On linking cognitive mechanisms to game play
A critique of Morikawa, Hanley, and Orbell

Peter Stone, PhD
Political Science Department
Stanford University
Stanford, CA 94305-6044
stone68@stanford.edu

Abstract. Tomonori Morikawa, James E. Hanley, and John Orbell have argued that natural selection leads populations who play Hawk-Dove, a game-theoretic stylization of confrontation, to develop the capacity for various "orders of recognition." Such an argument requires a model linking game play to the presence or absence of various cognitive mechanisms. Morikawa and colleagues present such a model but, I argue, leave it incomplete, unable to sustain the conclusions they wish to defend. The development of a more fully specified model would significantly assist future studies of cognitive structures related to game play.

In a recent article in *Politics and the Life Sciences*, Tomonori Morikawa, James E. Hanley, and John Orbell imagine a population whose members periodically pair off to play Hawk-Dove games.[1] The payoffs of these games consist of resources (such as food) that can increase the survival odds of the player receiving them; the higher the payoff, the greater the chance the individual receiving the payoff will survive and produce offspring. Morikawa and colleagues consider how natural selection might work on such a population, with greater evolutionary success rewarding players who average high payoffs.

In seeking to explain what impact natural selection might have on game play, Morikawa and colleagues follow a path trod by many social scientists and natural scientists alike. Indeed, a lively field known as *evolutionary game theory* exists to explore just this question. The primary results generated in this field have been surveyed in a number of recent works.[2,3] However, the approach taken by Morikawa and colleagues is different from that employed by most students of evolutionary game theory. The latter examine the likely evolutionary success of various *strategies* that players may employ while playing games. They do not investigate why the players play these strategies; instead, they treat the process by which strategies are selected as a "black box" and consider only whether the strategies lead the players using them to thrive or starve. But if this black box could be opened, useful information could be derived about the cognitive mechanisms driving this strategy-selection process. This is the goal of the research project to which Morikawa and colleagues belong. Leda Cosmides and John Tooby have put the point as follows:

> Every economic model entails theories about these computational devices, but they are usually left implicit, buried in the assumptions of the model. At the moment, most economists rely on the implicit (and somewhat vague) theory that these computational devices somehow embody 'rational' decision rules. But developing a more accurate, useful, and well-defined substitute for this black box is now a realistic goal.[4]

Morikawa and colleagues thus wish to open the black box in evolutionary game theory and examine the cognitive mechanisms that lead players to play games in one manner or another. Some cognitive mechanisms, they argue, will lead players to play strategies that generate high payoffs; players without these mechanisms will be stuck following lower-payoff strategies. In this manner, Morikawa and colleagues believe that natural selection can explain the development of various cognitive mechanisms in game-playing populations, such as the human race. In particular, natural selection can account for the development of the capacity for various "orders of recognition" in game players. Such a capacity consists of

the ability to recognize certain types of information that might prove useful in adopting a game strategy.

The study of natural selection's influences on the cognitive mechanisms of game play is still in its infancy. Up to now, most of the work done on this subject has revolved around Prisoners' Dilemma, a game that many evolutionary theorists believe captures the logic of social exchange.[5] Morikawa and colleagues wish to extend this work to cover Hawk-Dove, which they believe effectively models another evolutionarily significant class of social interactions. This extension should intrigue anyone interested in understanding the cognitive structure of the human mind, and the authors are to be commended for undertaking it.

If Morikawa and colleagues are to succeed, however, in arguing that certain cognitive capacities would emerge due to evolutionary pressure generated by Hawk-Dove interactions, their argument must satisfy certain conditions. For a given cognitive mechanism, *X*, the argument must first specify the payoff players lacking *X* can expect to get, on average, while playing other players also lacking *X*. Then it must show that a player entering the game bearing *X* would in fact derive a higher average payoff from playing against players lacking *X* than the players lacking *X* would derive. Finally, it should show that a player with *X* would continue to derive higher payoffs than players lacking *X* even if the proportion of players with *X* in the population increases. (This condition might not hold if, for example, players without *X* did significantly better against each other than players with *X* did when playing each other.)

The third condition seems intuitively very likely, as Morikawa and colleagues recognize when they assume that cognitive capacities are "upward-ratcheted," meaning that capacities that confer an advantage once gained never become a disadvantage later on.[6] The first two conditions, however, are critically important. There, conditions require a clear specification as to how both players with *X* and players without *X* will play the game. This means linking the cognitive mechanisms in question to game play. In game theory, this is normally done through some sort of model depicting the actors in a game and deriving conclusions about their behavior from various assumptions about their characteristics (especially cognitive capacities). Morikawa and colleagues attempt to do this, but their argument is incomplete in important ways; moreover, successfully completing this argument is a formidable task.

In the sections that follow, I first lay out the basic characteristics of Hawk-Dove as described by Morikawa and colleagues. I then briefly restate some of their preliminary conclusions in ways that clarify my concerns. Next, I demonstrate that Morikawa and colleagues underestimate the incompleteness of their argument and, as a result, draw conclusions they cannot defend. Finally, I describe some of the difficulties likely to confront those who might wish to further this work.

Hawk-Dove

In Hawk-Dove, two individuals vie for control of a resource that can contribute to evolutionary success. Each player must choose one of two strategies, "hawk" or "dove." If both players play dove, neither obtains the resource, and both receive a payoff of 0. If one player plays hawk and the other plays dove, the one playing hawk receives the value of the resource, *V*, and the other player receives nothing. If both play hawk, there is a fight between the players for control of the resource, a fight whose outcome is determined by the respective strengths of the players. Let the strengths of players 1 and 2 be represented by S_1 and S_2, respectively. Then the probability that an individual, *i*, will win the fight equals *i*'s proportion of the two players' total strength, or $S_i/(S_1 + S_2)$. This probability can be designated $p_i(win)$. The player winning the fight receives a benefit, *V*, while the loser pays a cost, *C*. (As noted before, both the benefit and the cost are measured in terms of the contribution, whether positive or negative, made to the survival chances of an individual, *i*.) Assuming, as Morikawa and colleagues do, that players have Von Neumann-Morgenstern utility functions, the players will treat their expected payoffs from a fight as *real* payoffs. These expected payoffs are $(S_1/(S_1+S_2))V-(S_2/(S_1+S_2))C$ for player 1 and $(S_2/(S_1+S_2))V-(S_1/(S_1+S_2))C$ for player 2. Call these expected payoffs EP_1 and EP_2, respectively. The payoff matrix for this game can then be depicted as follows:

		Player 2: Hawk	Player 2: Dove
Player 1	**Hawk**	EP_1, EP_2	V, 0
	Dove	0, V	0, 0

For Morikawa and colleagues, the game-playing

advantage imparted by a given cognitive mechanism will determine the likelihood of its conservation. Let us turn, then, to the cognitive mechanisms they have discussed.

Playing Hawk-Dove

Morikawa and colleagues divide cognitive mechanisms into two categories. Mechanisms of the first type they describe as mechanisms of *recognition*. These mechanisms allow players to perceive various features of the world, themselves, and other players. Mechanisms of the second type they call mechanisms of *processing*. These mechanisms transform the information recognized into forms useful in playing games. The authors' central concern is with mechanisms of recognition, and in particular with the various *orders of recognition* that these mechanisms make possible. First-order recognition consists of the recognition of features of the environment that can be described without reference to the recognition of anything else, by anyone else. Examples include recognition of the values of *V, C,* S_1, and S_2. Second-order recognition consists of the recognition of acts of first-order recognition. If player 1 discerns what player 2 believes to be the values of S_1 and S_2, for example, then that discernment would be an act of second-order recognition. Third-order recognition consists of the recognition of acts of second-order recognition, and so on. Anyone capable of *n*th-order recognition is presumed capable of some form of *n*-1th-order recognition, *n*-2th-order recognition, and so on.

Obviously, the capacity for a certain order of recognition need not be general. To say that a person can recognize the value of parameter *X* (first-order recognition) is not to say that this person can recognize the value of parameter *Y* (also first-order recognition). Similarly, just because one game player can recognize that another player recognizes the value of *X* (second-order recognition) does not mean that this player can see that this same player recognizes the value of *Y*, and so on. This complication Morikawa and colleagues avoid by studying only one particular form of recognition with an order higher than first.

Morikawa and colleagues take as their starting point a player capable of several types of first-order recognition. They assume that this player — say, player 1 — can estimate the payoff parameters *V* and *C*, as well as the strengths of the players S_1 and S_2. (For the sake of exposition, I shall assume throughout the paper that the analysis is concerned with the capacities and behavior of player 1. The exact same considerations apply, *mutatis mutandis*, to player 2.) They also assume player 1 will possess processing mechanisms capable of, first, estimating the probability he or she will win a confrontation resulting from both players playing the hawk strategy, $p_1(win)$, and, second, calculating, given this probability, the expected payoff he or she will enjoy as a result of such a confrontation, EP_1. Finally, they assume that players seek to maximize expected utility given the beliefs and preferences indicated to them via their mechanisms of recognition and processing.

Morikawa and colleagues make several claims about the behavior to be expected from player 1, given the cognitive mechanisms attributed above. Some of these claims are valid, others not, in the sense that the argument given does not sustain the conclusion. Unfortunately, the valid and invalid claims are somewhat tangled together. For this reason, I shall restate the valid results to be derived from the authors' investigation of first-order recognition and its effects on behavior, before proceeding to consider the invalid ones in the next section.

In Hawk-Dove, what counts as utility-maximizing behavior for player 1 depends critically on the expected utility to be derived when a fight breaks out (*i.e.*, when both players play hawk). As noted above, this expected payoff can be written as $EP_1 = (S_1/(S_1+S_2))V - (S_2/(S_1+S_2))C$. Whenever $EP_1 > 0$, hawk is a strictly dominant strategy, and player 1 will wish to choose it regardless of what player 2 might choose to do. In this regard, Hawk-Dove resembles Prisoners' Dilemma to player 1 as long as $EP_1 > 0$. When $EP_1 = 0$, hawk remains a dominant strategy for player 1. So, while there will be circumstances in which player 1 can do as well playing dove as playing hawk, hawk will sometimes be better (depending on what player 2 does), and it will never be worse. Either way, as long as $EP_1 \geq 0$, player 1 can never go wrong playing hawk.

The implications for this conclusion in terms of natural selection are important. Player 1 can always maximize utility by playing hawk whenever $EP_1 \geq 0$. Player 1 can never improve his or her play through the capacity for higher orders of recognition. There is no additional information that could make dove more attractive than hawk; therefore, natural selection will never pressure player 1 to develop further cognitive

capacities. If natural selection is to lead player 1 or his or her descendants to develop cognitive mechanisms capable of higher orders of recognition, then this development must depend on an improved ability to play the game whenever $EP_1 < 0$.

This conclusion is implicit in the analysis offered by Morikawa and colleagues but never stated explicitly. For example, Morikawa and colleagues provide a figure depicting a two-dimensional space. One dimension of this space represents the ratio of the value of victory to the cost of defeat (V/C) while the other represents player 1's estimated probability of winning the fight: $p_1(win)$, or $S_1/(S_1 + S_2)$.[7] Using this graph, the authors suggest that when $p_1(win)$ and V/C fall in certain regions of the two-dimensional space, hawk is the recommended strategy. They do not indicate that this is because, in those regions, hawk is either a dominant or strictly dominant strategy, depending on whether the point in question is on the boundary or within the interior of the region in question. In addition, at the very beginning of the paper they claim that Hawk-Dove is distinguishable from Prisoners' Dilemma because "there is no dominant incentive" in the former, as opposed to the latter. "What course of action is best for each individual can depend critically on the other's intentions and capacities."[8] It *can* so depend, to be sure, but it need not. Indeed, whenever $EP_1 \geq 0$, it does not hold for player 1, who does indeed have a "dominant incentive" after all.

Willing to fight?

Thus, although the argument Morikawa and colleagues make has the clear implication that player 1 will have a dominant strategy whenever $EP_1 \geq 0$, they fail to make this point explicitly. This failure has an unfortunate consequence. It blurs a valid prediction for game play whenever $EP_1 \geq 0$ with an invalid prediction for game play whenever $EP_1 < 0$. I shall therefore first review the structure of the game whenever $EP_1 < 0$ and then demonstrate how Morikawa and colleagues fail to take this structure properly into account in their analysis.

If $EP_1 < 0$, then player 1 expects to lose from a fight. That is, he or she will do worse on average than if no fight occurred. However, player 1 will not necessarily play dove under these conditions; hawk might still be chosen if the opponent is believed likely to play dove. If player 2 *does* play dove, player 1 could capture the contested good without a fight by playing hawk. Player 1 might therefore be willing to risk negative returns by playing hawk so long as the risk is not too great. This willingness can be made precise. Assume that player 1 estimates the probability that player 2 will play hawk at p_2. Then player 1 would enjoy a higher expected return playing hawk if and only if

$$p_2 \leq \frac{(S_1 + S_2)V}{S_2(V + C)}. \tag{1}$$

Indeed, when $EP_1 < 0$, player 1 could have one of three expectations, each of which would dictate a different course of action. If he or she expects player 2 to play hawk, he or she should play dove. If he or she expects player 2 to play dove, he or she should play hawk. (The similarities between Hawk-Dove with this payoff structure and Chicken are obvious.) If player 1 expects player 2 to play hawk with probability p_2, player 1 will play hawk if equation (1) is satisfied with strict inequality; dove if it is not satisfied at all; and if (1) is satisfied with equality, then player 1 will be indifferent between hawk, dove, and any probabilistic combination of the two.

If player 1 is indeed a utility maximizer and if $EP_1 < 0$, then choice of strategy can be determined in one of two ways. If he or she can formulate a belief as to the value of p_2, then player 1 can decide how to act based on the considerations above. If not, then he or she faces a form of *uncertainty*, the absence of any probability estimates upon which to base a decision, as opposed to *risk*, which allows for such estimates.[9] (Formally, the term "uncertainty" is applied only when probability estimates are lacking in parametric, rather than strategic, decision-making situations. However, the problem faced is essentially the same in both contexts,[10] and so this complication can be ignored here.) The rational course of action in such cases is not clear. Kenneth J. Arrow and Leonid Hurwicz have forcefully argued that in cases of uncertainty, the rational actor can take into account only the best- and worst-case outcomes for each course of action.[11] In cases such as this, however, where one course of action offers both a better best-case outcome and a worse worst-case outcome than another, this restriction provides little guidance. But, in any event, without an estimate for p_2, player 1 needs some alternative means of resolving upon a course of action consistent with utility maximization.

Thus, to make a utility-maximizing decision when

$EP_1 < 0$, player 1 needs either an estimate for p_2 or a defensible means for making decisions under uncertainty. Either way, some additional cognitive apparatus is required. The first-order cognitive mechanisms that tell player 1 how to play when $EP_1 \geq 0$ are not enough when $EP_1 < 0$.

Morikawa and colleagues recognize this, so one would expect them to explain how an actor capable only of first-order recognition would decide to act when $EP_1 < 0$. However, they do not do this. Instead, they develop their argument with a solution to this problem already assumed. They assume, without argument, that a player capable only of first-order recognition will play dove whenever $EP_1 < 0$. This claim creeps into their argument early on, stays with them as they move from first-order to higher-order recognition, and remains undefended. They introduce it in a passage explaining the generation of EP_1, arguing that this process would use $p_1(win)$ together with V and C "to 'recommend' a fitness-maximizing choice between hawk and dove strategies" by calculating EP_1. They go on to claim of EP_1 that "[i]f positive, the individual should be willing to fight. If not, then not."[12]

"[W]illing to fight" here must have one of two meanings:

(2) Player 1 is willing to fight if he or she prefers playing hawk over playing dove but is unwilling to fight with the opposite preference.
(3) Player 1 is willing to fight if he or she expects a higher return from engaging in a fight (*i.e.*, when both players play hawk) than from not engaging in a fight at all (*i.e.*, by playing dove). In other words, being willing to fight simply means that $EP_1 > 0$, and being unwilling to fight means that this is not the case.

Each possible meaning of "willing to fight" points to a different evaluation of the claim that player 1 should be willing to fight when $EP_1 > 0$ and unwilling otherwise. If Morikawa and colleagues simply mean claim (3), then their claim is trivially true. In this case, however, the claim does not say anything about how player 1 should behave when $EP_1 < 0$, for the reasons given above. The quotation marks they place around "recommend" are justified indeed; the knowledge that $EP_1 < 0$ does not in itself point to either strategy in any meaningful way. If, on the other hand, the authors mean claim (2), then their claim is partially false and partially undefended. On the one hand, it is false to say that player 1 will have any reason to play dove when $EP_1 = 0$. The opposite is the case; hawk is a dominant (though not strictly dominant) strategy. On the other hand, the claim that player 1 will play dove when $EP_1 < 0$ is undefended. Whether player 1 will do this or not must depend either on the assessment of p_2 or on his or her mechanism for decision-making under uncertainty. The authors attribute neither an assessment nor a mechanism to player 1, and so in no way motivate the behavior attributed.

Thus, as stated, the authors speak imprecisely about "willingness to fight," making either a problematic claim (2) or a trivial claim (3) about the relationship between EP_1 and the optimal strategy to be pursued by player 1. Within short order, however, they have clarified the situation: they intend the problematic rather than the trivial claim. In the course of a discussion of what I as player 1 might do once I have calculated EP_1, they write the following:

> That value might be positive, in which case the recommendation from my cognitive apparatus is that getting into a fight would be worthwhile, thus that hawk will be a rational choice whether or not you do decide to fight. Or that value might be negative, in which case the recommendation would be to avoid the fight.

Thus far, their claim remains ambiguous. Do they intend to suggest that the value of EP_1 always recommends a strategy (2), or do they intend merely to describe what EP_1 says (3)? They go on to write the following about the case where $EP_1 = 0$:

> In the threshold case, the expected value of a fight might be exactly zero. Formally, of course, with the expected value at zero, I should be indifferent between fighting and not fighting, tossing a coin to decide what to do.[13]

In effect, Morikawa and colleagues claim that if player 1 is indifferent between the expected value of a fight ($EP_1 = 0$) and the outcome ensuing when choosing not to risk a fight (0), then he or she must also be indifferent between playing hawk and playing dove. But this claim would be false. If playing hawk, player 1 might get $EP_1 = 0$, which is no better and no worse than playing dove. But player 1 might also get V if

player 2 plays dove. As noted before, playing hawk is a dominant strategy when $EP_1 = 0$; a coin toss would make no sense.

The authors are conflating claims (2) and (3), in effect suggesting that knowing that $EP_1 < 0$ is enough to recommend dove as the preferred strategy to player 1, just as hawk is the preferred strategy when $EP_1 \geq 0$. Yet, while hawk is a dominant strategy in the latter case, dove is not dominant in the former case; the borderline case, where $EP_1 = 0$, is also a point of confusion. Perhaps an ambiguity in the attractively intuitive term, "willing to fight," led the authors astray here — and throughout the balance of paper.

When they consider the possible value to player 1 of a second-order recognition mechanism — specifically, one for recognizing player 2's assessed probability of winning a fight, $p_2(win)$ — Morikawa and colleagues inquire whether such a mechanism would lead players to behave differently. If not, then it could provide no survival benefit and would not be conserved as a trait. To assess whether the development of this capacity would alter behavior, however, they need to be able to say how players without the capacity — with only the cognitive mechanisms specified earlier — would play the game. This is not a problem for players when $EP_1 \geq 0$; it is clear how such players would play the game and equally clear that no second-order recognition capacity could improve expected payoffs for such players. But there still is no effective argument as to how such players would play the game when $EP_1 < 0$. The authors assume that such players would play dove no matter what, and, as a result, the conclusions they draw at this stage are also flawed.

Thus, Morikawa and colleagues claim that the benefit player 1 can gain from the ability to recognize player 2's assessment of $p_2(win)$ will depend on EP_1. They correctly note that the additional knowledge second-order recognition would generate could not possibly benefit player 1 when $EP_1 \geq 0$. (Again, they state this somewhat indirectly, using their graph, as they make no direct references to the presence or absence of dominant strategies depending on EP_1.) They then, however, consider the case where $EP_1 < 0$. Following the authors, suppose this condition holds, and that player 2's estimate of $p_2(win)$, together with 2's estimates of V and C, leads player 2 to believe that $EP_2 \geq 0$. Then if player 1 could identify player 2's estimate of $p_2(win)$, player 1 would probably also conclude that player 2 believes that $EP_2 \geq 0$. (The only complicating factor here is that player 1 and player 2 might have different estimates of V and C. For the sake of argument, suppose that player 1 knows player 2's estimates of these parameters as well.) Morikawa and colleagues conclude that, if all these conditions hold, then if I were player 1, the second-order knowledge could not possibly "change my *un*willingness to fight."[14] They therefore conclude that, in this case as well, second-order recognition capacity could serve no purpose.

Two things are worth noting here. First, Morikawa and colleagues now explicitly assume that player 1's "unwillingness to fight" translates directly into a decision to play dove as long as player 1 has only first-order recognition capacity. There is no other way to make sense out of this passage, but there is also no argument as to why player 1 would play this way. Second, this assumption by Morikawa and colleagues does real work here. Assume that a plausible theory as to how player 1 should play under uncertainty is developed, and suppose that it suggests that when player 1 possesses only the first-order recognition capacities player 1 should play hawk with some positive probability when $EP_1 < 0$ and player 2's course of action is unknown. Now suppose that player 1 obtains the second-order ability to discern player 2's estimate of $p_2(win)$ and using it deduces that player 2 believes $EP_2 \geq 0$. In this case, player 1 knows that player 2 plans to play hawk. But if $EP_1 < 0$ and player 2 plays hawk, player 1 will (by iterated dominance) want to play dove. In other words, player 1 will play the game differently depending on whether he or she does or does not possess second-order recognition capacity. But this means that the development of second-order capacity, *contra* Morikawa and colleagues, does real work in the case where $EP_1 < 0$ and $EP_2 \geq 0$. This development will aid player 1 as long as one assumes that when player 1 lacks second-order recognition capacity he or she plays hawk with some positive probability. This assumption is denied by Morikawa and colleagues, but it is just as plausible as their own contrary assumption.

Thus, when the authors assume that players with first-order recognition play dove when $EP_1 < 0$, their assumption has consequences for their argument regarding second-order and higher-recognition capabilities. Their assumption therefore cannot reasonably be made without a supporting argument of some kind. And this Morikawa and colleagues nowhere provide.

Conclusion

To predict what sort of cognitive abilities will be generated via natural selection in a game-playing population, Morikawa, Hanley, and Orbell must predict how individuals with different abilities will play Hawk-Dove. These predictions require a model that links the cognitive mechanisms available to the game player (mechanisms allowing the performance of acts both of recognition and of processing) to the predicted actions. Morikawa and colleagues make predictions with regard to actors capable only of certain forms of first-order recognition together with the relevant processing mechanisms. The model of these actors that they provide, however, is incomplete, in that it provides no predictions as to how an actor with the model-specified capacities will behave under certain well defined circumstances. Their argument regarding the development of higher-order recognition abilities builds on this model; there is no way to say whether the development of second-order recognition ability will help an actor without knowing how that actor is doing with first-order ability alone. Morikawa and colleagues overlook the incompleteness of their model, and as a result their conclusions regarding higher-order recognition abilities cannot be established.

The incompleteness of the model upon which Morikawa and colleagues build, however, is no cause for shame on their part. In effect, actors with only limited capacities for orders of recognition live in a world of *bounded rationality*. Standard game theory analysis builds into its models of human behavior the assumption that this behavior satisfies several highly demanding conditions. For example, game theorists typically assume that both *maximization* (players maximizing expected utility) and *consistency* (each player's beliefs about the behavior of other players being consistent with the behavior actually displayed by those players) characterize player behavior.[15] These highly demanding assumptions are usually, but not always, enough to generate specific predictions regarding game play. (Exceptions exist; for example, when a game is infinitely repeated under the right conditions, players could display an infinite variety of behaviors consistent with maximization, consistency, and so forth.[16]) Models of bounded rationality, however, are less precisely formulated and, so, yield less testable predictions.

Morikawa and colleagues doubtless would argue that the standard model of the strategic actor is irrelevant for their purposes. After all, the whole point of their paper is to study whether or not real people playing real games would ever, thanks to natural selection, attain cognitive abilities that are anywhere near those upon which standard game theory analysis relies. Morikawa and colleagues are content to retain the maximization assumption but clearly want to discard the assumption of consistency. Utility-maximizing players whose behavior satisfies the consistency condition always have correct beliefs about the strategies to be employed by other players. As utility-maximizers, they therefore adopt strategies that provide them with the highest possible expected utility given the strategies undertaken by others. In effect, players whose behavior satisfies the maximization and consistency conditions always play in Nash equilibria. However, since a player at such an equilibrium is already receiving his or her highest possible expected utility, given the strategies employed by other players, there is no way that the development of new capacities could assist that player in obtaining higher utilities. In short, maximizing and consistent players are playing games as well as they can; if the players in the paper can improve their payoffs through the development of new capacities, and if they are utility-maximizers, then their behavior cannot be guaranteed to satisfy the consistency condition.

Thus, the authors' desire to model players as possessing weaker capabilities than those possessed by actors in standard game theory analysis is unobjectionable. In effect, it is a call for the creation of a model more general than the standard one, perhaps with the standard model as a limiting case. Such a general model would specify, for a given level of cognitive abilities, precisely how players with those abilities would play games. This model would allow students of cognitive mechanisms to show in a precise manner that such-and-such cognitive mechanism makes such-and-such a contribution to the evolutionary fitness of game-playing actors. Unfortunately, no such model exists now, and the result is indeterminacy of prediction whenever models relaxing these conditions are created. This is precisely the problem encountered by Morikawa and colleagues.

In another context, Jon Elster acknowledged the limits of the standard model of strategic action, particularly the unrealistic nature of its assumption of utility maximization. He further noted that models that assume

bounded rationality, such as the model of "satisficing" developed by Herbert Simon,[17] avoid this lack of realism — but only at a cost. "The reason," he writes,

> why the "satisficing" models have not replaced maximizing models in economics, in spite of very powerful *a priori* arguments, is that they offer no hope of arriving at a determinate explanation. It is no doubt true that we often go for what is "good enough" rather than for the "best", but this is of little help as long as we do not possess a simple and general theory to tell us how people arrive at their estimates of what is good enough.[18]

Much the same can be said of the problem posed by the argument of Morikawa and colleagues. It is no doubt true that human beings developed their present set of cognitive abilities over time, and that natural selection promoted the expansion of this set (at least up to a point) due to the contributions the expansion made to game-playing ability. But this is of little help as long as we do not possess a general theory to tell us how people with cognitive abilities weaker than those of the standard strategic actor will play games. The need for a theory of this nature is, I suggest, the most important task today facing students of natural selection's influence on mental performance.

References

1. Tomonori Morikawa, James E. Hanley, John Orbell, "Cognitive Requirements for Hawk-Dove Games: A Functional Analysis for Evolutionary Design." *Politics and the Life Sciences,* March 2002, 21:1, 3–12.

2. Jörgen W. Weibull, *Evolutionary Game Theory* (Cambridge, MA: MIT Press, 1997).

3. Larry Samuelson, *Evolutionary Games and Equilibrium Selection* (Cambridge, MA: MIT Press, 1998).

4. Leda Cosmides John Tooby, "Better than Rational: Evolutionary Psychology and the Invisible Hand." *American Economic Review* (May 1994), 84:2, 327–332, 327.

5. Leda Cosmides John Tooby, "Evolutionary Psychology and the Generation of Culture, Part II. Case Study: A Computational Theory of Exchange." *Ethology and Sociobiology,* (1989), 10:1–3, 51–97.

6. Morikawa *et al.*, p. 3.

7. Morikawa *et al.*, p. 6.

8. Morikawa *et al.*, pp. 3–4.

9. Frank Knight, *Risk, Uncertainty and Profit* (Chicago: University of Chicago Press, 1971).

10. Jon Elster, *Explaining Technical Change* (New York: Cambridge University Press, 1983), p. 12.

11. Kenneth J. Arrow Leonid Hurwicz, "An Optimality Criterion for Decision-Making under Ignorance," in *Uncertainty and Expectation in Economics,* C.F. Carter and J.L. Ford, editors (Oxford: Basil Blackwell, 1972).

12. Morikawa *et al.*, p. 6.

13. Morikawa *et al.*, p. 7.

14. Morikawa *et al.*, p. 8.

15. George J. Mailath, "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory." *Journal of Economic Literature,* September 1998, 36:3, 1347–1374, 1347.

16. Drew Fudenberg Eric Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica,* May 1986, 54:3, 533–554.

17. Herbert Simon, "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics,* February 1955, 69:1, 99–118.

18. Jon Elster, "A Paradigm for the Social Sciences?" *Inquiry,* September 1982, 25:3, 378–385, 379.