods adopted in our ACT-R model. Finally, we present and discuss the outcomes of a comprehensive simulation, comparing model and human strategic interaction under different levels of information availability, across multiple trials and conditions.

[1] Some approaches question the validity of learning *per se*: in evolutionary theories, for example strategies are assumed to be innate and constant over time.

Behavioral Aspects

CG represents an interesting framework for studying the dynamics of strategic interaction in situations where two entities (e.g., humans, organizations, agents) are competing over a single limited resource (e.g., energy, market share, etc.). However, as in many real-world situations, collaboration by sharing the resource and taking turns when using it can be beneficial to all the competing entities. This competition can be modeled in a fairly simple scenario, in which two cars travel fast towards one another from opposite directions in a single-lane road: each driver has to choose between driving straight towards a possible collision (i.e., Dare) or turning the steering wheel (i.e., Swerve) and avoid the crash. In a non-zero sum game like CG (see Table 1 for a complete payoff matrix) the best result is for the player that chooses the most hazardous move (Dare) while the opponent simultaneously chooses to avoid risk (Swerve). In a one-shot CG, the outcome is best for a player that Dare while the other player Swerve [10,-1]; the second-best for each is if both Swerve [1,1]; the third-best is for a player that Swerve while the other player Dare [-1,10]; and the worst for each is if both Dare [-10,-10], since the outcome becomes mutually destructive.

When shifting from one-shot CG to repeated CG, the dynamics of strategic interaction scales up in complexity. In particular, the strategy of alternating asymmetric moves clearly leads the players to achieve maximal joint payoff: more specifically, we talk about "mutual alternation", namely a pattern "Swerve-Dare->Dare-Swerve" (or its inverse Dare-Swerve->Swerve-Dare) where both drivers subsequently get the highest payoff [10] and an acceptable loss [-1]. As a matter of fact, to steadily embrace this long-term strategy and prefer it over a selfish behavior (which is more efficient at short term), a player needs to build on her experience, comparing the outcomes of her current moves with the outcomes of her previous moves, gauging the opponent strategy and understanding how the payoff matrix impacts on the decision-making activity. In a laboratory experiment reported in (Ben-Asher, Lebiere, Oltramari & Gonzalez 2013), we referred to this ideal situation where the players have full knowledge of the payoff matrix beforehand as *Descriptive condition* (the standard condition in experimental game theory). When a player doesn't have explicit access to the complete payoff matrix, though she can still rely on game history to recall her own previous moves and payoffs as well as the opponent's, we talk about *Experiential condition*. When the information about opponent's payoffs and moves is removed and a player is only aware that the opponent is a human: we refer to this situation as *Minimal condition*. Finally, when a player is not informed that she is interacting with another player, we describe it as the *Individual condition,* where the selection of an action in each round is perceived as a binary choice between two moves with probabilistic payoff (as if the trials were independent and making projections of the next outcome is impossible). Summarizing the results of the above-mentioned experiment, we notice that under full-information conditions (Descriptive and Experiential), the players tend to cooperate most of the time and prefer to mutually alternate in order to maximize their gains. On the contrary, when there is no awareness of the opponent's actions and outcomes, as in the Individual and Minimal conditions, asymmetric and mutually destructive outcomes soar at the expense of cooperation and fairness[2]. In the next sections we illustrate how an ACT-R cognitive model of repeated CG can account for these behavioral differences. First of all, let us examine what we generally mean by emphasizing the "compositional" features of ACT-R and in which sense "compositionality" differs from "modularity".

Table 1. Chicken Game payoff matrix. The cells show a pair of outcomes (x, y) where x is the payoff to Player 1 and y is the payoff to Player 2[3].

		Player 2 Action	
		A(Dare)	B (Swerve)
Player 1 Action	A (Dare)	-10, -10	10, -1
	B (Swerve)	-1, 10	1, 1

Compositionality in ACT-R

The ACT-R cognitive architecture distinguishes 'declarative knowledge' from 'procedural knowledge', the latter being conceived as a set of procedures (production rules) which coordinate information processing between its various modules: hence, agents accomplish their goals on the basis of knowledge elaborated through procedural steps. Subsymbolic mechanisms, driven by environmental stimuli and based on activation and utility, control both declarative information retrieval and procedural rule selection (Anderson & Lebiere 1998). ACT-R modular organization reflects the brain at a functional level (Anderson 2007): as the basal ganglia are considered to play a central role in selecting specific cognitive activities and enabling pattern recognition, the procedural module has been designed to bridge and synchronize the other modules (represented in Figure 1) through limited capacity buffers (Anderson et al. 2004). But modularity characterizes ACT-R far beyond the macro-level, as it "vertically" spans a wide range of narrower cognitive mechanisms: we refer to this phenomenon as "compositionality" of ACT-R cognitive mechanisms rather than "modularity", which is more exclusively correlated to the architectural level.

[2] The data set of the experiment includes 240 participants who were randomly paired to play 200 unnumbered rounds of repeated CG with payoffs as seen in Table 1. Participants were randomly assigned to one of the four information conditions. These conditions were modeled after the layers of the Hierarchy of Social Information (HSI) outlined by Gonzalez and Martin (2011).

[3] As far as the payoffs are kept consistent to the structure of the game, alternative values might also be used in the matrix.

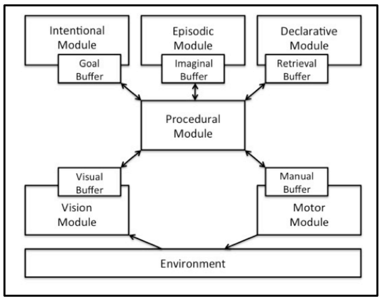

Figure 1. ACT-R Modular Structures.

As we pointed out in the introduction, the complexity of strategic interactions led us to explore cognitive mechanisms of decision-making. In particular, our model employs *reinforcement learning* (RL), *instance-based learning* (IBL) and *trust accumulation* (TA): each of these mechanisms can be regarded as an independent component associated to specific brain areas and computationally realized by ACT-R modules (see Figure 2). In general, both IBL and RL can produce different results depending on the input. Complementarily, recent literature (e.g., Castelfranchi & Falcone, 2010) suggests that trust can mitigates risk and is developed through risk-taking and reciprocity. Inspired by this perspective, we designed a new mechanism to implement "confidence" in the evaluation of outcomes (TA): the more positive is the outcome, the higher the trust a player has in the adopted strategy and in the opponent's intentions. When the outcomes are negative and the player is not satisfied by the development of the game, actual trust decreases but the readiness to risk increases (aiming at turning the game to a positive course). As this initial description suggests, the positive or negative trends in TA are tightly linked to the emotional responses of the player. In this sense, TA can be seen as a proxy of an "affective module" in the ACT-R architecture. Let us see how these mechanisms work in the model and how they can be de-activated according to the experimental requirements.

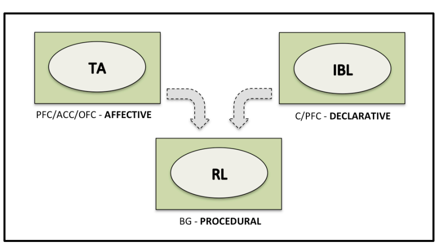

Figure 2. Mapping between IBL, RL, TA, ACT-R modules and brain regions. PFC = Prefrontal Cortex; ACC = Anterior Cingulate Cortex; OFC = Orbitofrontal Cortex; C = Cortex; BG: Basal Ganglia.

A Cognitive Model of Repeated CG

Our simulation of repeated CG under alternative information conditions is grounded in the work on generalization of learning presented in (Juvina, Lebiere, Gonzalez & Saleem 2012): more specifically, our compositional approach stems from the model for round-by-round human data in both CG and PD (Prisoner's Dilemma). The model is described "downward", from full-availability to absence of information.

Experiential Condition

Under the Experiential condition, each player can only see the actions and the outcomes of the other player, eventually learning from round to round how the payoffs are assigned according to specific interdependency between actions. From a cognitive modeling perspective, the model uses *(i)* IBL to predict the opponent's move on the basis of its past behavior and *(ii)* RL for action selection based on the utilities of own past moves. The core declarative structure of the model is the chunk-type *play*, whose slots denote the opponent's moves, respectively, in the current trial t and in the previous trial t-1[4]. The model generates expectations for the opponent's next move based on its history of sequential play. The most important procedural structures, conversely, correspond to the productions that control the decision-making process, namely action selection on the basis of predicted outcomes. While IBL is essentially realized at the declarative level, by evaluating the contents of the *play* slots in the *retrieval* buffer of multiple productions, RL focuses on the procedural level, where productions are selected according to utility computation (Fu & Anderson, 2006)[5]. In ACT-R each production has a utility value U associated with it: of the competing productions (those which match the current state) the one with the highest current U will be selected. The U value is based on the rewards that a production receives and will change as the model runs. The RL mechanism updates utility on the basis of reward. In general, the reward is a scalar quantity, whose value varies according to the difference between predicted and actual reward, which is consistent with the functional role played by dopaminergic signals in basal ganglia (Fu & Anderson 2006). In our model, reward in RL is determined by the state of TA resulting from previous interactions: the dynamics of reward is driven by two variables, "trust accumulator" and "willingness to invest in trust", as introduced in (Juvina, Lebiere, Gonzalez & Saleem 2012). 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 willingness to invest in trust is positive, the reward is based on the opponent's payoff (a player acts selflessly, trying to maximize the opponent's payoff). This can lead to mutual cooperation and development of trust (or players may

[4] This modeling feature is grounded in the learning of sequential moves, a method adopted in other ACT-R game playing models, e.g. paper rock scissors (West, Lebiere 2001).

[5] Sub-symbolic mechanisms are also important at the declarative level, e.g. in the calculation of chunk activations.

relapse into mutual destruction). When the trust accumulator is positive, the reward is equal to the difference between the current joint payoffs and the previous payoff of the opponent (a player tries to maximize joint payoff and avoid exploitation across multiple trials). Thus, when the model accounts for the Experiential and Descriptive conditions, IBL, RL and TA are enabled, as shown in Table 2[6].

Descriptive Condition

As previously mentioned, in the Descriptive condition each player has access to the complete pay-off matrix from the outset. Keeping the learning mechanisms *(i)* and *(ii)* enabled, we further modeled the Descriptive condition by increasing the base level activation of the *play* chunks for the mutual swerve strategy, reflecting the salience of mutually positive outcomes. This way we could recreate, at the computational level, a *knowledge priming* effect of the payoff matrix. Let us see how a principled subtraction of TA and IBL from the model recreates, correspondingly, the Minimal and Individual conditions.

Minimal Condition

The Minimal condition holds when a player can only access its own moves and payoffs, without knowing anything about the opponent but the fact that it is another human player and not a random strategy or a probabilistic algorithm. In this setting, the *play* chunk doesn't contain any slot denoting the opponent's moves, and refers just to the player's own current and previous moves. Therefore, the IBL mechanism is used in the model to select the current action just on the basis of a player's past behavior. RL is triggered by a basic payoff-centric rule: in particular, if the current payoff is higher than the previous one, the reward is equal to the difference between the two (the strategy is successful); else, if the current payoff is lower than the previous one and both are negative, the reward is set to the sum between the current payoff and the previous payoff (RL penalizes a failing strategy). Finally, when the current payoff is lower than the previous one and the latter is positive, the reward is set again as the difference between the two payoffs. In the last case, the value of the reward can be positive (t=1, t-1=10), neutral (t=1, t-1=-1) or negative (t=-10, t-1=1): this reward function captures the fact that when a player gets a positive outcome in a previous trial t-1, the effectiveness on the strategy depends on the consistency of the decision in the current trial t. In the Minimal condition TA is not considered to impact on the utility of productions because a player cannot build a mental model of the opponent's behavior (especially of the action it takes and the incentives it responds to): in general, when a person doesn't have any belief about another person, she won't engage in any further mental attitude about the other one, including trust or disposition of trusting in the near future (Ferrario &Oltramari 2004).

[6] We direct the reader to the last section of the paper for a discussion on the "Oracle" condition.

Individual Condition

In the Individual condition, both TA (as in the Minimal condition) and IBL are deactivated and only the RL mechanism is maintained. The model is operating on the basis of two simple productions for action selection in the current context (Dare or Swerve), without any consideration of the past behavior and the sole payoff received at each trial playing the role of reward in updating the utility value.

Table 2 – Combination between cognitive mechanisms and different information conditions (RL is not represented because it's active in all conditions).

		IBL	
		Activated	**Deactivated**
TA	**Activated**	Experiential/Descriptive	Oracle
	Deactivated	Minimal	Individual

Results

We present here the results of repeated CG in the four information conditions, comparing the model predictions with human data from the laboratory experiment, aggregated over 20 blocks of 10 trials for 30 pairs in each condition. The human proportion of alternating cooperation (see Figure 3) highly depended on the availability of information. In early stages of the interaction the proportion of alternating cooperation was low and similar in all of the information conditions. Participants in the Individual and Minimal conditions quickly established a relatively stable proportion of alternating cooperation and maintained this low proportion throughout the rest of the interaction. Furthermore, participants in the Minimal condition had a consistently higher proportion of alternating cooperation, compared to the Individual condition. In contrast to the stable behavior found in the less informative conditions, the proportion of alternating cooperation in the Experiential and Descriptive condition increased with repeated interaction. In general, the proportion of alternating cooperation increased faster and had higher values in the Experiential condition, compared to the Descriptive condition, and all other conditions. We were able to replicate this trend in the model by boosting the base-level activation of the mutual swerve[7] to reflect inspection of the payoff matrix in the Descriptive condition, *de facto* delaying the exploration of alternations and weakening the stability of the optimal strategy. Figure 4 clearly displays that the model was able to fit the trends of the human behavior with consistent adequacy. The dynamics in the proportion of mutual swerve also varied in the four information conditions. As we can see in Figure 5, participants in the Individual and Minimal condition demonstrated a relatively similar pattern. For these two conditions, the proportion of mutual swerve was low, and it

[7] In the declarative module, "mutual swerve" is represented as a specific subtype of the chunk-type *play* (whose slots denote a player and the opponent's move). It was manually set to 1.0.

decreased with the repeated interaction. From figure 6 we can also observe that the results for the model reflect the human pattern for Individual and Minimal conditions, where the proportion of mutual swerves generally remains stable, though with a slight increase across multiple trials. On the other hand, the ability to learn from experience increased the proportion of mutual swerve in the Experiential and Descriptive conditions. Furthermore, the availability of the payoff matrix in the Descriptive condition led to the highest proportion of mutual swerve. Regarding the dynamics of mutual swerve in the Experiential and Descriptive conditions, there was an initial increase, followed by a decreasing trend in the proportion of mutual swerve as participants collected more experience in the game. This is in parallel to the gradual increase in the proportion of alternating cooperation found for these two conditions. As Figure 6 shows, the model generally matches the behavioral trend for the Experiential and Descriptive conditions, though the overall proportion of mutual swerve tends to lightly increase (as for the less-informative conditions).

Unlike the other behaviors, the proportion of mutual dare was not sensitive to availability of information. In all conditions the proportion of mutual dare was relatively low and decreased with the progression in the game (Figure 7). This constancy in the mutual dare strategy was also replicated in the cognitive simulations for all the information conditions (Figure 8). Overall, the variants of model based on the availability of information match the trends in the human data adequately well.

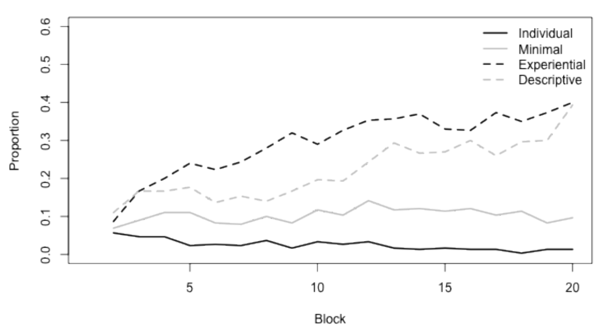

Figure 3. Human: average proportion of Mutual Alternation

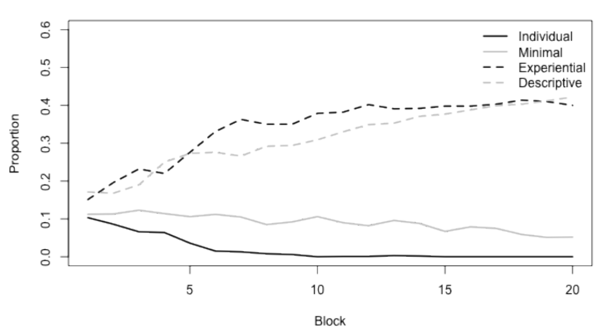

Figure 4. Model: average proportion of Mutual Alternation.

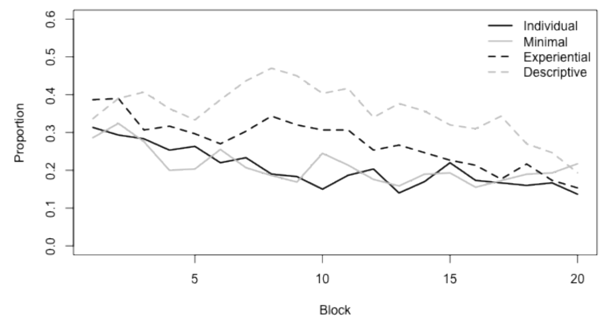

Figure 5. Human: Average proportion of Mutual Swerve.

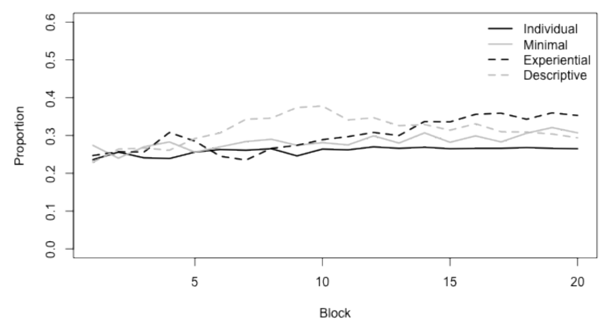

Figure 6. Model: average proportion of Mutual Swerve.

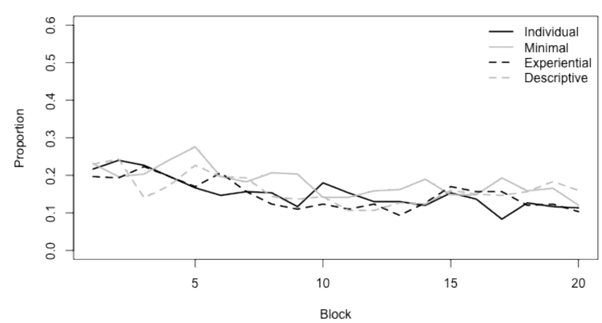

Figure 7 Human: average proportion of Mutual Dare.

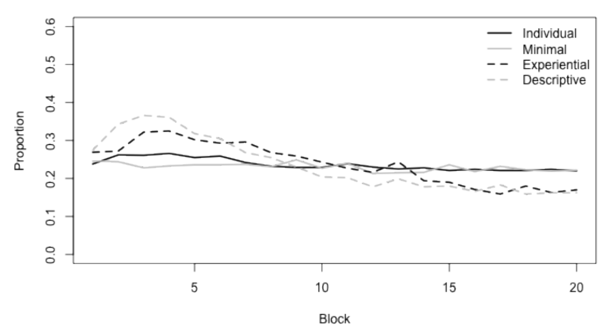

Figure 8 Model: average proportion of Mutual Dare.

Discussion and Conclusion

In this study we investigated how the strategic dynamics of social interaction vary under different information conditions. In particular we focused on the role of *instance-*

based learning, *reinforcement learning* and *trust accumulation* in a non-zero sum game of repeated CG. We illustrated how these cognitive mechanisms can be used as independent components in an ACT-R model of repeated social interaction. Final results show that the model, by means of its compositional features, successfully matches behavioral data. ACT-R's activation noise parameter was set to a) 0.1 for the less-informative conditions and to b) 0.02 for full information availability: this modeling choice reflects a) the higher stochasticity emerging when players have a blurred picture of the game-scenario and b) the solid prominence of rational strategies when they fully master the game, such as when knowing the opponent's actions and outcomes history and seeing the complete payoff matrix. For the sake of completeness, we intend to run further evaluations on alternative values of noise in our future work. Besides the Individual, Minimal, Experiential and Descriptive conditions, Table 2 shows that a fifth condition, labeled "Oracle", is hypothesized. No experimental data suggested the incidence of this condition, whereas it seems to be functionally legitimated by a combination of RL and TA in the cognitive modeling framework. Due to space limitations, we cannot examine this conjectural condition here, planning to devote future research to a deeper analysis. Preliminarily, we can say that the choice of the label was influenced by the peculiarities a player should manifest, such as obliviousness of the temporal structure of the game and of the opponent, yet holding a sense of trust. Like an "Oracle" that doesn't need to socially interact or inspect the world to come to a decision, under the fifth condition the model would exploit some kind of "introspection" as a means to decision-making. In this context, an intriguing problem is understanding which reward structures prompt self-analysis and, ultimately, how they can govern choices: the agent would have to build a sense of trust only on the basis of its internal mental states. We can speculate that these features evoke human *feelings*: in particular, as argued in (Damasio, A. 2003), feelings functionally arise from contents of thoughts. This interpretation is consistent with our view of trust accumulation as a proxy of an "affective" module in ACT-R, though we would need to investigate how emotional responses and feelings can be distinguished at the architecture level. In conclusion, our results demonstrate that a compositional approach to mechanisms within the ACT-R architecture matches experimental results found under various information conditions and suggests a general approach for building cognitive models that account for human performance in information-rich environments.

Acknowledgments

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

References

Anderson, J.R., Lebiere, C. (1998). *The Atomic Components of Thought.* NJ: Erlbaum, Mahwah.

Anderson, J. R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y. (2004). An integrated theory of the mind. *Psychological Review* 111, (4).

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

Ben-Asher, N. Lebiere, C., Oltramari, A. and Gonzalez, C. (2013). *Fairness and Efficiency Aspects of Cooperation in Repeated Social Interaction*. Proceedings of the 35th Annual Conference of the Cognitive Science Society.

Ben-Asher, N., Dutt, V., and Gonzalez, C. (2013). Accounting for the integration of descriptive and experiential information in a repeated prisoner's dilemma using an instance-based learning model. *Behavior Representation in Modeling and Simulation (BRiMS)*.

Bothell, D. (2004). ACT-R 6.0 Reference manual -Working Draft: http://act-r.psy.cmu.edu/actr6/reference-manual.pdf

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

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

Dancy, C. L., Ritter, F. E., & Berry, K. (2012). Towards adding a physiological substrate to ACT-R. *In Proceedings of the 21st Conference on Behavior Representation in Modeling and Simulation,* 78-85. Amelia Island, FL.

Damasio, A. (2003). *Looking for Spinoza. Joy, Sorrow and the Feeling Brain*. New York: Hartcourt Inc.

Ferrario, R., Oltramari, A. (2004). *Towards a Computational Ontology of Mind*. In Varzi, A. Vieu, L. (editors). Proceedings of FOIS 2004 4-6 November, Turin, Italy, pp. 267-277, IOS Press.

Fu, W-T. & Anderson, J. R. (2006). From recurrent choice to skill learning: A reinforcement-learning model. *Journal of Experimental Psychology: General, 135*(2), 184-206.

Gonzalez, C., & Martin, J. M. (2011). Scaling up instance-based learning theory to account for social interactions. *Negotiation and Conflict Management Research*, 4(2), 110-128.

Gonzalez, C., Ben-Asher, N., Martin, J. M., & Dutt, V. (2013). Emergence of cooperation with increased information: Explaining the process with an instance-based learning model. *Under Review*.

Juvina, I, Lebiere C, Gonzalez C, Saleem M. (2012). Generalization of learning in games of strategic interaction. CogSci. , Sapporo, Japan: Cognitive Science Society.

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

Russell, B., (1959). *Common Sense and Nuclear Warfare*. Simon & Schuster, New York.

West, R. L., & Lebiere, C. (2001). Simple games as dynamic, coupled systems: Randomness and other emergent properties. *Journal of Cognitive Systems Research*, 1(4), 221-239.