2017-07-20

Autonomous vehicle decision-making: Should we be bio-inspired?

Harris, Chris

10.1007/978-3-319-64107-2_25

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)

Springer International Publishing

Autonomous Vehicle Decision-Making: Should We Be Bio-inspired?

Christopher M. Harris(✉)

School of Psychology, Centre for Robotics and Neural Systems,
Plymouth University, Plymouth, Devon PL4 8AA, UK
cmharris@plymouth.ac.uk

Abstract. On our crowded roads, drivers must compete for space but cooperate to avoid occupying the same space at the same time. Decision-making is strategic and requires mutual understanding of other's choices. Fully autonomous vehicles (AVs) will need risk management software to make these types strategic decisions without human arbitration. Accidents will occur, and what constitutes rational and 'safe' decisions will be scrutinized by the legal system. It is far from clear how AV-Human and AV-AV interactions should be managed. Game Theory provides a framework for analyzing mutual 'games' with 2 or more players. It assumes that players mutually optimize their outcomes according to Nash equilibria (NE), but do humans follow Nash equilibria in Human-Human interactions? We implemented simple two-player competitive games to see whether people played rationally according to Nash equilibria. On each of 100 trials, each player was instructed to maximise their reward by pressing one of three buttons labelled "4", "6", and "12", without knowing the other players choice. If players pressed different buttons, they received a reward of 4, 6, or 12 points accordingly. If players pressed the same button, the reward was reduced depending on the game type. Results showed that players did not follow NE, but played a probabilistic game that included the "4" button, even though pressing this button is always suboptimal. We suggest that this may be an evolutionary strategy, but it clearly shows that people do not follow the 'rational' Nash strategy. It seems that AV-human interactions will be probabilistic. In AV-AV interactions, software may be playing itself, and may also require probabilistic optimal evolutionary-type strategies. We doubt that the full implications of autonomous decision-making have been fully worked out. Whether probabilistic decisions will tolerated legally and actuarially is doubtful. One way to avoid them would be to allow regulated AV-AV communications, and force software decisions to be deterministic according to some protocol. However, AV-Human interactions seem likely to remain problematic.

Keywords: Rationality · Decision making · Nash equilibrium · Reward · Matching law · Autonomous agents · Evolutionary game theory

Y. Gao et al. (Eds.): TAROS 2017, LNAI 10454, pp. 315–324, 2017.
DOI: 10.1007/978-3-319-64107-2_25

1 Introduction

We are on the cusp of a brave new world of fully autonomous robots including drones, missiles, ships, cars, and software robots. The future is difficult to predict, but driverless cars appear to be imminent and very much in the public eye. Hardly a month goes by without a car manufacturer or a software enterprise announcing their intention to develop an autonomous vehicle (cars) (AV). Initially, AVs will be semi-autonomous, but as technical and legislative issues are sorted out, AVs will become fully autonomous probably in 2020s. The commercial market is enormous with billions of AVs at stake, and competition among manufacturers will be fierce.

At the heart of a fully autonomous agent is the necessity to make decisions autonomously without direct arbitration by a human controller. Decision will be in real-time and have real life consequences, not only economically (cost, energy consumption, time, etc.) but also in terms of human injury. There are two broad categories of decisions: *non-strategic* and *strategic* games. In non-strategic games (sometimes called 'games against nature'), the agent makes a decision based on expected probabilities of outcomes. Any other agents are assumed to act independently. The traditional approach is to 'rationally' choose deterministically the alternative that optimises some decision criterion, such as maximising expected utility or payoff, or minimising maximum loss, etc. Non-strategic decision-making is dominant in low-density traffic where encounters with other vehicles are infrequent. The goal is to navigate the road, avoid obstacles, stop at traffic lights, and generally obey the rules of driving. There is a trade-off between journey time, safety, and risks from violating rules.

In strategic games, a decision needs to take into account the decisions of other agents (human or robot) who simultaneously make decisions based on the agent's expected decision. Such decisions are dominant in high-density traffic where there is contention for road space (slots in a moving queue). Such competition must be tempered with some degree of cooperation amongst drivers to avoid having (or causing) and 'accident'. Competition is most fierce when joining a queue at roundabouts, junctions, slip roads, and lane changes. Competition for road space lead to other frequency effects. A particular route may be the fastest and optimal, but if all drivers select the same route, it may become the slowest and suboptimal. Waiting at re-fuelling (re-charging) stations increases with the number of vehicles. Traditionally, analysis of strategic games comes under the rubric of "Game Theory", where agents are assumed to be rational and fully informed. The optimal decisions attempt to maximise individual gains in a stable way by finding Nash Equilibria (NE), which are the choices for which all agents cannot improve their outcomes (but not necessarily Pareto optimal). Solutions to strategic games may be deterministic, but may also be probabilistic ('mixed strategies'). Probabilistic plays are particularly relevant and intriguing. Should an AV manufacturer program random plays, and if so, how will the legal courts interpret liability in the event of an unlucky outcome?

How to program risk management in an AV is far from clear. Initially, most interactions will be between AVs and human drivers (AV-H interactions). The problem for the AV manufacturer is to be able predict the decisions of human decision-makers contingent on the AV decision options. Do humans follow Nash equilibria? Evidence,

based mostly on the prisoner's dilemma game and the ultimatum game, is mixed. Sone humans tend to be cooperative and do not follow NE, others are more individual and do follow NE. There is also considerable complication in interpreting results from games that are played more than once against the same 'opponent' (iterative games), as opposed to one-shot games. AV-H interactions are one-shot, although similar scenarios may arise with different opponents.

As AVs proliferate, interactions will become increasingly between AVs (AV-AV interactions). AVs with the same manufacturer (model) and software will presumably inherit the same decision making strategy leading to the strange situation where a decision strategy will effectively play itself – reminiscent of evolutionary game theory, where members of a species inherit the same strategy [1].

Nature has been making strategic decisions for eons via natural selection, and one wonders whether we could learn from her. An example is foraging where the gain from competition for resources decreases with number of competitors due to sharing or fighting. This is a simple frequency dependent game. When there are alternative food sources, animals distribute themselves probabilistically across the sources rather than all competing for the same source – called the matching law (ML) [2]. Thus, Nature seems to prefer a probabilistic solution. Some have argued that the ML is an evolutionary stable equilibrium strategy [3–6]. We are not aware of any game-theoretic studies on how humans compete for limited resources. We therefore set up a simple experiment to see how pairs of humans make Game Theoretic decisions in a simulated competition. Of course driving is much more complicated, but such games are simple and directly address the question of whether humans compete or cooperate and make deterministic or probabilistic decisions?

2 Implementation of Foraging Games

We implemented the foraging games in the following way. Two computers were synchronized via Ethernet. On each computer monitor, three buttons were displayed labelled "4", "6", and "12". On each trial, each player was instructed to choose a button to press (via a mouse). If players pressed different buttons they received the corresponding reward of, 4, 6, or 12 points. If both players pressed the same button (a clash), the reward was reduced depending on the type of game, which we call 'SPLIT' and 'ZERO'. Once both players had made a choice, the trial ended, and each player's running total of points was incremented and displayed to the player. Each player could only see their own display and their own total points – not the other player's points. The game was iterated over 100 trials. Eighty psychology undergraduate students were recruited, and randomly allocated to 40 pairs. Each pair played only once, either the SPLIT or ZERO game. The game type was randomly determined at the beginning of the game resulting in 21 SPLIT games and 19 ZERO games.

In the SPLIT game, when players clashed their reward was reduced by a half, receiving 2, 3, or 6 points depending on which button was pressed. Thus a clash is moderately expensive (e.g. sharing food, reduced journey time when same route is chosen). The payoff bimatrix is shown in Table 1. As can be seen for Player A, the maximum gain is maximised by choosing button 12 for any strategy by Player B.

Table 1. Top: Bimatrices of games. Bottom: Nash equilibria.

SPLIT	Player B			
		12	6	4
Player A	12	**(6, 6)**	(12,6)	(12,4)
	6	(6,12)	(3, 3)	(6, 4)
	4	(4,12)	(4, 6)	(2, 2)

ZERO	Player B			
		12	6	4
Player A	12	(0,0)	**(12,6)**	(12,4)
	6	**(6,12)**	(0,0)	(6,4)
	4	(4,12)	(4,6)	(0,0)

Player A				Player B			
4	6	12	PO	4	6	12	PO
0	0	1	12	0	1	0	6
0	1	0	6	0	0	1	12
0	0	1	6	0	0	1	6

Player A				Player B			
4	6	12	PO	4	6	12	PO
0	0	1	12	0	1	0	6
0	1	0	6	0	0	1	12
0	1/3	2/3	6	1/2	0	1/2	4
1/2	0	1/2	4	0	1/3	2/3	6
0	1/3	2/3	4	0	1/3	2/3	4

Similarly, for Player B, the maximum is also to choose 12 (the game is symmetric). Thus, the NE is [12, 12] with payoff (6, 6), which is also Pareto efficient. There are, however, two addition NE: when player A plays 6 and player B plays 12 [6, 12]; and vice versa: Player A plays 12 and player B plays 6 [12, 6]. Clearly, if player B never wavers from the 12 play, player A could also play 6 with same result. If player B does waver from the 12 play, then player A should not play 6, but only 12. It is also obvious that Nash players should never press the 4 button.

In the ZERO game, players received no reward when they pressed the same button. Thus, a clash is very expensive (e.g. fighting and being disabled, choosing the same traffic slot and crashing). In this case there is no dominant strategy but two pure NE at [6, 12] and [12, 6], which are contentious. Both playing 12 at [12, 12] is no longer optimal. There are also 3 additional mixed strategies (probabilistic) that are NE.

3 Results

In both game types (SPLIT and ZERO), the sequential pattern of button presses were highly variable across games (Fig. 1). Within a game, some players were highly variable and seemed to press all three buttons in a haphazard way (Fig. 1a), but others were much more consistent (Fig. 1b) with some pressing the "12" button on almost every trial. Across all games, a player's button press was significantly dependent on the players previous button press and also dependent on the other player's previous button press (χ^2, $p \rightarrow 0$). Thus, players' responses were contingent on the other players'.

We next computed the frequency of button presses for each button for each player and plotted them on triangle plots for comparison with the NE (Fig. 2). The majority of players did not align with an expected NE.

Plotting each player's strategy revealed some distinct patterns: (a) some played "12" mostly, "6" occasionally, and "4" rarely; (b) few played "6" more than "4" or "12"; (c) few played "4" more than "6" or "12"; (d) most played "4", "6" and "12" in

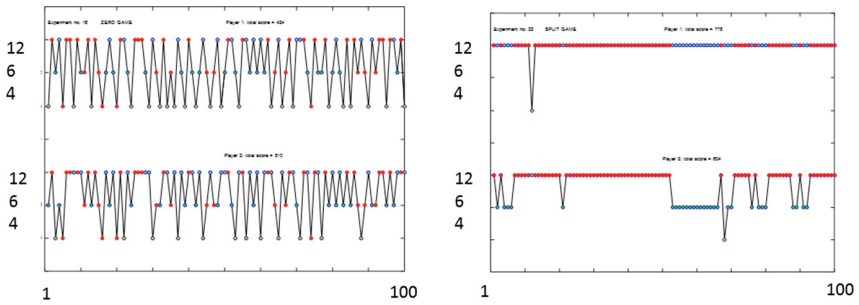

Fig. 1. Two examples of games played. Left: a typical game involving variable play. Note "4" button is frequently pressed. Right: a game with consistent play by one player, and intermediate variability by the other.

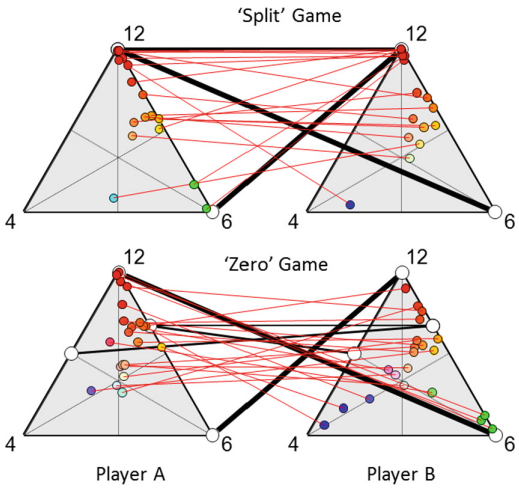

Fig. 2. Triangle plots of proportion of responses for each player A and B in the two game types. Black lines and white circles show Nash equilibria (NE). NE joining vertices are pure (deterministic) strategies; NE from edges are mixed (probabilistic) strategies. Small circles show each player's proportion of button presses for the "4", "6", and "12" buttons; thin lines join players in the same game. Note that most players do not align with NE.

increasing bands (Fig. 2). Patterns a and b were consistent with NE. Pattern c was clearly not consistent with NE. However, Pattern d seemed to approximate the ML with an increase in probability with button value.

3.1 Matching Law

Herrnstein's original matching law relates rate of behaviour to obtained reinforcement. We therefore plotted the proportion of button presses against the actual points awarded

per button press (i.e. taking clashes into consideration) (Fig. 4a). There was a clear linear trend consistent with the ML.

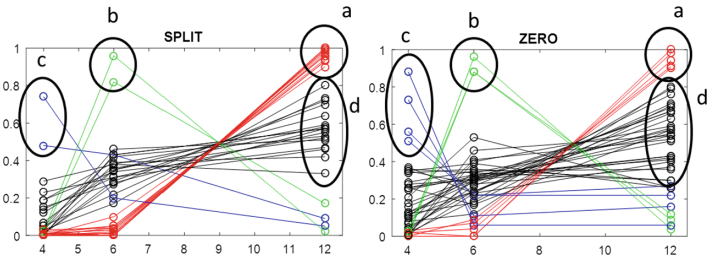

Fig. 3. Categorization of clusters of individual player strategies (see text).

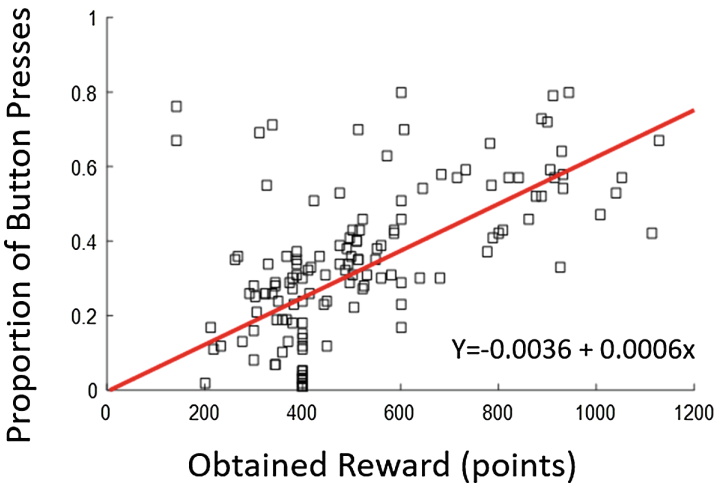

Fig. 4. Plot of proportion of button presses against actual obtained reward per button press for all pattern d responses across both game types (see Fig. 3). Plot is collapsed across the three buttons. Note approximate linear trend as expected from the matching law. Line is linear robust regression.

3.2 Evolutionary Game Theory Equilibria

An important insight can be gleaned from evolutionary game theory (EGT). The basic assumption is that strategies are inherited, and that successful strategies will dominate the gene pool through natural selection (presumably AVs will also inherit from their manufacturers). A consequence is that players will tend to adopt the same strategy. There may actually be a small set of stable strategies, but for the sake of argument, let us assume that players A and B share the same genes and always have the same strategy (Fig. 5). What is their optimal strategy? First consider the SPLIT game (Fig. 6a).

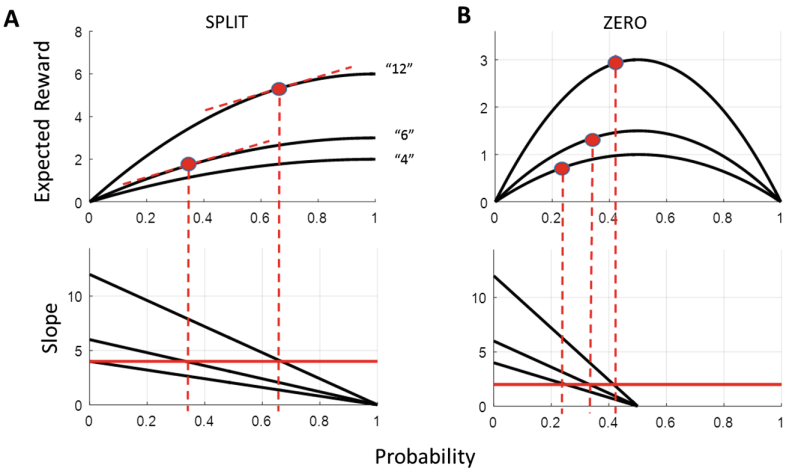

Fig. 5. Schematic to show how an equilibrium can be reached when players adopt the same strategy. At equilibrium, the slope of expected reward (dashed lines) become equal so that switching to a new button offers no advantage. The optimum strategy is then determined since the sum of probabilities (vertical dashed lines) add to unity.

If a player presses the "12" button with increasing probability, the expected payoff will increase as a compressive function such that the slope (rate of increase in reward with probability) decreases. This is because the other player is also pressing "12" with increasing probability. As the probability of playing "12" increases, there comes a point where the players are better off switching to the "6" button because the slope on the "6" button is greater than the "12" button. (This is equivalent to switching habitats in EGT). The slope of the "6" button will also decrease and eventually it will pay to switch to the "4" button. The process will stabilise when the slopes of buttons become the same, since then there is nothing to be gained by switching. Because the sum of probabilities must always add to unity, the final equilibrium point is given by the horizontal line in Fig. 6a. For the SPLIT game, the equilibrium strategy is (0, 0.33, 0.67) (for both players). Thus, it still does not pay to press "4", but the equilibrium point is very close to zero and any fluctuations would involve the "4" button.

For the ZERO game, the equilibrium is different and is (0.25, 0.33, 10/24) (Fig. 6b), and does require "4" presses. These equilibria are optimal but not at a NE. In Fig. 6, these optimal strategies are plotted on dual triangles and compared to observed strategies (pattern d in Fig. 3). They are mixed strategies and similar to, but not precisely the same as, the ideal ML (4/22, 6/22, 12/22). There is some agreement, but it is not perfect especially for the ZERO game. However, there is considerable variability in observed data, and clearly the optimal strategy would depend on how well players could determine their expected payoffs.

This is based on the EGT assumption of identical strategies, which is open to question for human behaviour. However, it demonstrates the key point that when expected reward for each choice is a compressive function (decreasing slope) of probability/frequency of play (Fig. 6), it may pay to switch to a less rewarding choices

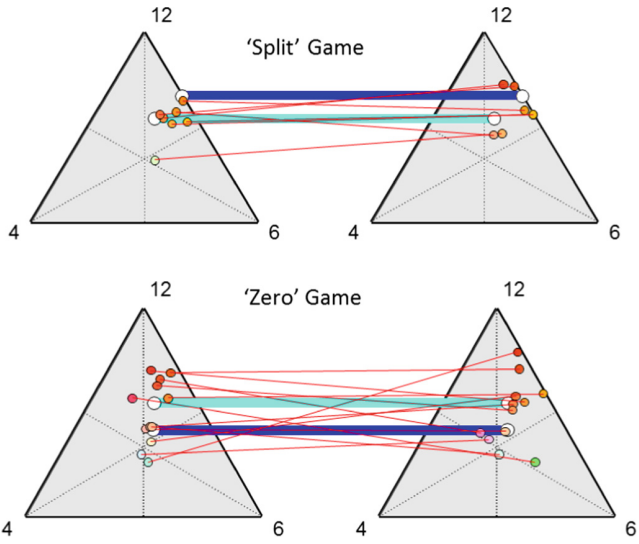

Fig. 6. Optimal strategy when players have same strategy (dark blue line) and the ideal matching law (light blue line) compared to observed players' strategies (for pattern d). (Color figure online)

(depending on their slopes). This will lead to non-NE mixed strategies. It must be emphasised, though, that players do not simply play a fixed mixed strategy independent of the other player, but their choices do depend on the opponent's choices. Thus, a player's expected reward would need to depend on the other player's choices. At present we do not know how players derive expected reward, but if it is based on past experience, then it is plausible that stable compressive functions similar to that in Fig. 3 could emerge. This is a complex problem that we have not yet explored.

4 Discussion

It is clear from this simple experiment that humans do not as a rule adopt NEs. For either the SPLIT or ZERO games, players did not adopt the same strategy, and most pairs of players did not converge on any Nash equilibrium. In the SPLIT game, 5 pairs approximately played the optimum [12, 12] strategy, and 2 pairs approximated the [12, 6] or [6, 12] NE, but 10 pairs appeared to choose a mixed strategy with no NE alignment (Fig. 2) (there are no mixed NE in this game). For the ZERO game, there are two pure NE, [6, 12] and [12, 6], which were approximated by 4 pairs. The remaining pairs, however, chose a mixed strategy. In this game, there are mixed NE (see Table 1); two involved playing the "4", but no pairs adopted these. The other required a mixture of 6 and 12, and it is possible that some pairs approximated this strategy, but we are doubtful as players also approximated this strategy in the SPLIT game which is not a NE. Thus, we conclude that some but most do not adhere to NE.

It could be argued that one-shot NEs are not applicable to iterative games (many trials with the same players), but this is not the case for our games. It is easy to see that for the SPLIT game, playing "12" is always the optimal strategy regardless of the other player's strategy. Playing "6" is risky as the reward could drop to 3 points if the other player also plays "6", and playing "4" is always suboptimal. We also thought that this optimal strategy would be obvious to any player, but evidently this was not the case. In the ZERO game, the optimal strategy is less obvious and requires negotiation between playing "6" and "12": if one player chooses "12", the other should choose "6" and vice versa. So it possible that a player could learn the other's preference and adapt to it or interfere with it. Playing "4" is always suboptimal (for two players).

A few players chose "4" with the highest probability. It is possible that it was strategic if the player assumed that the other player would play "6" or "12" and believed that "4" was the safest option. This would fail, of course, if the other player adopted the same strategy.

In both game types, most players clustered around a mixed strategy with increasing frequency with "4", "6", and "12", although some chose "6" more often than "12". This pattern is reminiscent of the Matching Law, and there is clear trend of a button being pressed increases with the amount of actual reward obtained (Fig. 4a). The ML has long been a contentious issue, and often considered as irrational, or at least a non-maximising strategy.

4.1 Implications for AVs

We need to consider AV-H and AV-AV interactions separately. Based on the results of this experiment with human-human interactions, we cannot assume that a human will act deterministically or even follow a mixed Nash equilibria. Instead, humans appear to adopt probabilistic decision-making at is, an alternative with low expected pay-off is sometimes chosen, but there are individual differences. It seems that Nature prefers probabilistic plays. Should we, therefore, be bioinspired and incorporate such a strategy in decision-making software? There are two problems.

First, we do not know why humans (and animals) are probabilistic. It may be an evolutionary stable strategy, but we cannot be sure. If we assume that it is nevertheless an optimal strategy, there is no guarantee that it would optimal for a man-made AV machine. That is, is it optimal for any decision-making machine or is it peculiar to biological organisms (see Harris [7]). Given that unlucky outcomes are likely to have serious health and financial outcomes, perhaps the gamble of bio-inspiration is a step too far. This brings us to the second problem. A probabilistic strategy will inevitably have unlucky outcomes. Will it be acceptable by the legal system and insurance companies that an 'accident' is perceived to have occurred because of a random number generated in AV software? It is doubtful. The problem for AV risk management software is predicting what a human will do. It will need to make some legally defensible assessment of human behaviour and arrive at a defensible deterministic decision. An AV's speed of processing and response to external events will be much faster than a human's. This may provide some advantage for an AV evading a collision,

but it may influence the ongoing human decision-making process and possibly cause unexpected human behaviour.

A different scenario occurs in AV-AV interactions, which will become increasingly common in the next 20 years. AVs will have the same (or similar) technology and decision-making strategies (presumably depending on the AV brand). Their interactions will inevitably be different from AV-human interactions. Presumably, AVs will need to broadcast their autonomous status so that other AVs can make decisions accordingly. Vehicles that do not broadcast will be assumed to have human drivers. A potential problem arises when AVs make the same decision in a conflict scenario, so that the decision-making software plays against itself, as in evolutionary game theory. It is not possible to predict the outcome at present, but deterministic decisions could be uneconomic as all AVs could make the same error. One way to avoid this scenario would be for AVs to communicate with each other in order to resolve competitive/conflict situations. However, this would need enforcement via some a regulatory body, as seen in air traffic control.

AV-H and AV-AV interactions will be inevitable in the near future. Although currently challenging, it seems likely that autonomous non-strategic driving will become at least as safe as human driving. On the other hand, how autonomous strategic decision-making in high density traffic will evolve remains unclear and could remain persistently problematic until the game - theoretic implications are better understood.

References

1. Maynard Smith, J.: Evolution and the Theory of Games. Cambridge University Press, Cambridge (1982)
2. Herrnstein, R.J.: Relative and absolute strength of responses as a function of frequency of reinforcement. J. Exp. Anal. Behav. **4**, 267–272 (1961)
3. Mailath, G.J.: Do people play Nash equilibrium? Lessons from evolutionary game theory. J. Econom. Lit. **36**, 1347–1374 (1998)
4. Seth, A.K.: The ecology of action selection: insights from artificial life. Philos. T. Roy. Soc. B. **362**, 1545–1558 (2007)
5. Houston, A.I., McNamara, J.M., Steer, M.D.: Do we expect natural selection to produce rational behaviour? Phil. Trans. R. Soc. B **362**, 1531–1543 (2007)
6. Fretwell, S.D., Lucas Jr., H.L.: On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. Acta Biotheor. **19**, 16–36 (1970)
7. Harris, C.M.: Biomimetics of human movement: functional or aesthetic? Bioinspiration Biomimetics **4**, 33001 (2009)