On the Evolution of Choice Principles

Michael Franke[1] and Paolo Galeazzi[2]⋆

[1] Department of Linguistics, University of Tübingen
[2] Institute for Logic, Language and Computation, Universiteit van Amsterdam

Abstract. We compare level-k utility maximization and level-k regret minimization in evolutionary competition based on averages of one-shot plays over many randomly generated strategic games. Under the assumption that Theory-of-Mind-reasoning of depth k incurs a cost monotonically increasing with k, our results show that mixed states with low levels of reasoning depth k can be evolutionary stable with substantial basins of attraction under the replicator dynamics and that regret minimization can outperform utility maximization.

The widely accepted normative standard of individual decision making under uncertainty is expected utility maximization. An agent is rational only if his choices maximize expected utility. In games with several players, agents who believe that their opponents are rational might thus rule out certain of the opponents' choices. This kind of reasoning can be iterated, giving rise to predictions of rational choice under arbitrarily deep nestings of mutual beliefs in rationality.

Empirical data on human solitary or interactive decision-making is at odds with this idealized picture. For one, human decision makers systematically deviate from the predictions of expected utility theory (e.g. Tversky and Kahnemann, 1974, 1981). For another, human capability for higher-order Theory of Mind (ToM) reasoning appears to be rather limited (e.g. Ho, Camerer, and Weigelt, 1998; Keysar, Lin, and Barr, 2003; Verbrugge and Mol, 2008; Degen, Franke, and Jäger, 2013, inter alia).

The usual reactions to this discrepancy between idealized norm and empirical observation is to either rethink the prescriptive axioms of rational choice or, alternatively, to devise a descriptive theory to capture the empirical facts of human psychology of decision-making and strategic reasoning (e.g. Camerer, 2003; Glimcher et al., 2009). The approach taken here is different from either. We refer to a general decision making rule as a *choice principle*. Formally, a choice principle is a function that takes a game as input and returns a set of acts as the decision maker's choice. We then adopt an evolutionary point of view and ask: which choice principles are successful on average in recurrent competition with alternative choice principles (including more or less limited ToM reasoning)?

de Weerd, Verbrugge, and Verheij (2013) address a part of this question by looking at agent-based simulations of utility maximizers with different ToM reasoning capabilities when playing repeatedly selected zero-sum games against each other. The simulations of de Weerd, Verbrugge, and Verheij suggest that there is a benefit to deeper ToM reasoning only up to a fairly limited depth of reasoning.

Extending and generalizing this line of investigation, we use numerical simulations to approximate the expected payoff of agents with a fixed choice principle who play many *arbitrary* 2-player one-shot strategic games, not just a small hand-picked selection. This way we try to assess the general evolutionary benefit of a choice principle across an unbiased sample of games, not just for those games that are of particular technical, philosophical or historical importance. Additionally, we compare not only different ToM reasoning capability of utility maximizers, but also variation in the underlying method of choice. Concretely, we look at level-k utility maximization and level-k iterated regret minimization (IRM) (Halpern and Pass, 2012). The motivation for looking at IRM is threefold. For one, regret minimization has been suggested as a conceptually appealing alternative to utility maximization and a potentially promising descriptive theory of individual decision-making (Loomes and Sugden, 1982). For another, we are able to show that non-iterated regret minimization outperforms the closely related MaxiMin security strategy (see below) in evolutionary settings like the one investigated here, making regret minimization the evolutionarily best representative of a security strategy that we are aware of.[3] Finally, IRM usually requires only few iteration steps, making it a serious challenger of the standard theory in terms of limited ToM reasoning as well (c.f. Halpern and Pass, 2012).

The choice principles we compare are defined as follows. Look at level-k utility maximizers first. L-0 utility maximizers have an arbitrary probabilistic belief about the opponent's behavior and are indifferent between any act that maximizes expected utility under this belief. L-$k+1$ utility maximizers have an arbitrary probabilistic belief with support only on acts that an L-k opponent might choose, and are indifferent between any act that maximizes expected utility under this belief. Notice that this construction is different from that of de Weerd, Verbrugge, and Verheij (2013) in several respects. Firstly, we look at level-k reasoning models (e.g. Crawford, 2003, 2007), not cognitive hierarchy models (e.g. Camerer, Ho, and Chong, 2004), like de Weerd, Verbrugge, and Verheij do. The former models assume that agents of level $k+1$ believe that the opponent is of level k, while the latter have a more general belief that the opponent is of some level-l with $l \leq k$. Moreover, we assume all games to be one-shot encounters, while de Weerd, Verbrugge, and Verheij look at repeated encounters with the same opponent, and the concomitant possibility of strategic learning.

---

[3] More concretely, so far we have been able to prove this result for the class of 2-player symmetric games, and to demonstrate that it holds more generally using numerical simulations like the one reported here.

	EU-L0	EU-L1	EU-L2	EU-L3	EU-L4	RM-L0	RM-L1	RM-L2
EU-L0	6.735	6.749	6.739	6.873	6.755	6.742	6.733	6.734
EU-L1	7.126	6.650	6.649	6.610	6.954	6.969	7.016	7.017
EU-L2	6.642	7.413	6.979	6.857	6.961	6.414	6.494	6.495
EU-L3	6.636	6.946	7.627	7.152	7.162	6.430	6.465	6.465
EU-L4	6.780	7.038	7.314	7.734	7.388	6.592	6.620	6.619
RM-L0	6.686	6.684	6.700	6.830	6.695	6.681	6.676	6.675
RM-L1	6.853	6.768	6.782	6.913	6.806	7.051	7.046	7.043
RM-L2	6.856	6.770	6.784	6.916	6.810	7.052	7.057	7.053

Table 1. Average utilities of level-k utility maximizers and level-k regret minimizers over 10.000 arbitrary games (see main text). In our sample, regret minimizers of level $k > 2$ are behaviorally equivalent to RM-L2 and are therefore omitted in this table.

Let's turn next to the definition of level-k regret minimizers. Formally, if U is a matrix of utilities for the row player, regret minimization is equivalent to playing MaxiMin on the derived matrix U', where $U'_{ij} = U_{ij} - \max_k U_{kj}$ is the negative regret of choosing act i when the opponent chooses j. L-0 regret minimizers are indifferent between any act that minimizes regret in this sense. An L-$k+1$ regret minimizer chooses like an L-0 regret minimizer in the reduced game in which only acts remain that row and column players would choose that are L-k regret minimizers (see Halpern and Pass, 2012).

We randomly generate arbitrary strategic games and record the payoff earned by each choice principle when playing against any other, including itself. In general, the details of generating arbitrary games are important to this approach. Some choice principles might be better in certain classes of games, but not others. For that reason, we used several algorithms for creating games that are as unbiased and neutral as possible. We choose neutrality in the selection of games, because we have at present no better answer to the main empirical question that is relevant here, namely which kinds of interactive decision-making situations were most frequent during the critical stages of cognitive development. For illustration, this abstract focuses on the results of a simple generic algorithm for constructing arbitrary asymmetric strategic games with 2 players. The procedure starts by first picking a random number between 2 and 10 as the number of acts for each player, say n and m. Then, we determine each entry in the $n \times m$ and $m \times n$ utility matrices of each player as an integer between 0 and 10, all independently and uniformly at random.

Table 1 gives the accumulated payoff averaged over 10.000 games sampled in this way, for $0 \leq k \leq 4$. The table lists the average payoff for the row principle when playing against the column principle and can be regarded as a symmetric "meta-game" that captures the evolutionary competition between choice principles. The only evolutionarily stable strategy (ESS) of this meta-game is EU-L4, but it is clear (from more extensive simulations) that this is just because we chose an arbitrary cutoff for depth of reasoning. Generally, it appears that pure populations of EU-Lk players can always be invaded by EU-L$k+1$ players. But,

interestingly, some pure populations of EU-Lk players can also be invaded by EU-Ll players, with $l < k$ (see Table 1). On the other hand, if $k \geq 2$, then RM-Lk is a neutrally stable strategy (NSS): regret minimizers of a different level $l \geq 2$ would not be driven to extinction, but would also not take over the population (because they are behaviorally equivalent in all games from our random sample).

The results from this example generalize to other game-sampling routines, as long as they are general and encompassing enough. While there is usually no k at which EU-Lk is an ESS, there is a small k (almost always $k \leq 3$) for which IRM has reached its fixed point in all sampled games and RM-Lk' is an NSS for all $k' \geq k$. This is a noteworthy result: on randomly sampled games, it is always beneficial for utility maximizers to outsmart the opponent by ideally one level of ToM reasoning, but too much ToM outsmarting is harmful. For IRM, levels of ToM reasoning higher than a few steps are evolutionarily pointless.

More fine-grained results obtain when we look at the predictions of the replicator dynamics on the simulated "meta-games" together with the natural hypothesis that ToM reasoning incurs a cost $c(k) \in \mathbb{R}$, which is monotonically increasing with k. For concreteness, we look at a parameterized cost function $c(k\,;\,s,a) = \sum_{i=0}^{k} a^i s$ where s is the initial step of growth, and a is a parameter that controls for superlinear and sublinear growth. For at least linear growth ($a \geq 1$), there is a threshold $\theta \approx .391$ such that if $s > \theta$, the only attractor is EU-L0. With s smaller or not too much bigger than θ and sublinear growth (e.g., $a = 0.25$), RM-L1 is a strong attractor, as well as mixed states of low level EU-maximizers. For step $s < \theta$ and $a \leq 1$ results are more varied. If $.313 < s < \theta$, the only attracting state is a mix of EU-L0 and EU-L1. If $.011 < s < .313$, RM-L1 is a strict symmetric Nash equilibrium and therefore an attractor, but also, depending on s, mixed states of low-level EU-maximizers are attracting. If $0 < s \leq .011$ RM-L2 is an attractor, together with mixed states of low level EU-maximizers.

In sum, our simulation data show that shallow depth of reasoning is plausibly evolutionarily dominant in long run competition with "deep ToM" strategies. Regret minimization often outperforms utility maximization in this evolutionary competition, especially when reasoning costs are added. Although already insightful, our results are only partial, because many interesting questions have not yet been addressed. We mention just three obvious extensions for future work. Firstly, sequential games might yield different results, because they would induce more ties in strategic form, and hence more reasons to apply deeper ToM reasoning. Secondly, the payoff of utility maximization heavily depends on the type of belief about the opponent's choice. Here we used arbitrary beliefs compatible with the opponent's assumed type, but certain restrictions on belief formation, e.g., using maximum-entropy beliefs, might give utility maximizers a better payoff on average. Finally, it would be interesting to extend our comparison to more choice principles, e.g., ones including learning from repeated interactions with the same players (e.g. de Weerd, Verbrugge, and Verheij, 2013), or choice principles building on prospect theory, for example.

References

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

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2004). "A Cognitive Hierarchy Model of Games". In: *The Quarterly Journal of Economics* 119.3, pp. 861–898.

Crawford, Vincent P. (2003). "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions". In: *American Economic Review* 93.1, pp. 133–149.

— (2007). "Let's Talk It Over: Coordination Via Preplay Communication With Level-k Thinking". Unpublished manuscript.

Degen, Judith, Michael Franke, and Gerhard Jäger (2013). "Cost-Based Pragmatic Inference about Referential Expressions". In: *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Ed. by Markus Knauff et al. Austin, TX: Cognitive Science Society, pp. 376–381.

Glimcher, Paul W. et al., eds. (2009). *Neuroeconomics: Decision Making and the Brain*. Amsterdam: Elsevier.

Halpern, Joseph Y. and Rafael Pass (2012). "Iterated Regret Minimization: A New Solution Concept". In: *Games and Economic Behavior* 74, pp. 184–207.

Ho, Teck-Hua, Colin Camerer, and Keith Weigelt (1998). "Iterated Dominance and Iterated Best Response in Experimental 'p-Beauty Contests'". In: *The American Economic Review* 88.4, pp. 947–969.

Keysar, Boaz, Shuhong Lin, and Dale J. Barr (2003). "Limits on Theory of Mind Use in Adults". In: *Cognition* 89.1, pp. 25–41.

Loomes, Graham and Robert Sugden (1982). "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty". In: *The Economic Journal* 92, pp. 805–824.

Tversky, Amos and Daniel Kahnemann (1974). "Judgement under Uncertainty: Heuristics and Biases". In: *Science* 185, pp. 1124–1131.

— (1981). "The Framing of Decisions and the Psychology of Choice". In: *Science* 211.4481, pp. 453–458.

Verbrugge, Rineke and Lisette Mol (2008). "Learning to Apply Theory of Mind". In: *Journal of Logic, Language and Information* 17, pp. 489–511.

de Weerd, Harmen, Rineke Verbrugge, and Bart Verheij (2013). "How much does it help to know what she knows you know? An agent-based simulation study". In: *Artificial Intelligence* 199–200, pp. 67–92.