†These authors contributed equally to this study.

Bounded rational response equilibria in human sensorimotor interactions

Cecilia Lindig-León†, Gerrit Schmid† and Daniel A. Braun

Institute of Neural Information Processing, Faculty of Engineering, Computer Science and Psychology, Ulm University, Germany

The Nash equilibrium is one of the most central solution concepts to study strategic interactions between multiple players and has recently also been shown to capture sensorimotor interactions between players that are haptically coupled. While previous studies in behavioural economics have shown that systematic deviations from Nash equilibria in economic decision-making can be explained by the more general quantal response equilibria, such deviations have not been reported for the sensorimotor domain. Here we investigate haptically coupled dyads across three different sensorimotor games corresponding to the classic symmetric and asymmetric Prisoner's Dilemma, where the quantal response equilibrium predicts characteristic shifts across the three games, although the Nash equilibrium stays the same. We find that subjects exhibit the predicted deviations from the Nash solution. Furthermore, we show that taking into account subjects' priors for the games, we arrive at a more accurate description of bounded rational response equilibria that can be regarded as a quantal response equilibrium with non-uniform prior. Our results suggest that bounded rational response equilibria provide a general tool to explain sensorimotor interactions that include the Nash equilibrium as a special case in the absence of information processing limitations.

1. Introduction

Sensorimotor interactions in humans include cooperative examples like carrying a table together across the room or dancing as well as competitive examples like arm wrestling, tug-of-war or playing tennis. There are many different frameworks to study sensorimotor interactions, including the sport sciences [1–3], the psychological sciences [4–18], the neurosciences [19,20] and even engineering when it comes to replicating successful sensorimotor interactions in human–machine interactions [21–24]. Quantitative concepts to study strategic interactions often originate from the decision sciences and include game theory [25–28] and reinforcement learning models [29,30] that were mainly developed in economics and computer science, respectively, but have also found application in studying sensorimotor interactions [24,31–35]. Without a doubt the central solution concept in the decision sciences to understand stable interaction patterns between different agents is the concept of the Nash equilibrium [36]. Roughly, a Nash equilibrium corresponds to a combination of strategies where no agent has anything to gain by deviating unilaterally from their strategy. Abstractly, a strategy can be conceived as a probability distribution over actions, so that Nash equilibria are in general pairs of distributions (mixed equilibria), or in special cases pairs of actions (pure equilibria), when the distributions concentrate their probability mass on a single action.

Previously, it was shown that sensorimotor interactions are amenable to the decision-theoretic framework of Nash equilibria [31]. By designing continuous sensorimotor versions of classic 2 × 2 matrix games like the infamous Prisoner's Dilemma, where players are haptically coupled and experience sensorimotor forces as pay-offs, it could be shown in several studies that human subjects naturally converge to Nash equilibria without verbal descriptions of the game [31–34]. While classic games are typically studied in behavioural

economics in cognitive decision-making tasks with explicitly communicated monetary pay-offs as utilities and clearly defined and known uncertainties, sensorimotor tasks typically involve implicit, action-related utilities such as motor effort or task accuracy and experiential probabilities that have to be learnt from many repetitions. Moreover, motor tasks often involve implicit learning, in contrast to explicit learning. When comparing the results of the sensorimotor games to the corresponding cognitive games, interesting differences can arise, as for example in the Prisoner's Dilemma where sensorimotor interactions regularly converge to the predicted Nash solution of defecting, whereas cognitive versions of the Prisoners' Dilemma regularly lead to some level of cooperation. Other studies have also found interesting differences between economic decision tasks and their equivalent sensorimotor tasks [37]. In particular, it has repeatedly been suggested that human sensorimotor behaviour abides by rational decision-making models [38], whereas for economic studies deviations from rational behaviour have been reported more frequently—although this idea has also been contested [39], and therefore requires further investigation.

While the Nash equilibrium is one of the most successful concepts in the decision sciences, it is also a well-known fact, in particular in behavioural economics, that human behaviour does not always perfectly align with predicted Nash equilibria [40,41]. It is safe to assume that there are multiple reasons for this failure depending on the exact tasks that are investigated, but one prominent reason that has been repeatedly proposed and quantitatively investigated in economic decision-making tasks is bounded rationality [42]. Players that are bounded rational are lacking perfect rationality required to maximize expected utility [43,44] in that they may not know all possible outcomes or the utility functions of the other players, they may have incomplete knowledge, model uncertainty or lack computational resources. One way to model limited information processing capabilities is to assume an information bound on how much players can change an *a priori* agnostic strategy (e.g. a uniform distribution over actions) to an expected utility maximizing strategy [45–48]. In the game-theoretic literature such information bounds on players' strategies have been investigated in the context of quantal response equilibria [49,50], which correspond to Nash equilibria in the unbounded limit, but can otherwise deviate significantly from Nash equilibrium solutions.

In behavioural economics, several studies have confirmed deviations from Nash equilibria in economic decision-making tasks that could be explained by quantal response equilibria [51–53], but so far it is unknown whether similar deviations can be observed in sensorimotor interactions. To this end, we designed three continuous sensorimotor versions of the traditional two-player matrix game of the Prisoner's Dilemma [54], corresponding to the classic symmetric form and two asymmetric variations. Crucially, all three versions of the game have the same single pure Nash equilibrium, but have different quantal response equilibria. In the classic symmetric form of the Prisoner's Dilemma it is assumed that both players can decide to either cooperate or to defect, but that regardless of what the other player decides, defecting is always associated with a better pay-off. The dilemma arises when both players follow this reasoning and end up with a pay-off that is worse compared to a situation where both players cooperate, but cooperation is unstable because each

player can improve their pay-off unilaterally by defecting. Intuitively, in the asymmetric version of the Prisoner's Dilemma, we can imagine that one of the prisoners has some form of weak alibi [55,56], that means that one player has more or less to lose than the other player when deviating from the stable Nash solution. Assuming that players have limited precision when maximizing expected utility due to bounded rationality, the quantal response equilibria in the asymmetric games differ systematically with the extra cost, even though the Nash equilibrium remains the same. This allows for a simple hypothesis test: does behaviour change across the three games or does it stay the same?

2. Results

In our sensorimotor version of the Prisoner's Dilemma, both players were sitting next to each other and used the handles of a robotic interface that each player could move freely in the horizontal plane—see figure 1*a* and electronic supplementary material, methods. During each trial, players were instructed to move their handle to touch the target bar that was projected onto a mirror above the plane of movement. The lateral position of both handles specified the individual magnitude of a resistive force exerted by the robot handles to oppose players' forward motion. Thus, we could induce a haptic coupling where the movement of both players affected the force as a form of pay-off experienced individually by each player in a continuous fashion. By imposing the three Prisoner's Dilemma pay-off landscapes shown in figure 1*b*, player 2 was always exposed to the same force landscape, whereas player 1 experienced more or less costs for deviating from the Nash solution depending on the condition. Accordingly, the quantal response equilibrium predicts a shift in player 1's equilibrium strategy, but not in player 2's, even though the Nash equilibrium is the same for all three games. To allow for learning of the different force landscapes, a particular haptic coupling was kept for a set of 40 trials. For our analysis in the following, we therefore focus on trajectory endpoints over such sets of 40 trials, where players can co-adapt.

Figure 2*a* shows a scatter plot of players' endpoint positions of the last 30 trials of all trial blocks in all three Prisoner's Dilemma games in relation to the Nash equilibrium at the position (1, 1) in the top right corner. The observed spread in the scatter plots suggests that behaviour differs across the three games. In particular, in the asymmetric high cost condition, deviations from the Nash equilibrium seem to be less than in the asymmetric low cost condition, where endpoints spread more freely in the entire plane. This impression from the scatter plots is confirmed when looking at the two-dimensional histograms in figure 2*b* that show a steeper increase in response frequencies for the asymmetric high cost condition and a shallower increase for the asymmetric low cost condition. This shift in the distribution across conditions can be quantified when determining the mean response of player 1 and player 2 from figure 2*a* and comparing it across conditions—see electronic supplementary material, figure S3. In particular, we find that the shift in player 1's strategy across conditions is highly significant ($p < 0.001$, rm ANOVA, $F = 24.86$, $d_1 = 2$, $d_2 = 14$), while player 2's strategy remains the same ($p > 0.1$, rm ANOVA, $F = 0.84$, $d_1 = 2$, $d_2 = 14$).

To analyse the results in terms of the classic Prisoner's Dilemma responses, we categorized the endpoints in figure 2*a*
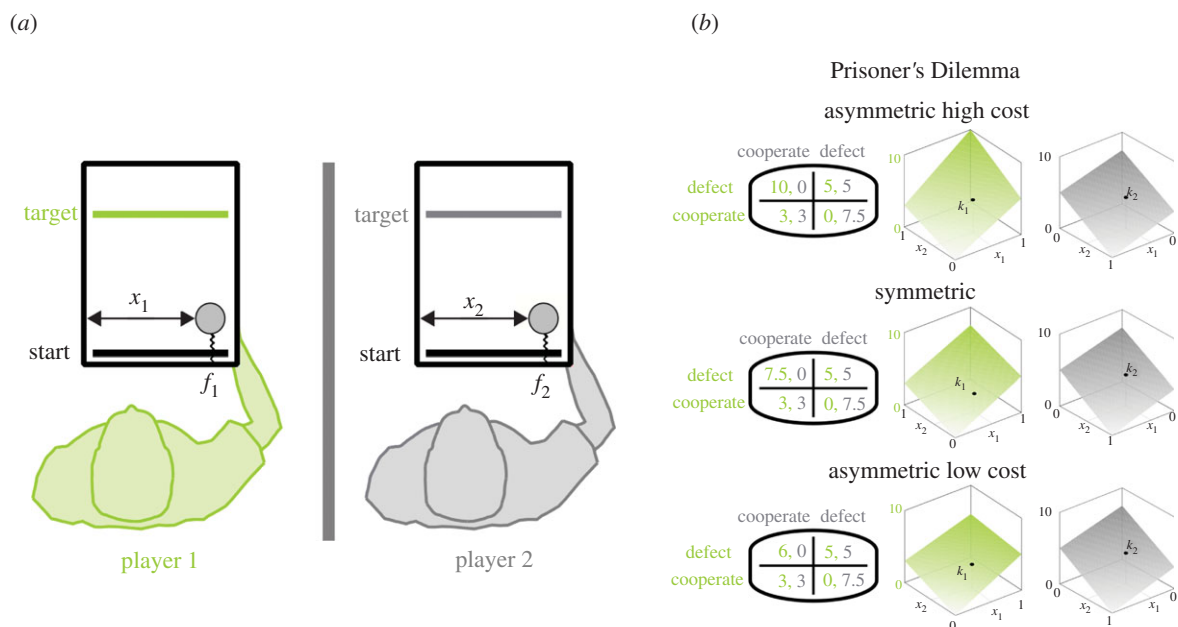
Figure 1. Experimental set-up. (*a*) Sensorimotor game. Players were haptically coupled by two handles that generated a force on their hand resisting their forward movement. Importantly, each player's force depended on both players' lateral positions. (*b*) Pay-off matrices for each one of the three games including pay-offs for player 1 and player 2. The matrix values establish the boundary values of the force pay-offs at the limits of the $x_1 x_2$-space (left). Interpolation of the pay-offs defines a continuous pay-off landscape for a continuum of actions (right). (Online version in colour.)

according to their quadrants into Nash responses (defect–defect), cooperative responses (cooperate–cooperate) and exploitative responses (defect–cooperate or cooperate–defect). As can be seen in figure 2*c*, the Nash solution is predominant in all cases, although the exploitative response (cooperate–defect) becomes increasingly common across conditions as player 1's costs for deviating from the Nash equilibrium decreases. This change in strategy can also be seen in individual subject pairs in electronic supplementary material, figure S1, where the predominance of the Nash equilibrium response decreases across conditions.

To quantify this strategy shift in the probability space, we can determine how each player's response frequency λ, to choose the defect action, changes across conditions. As illustrated in figure 2*d*–*f*, player 1's shift in response frequency λ is highly significant ($p < 0.001$, rm ANOVA, $F = 21.67$, $d_1 = 2$, $d_2 = 14$), while player 2's strategy remains the same ($p > 0.1$, rm ANOVA, $F = 0.48$, $d_1 = 2$, $d_2 = 14$). The response shift can also be observed at the level of individual pairs, although with higher variability—see electronic supplementary material, figure S2. In summary, we can conclude that the observed behavioural change across the three games is incompatible with the Nash equilibrium prediction of no change.

To see whether the observed deviations from the Nash equilibrium are consistent with the quantal response equilibrium hypothesis assuming limited information processing capabilities, we can compare the marginal distributions over responses of each player to the equilibrium distribution equations (2) and (3) in the electronic supplementary material predicted by the quantal response equilibrium when adjusting the precision parameters to fit the behaviour of subjects. Figure 3*a* shows the frequencies over players' responses in all games independent of the other player, separate for player 1 and player 2. As the cost structure for player 2 does not change, the quantal response equilibrium predicts that there should be no changes in the response distribution of player 2, which is in accordance with our data. For player 1, the quantal

response equilibrium predicts that in the asymmetric high cost condition, response frequencies should be elevated near the Nash equilibrium, but decreased further away from the Nash equilibrium compared to the classic symmetric Prisoner's Dilemma. Similarly, in the asymmetric low cost condition, the response frequencies around the Nash equilibrium should be suppressed, and instead elevated further away from the Nash solution. This prediction is confirmed, too, by the trends in the data. The only observation that is not predicted by the quantal response equilibrium is a border effect for player 1, where the response frequency in the direct neighbourhood of the Nash equilibrium declines sharply. However, such a border effect can be taken into account in the quantal response equilibrium model when considering non-uniform priors for the response distribution. To fit the players' empirical priors, we obtained a histogram over initial positions in the first trials of each block of 40 trials—figure 3*b*. It can be seen that player 1 assigns lower probabilities to the corners of the workspace and concentrates probability mass in the centre, whereas player 2 has a more uniform prior distribution. Taking into account these priors, the bounded rational equilibrium fits the data significantly better—compare figure 3*c*. The predicted mean of these fitted equilibrium distributions is also in good agreement with the data, as shown in electronic supplementary material, figure S3. In terms of the categorical response frequencies, the quantal response equilibrium predicts an up-shift for the asymmetric high cost condition and a down-shift for the asymmetric low cost condition for player 1, again in good agreement with the data—compare figure 2*d*.

To investigate further how the equilibrium distributions are reached over time, figure 4*a* shows how the categorical response frequencies change on average over a course of 40 trials. For both players, the initial probability to select either defect or cooperate is 0.5. Over the next 10 trials, this probability gets biased towards the defect action. Crucially, the resulting learning curves for the three different games for player 2 are identical, whereas for player 1 learning is sharper for the asymmetric high cost
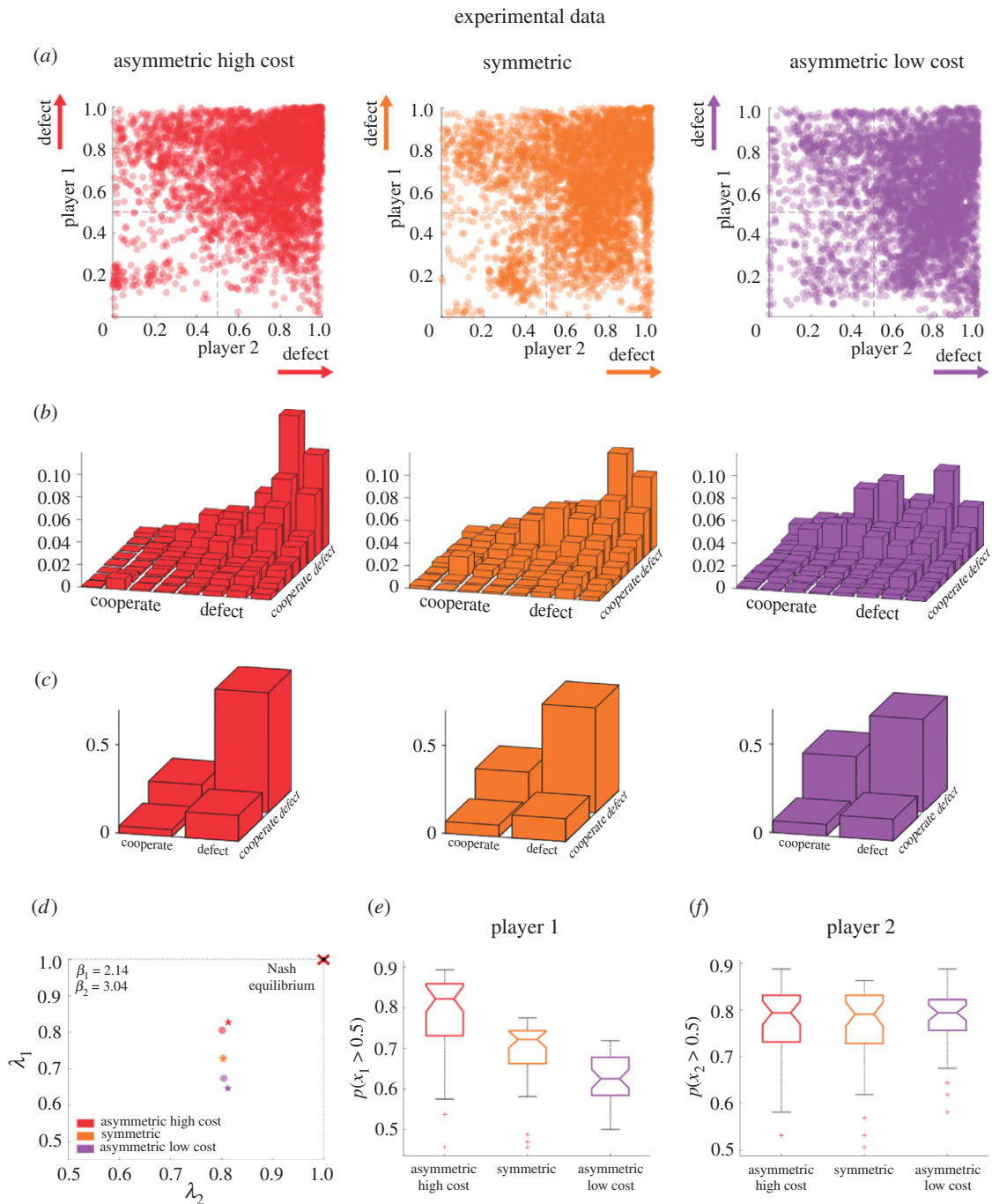
experimental data

Figure 2. Experimental data. (*a*) Scatter plots of the final decisions of the last 30 trials in the x_1x_2-plane for the three Prisoner's Dilemma games, where subjects' actions are expected to cluster around the single pure Nash equilibrium located in the top-right corner at position (1,1). (*b*) Two-dimensional histograms of the experimental scatter plots. (*c*) Final decisions with binary categorization. (*d*) Both players' response frequency λ calculated from the categorized actions of the last 30 trials. The experimental data are shown with stars, whereas the circles represent the quantal response equilibria fitted to the data according to equations (6) and (7) with $\beta_1 = 2.14$ and $\beta_2 = 3.04$. (*e*) Player 1's response frequency of defecting across game conditions. The central marks indicate the median and the edges of the box are the 25th and 75th percentiles considering all trials and subjects. (*f*) Box plots for player 2's response frequency across conditions. (Online version in colour.)

condition and flatter for the asymmetric low cost condition. Accordingly, player 1 is more indifferent between cooperating and defecting in the asymmetric low cost condition, and more prone to defect in the asymmetric high cost condition, which can also be seen in the temporal evolution of the combined choices—see electronic supplementary material, figure S4. When simulating these learning curves with a pair of continuous reinforcement learning agents employing Q-learning as defined in equation (8) in the electronic supplementary material, this pattern of differentiated learning curves for player 1 across the three games is reproduced—figure 4*b*.

When analysing the behaviour of the simulated reinforcement learners in the same way as the human players, we can see in the scatter plots the same trend in that the asymmetric high cost condition is more concentrated towards the Nash equilibrium than the asymmetric low cost condition—figure 5*a*. As for the human subjects, the two-dimensional histograms show a steeper increase in response frequencies for the asymmetric high cost condition and a shallower increase for the asymmetric low cost condition—figure 5*b*. Finally, when categorizing the responses into (defect–defect), (cooperate–cooperate) and (defect–cooperate or
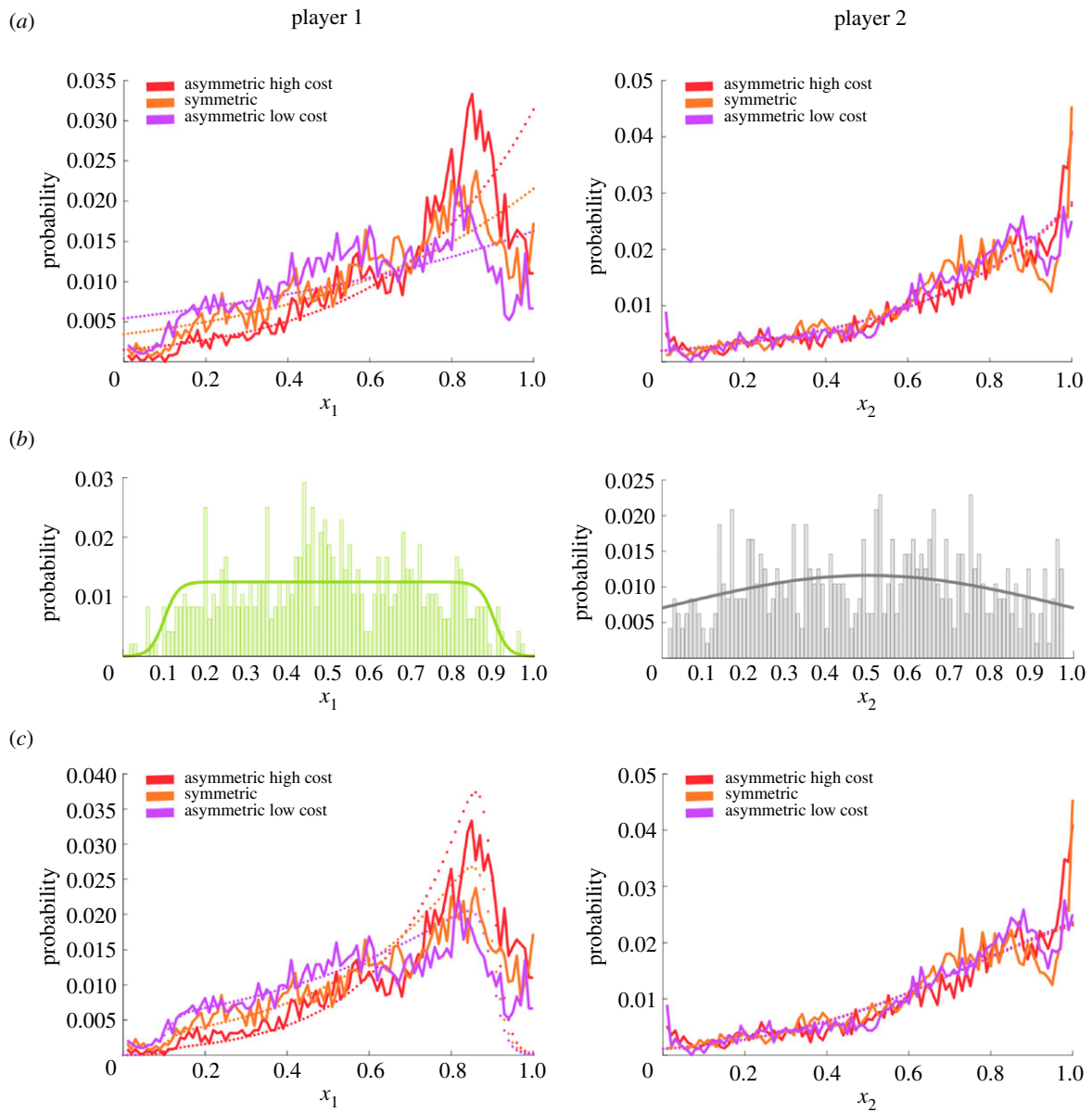
Figure 3. Comparison of the marginal distributions over responses of each player to the equilibrium distributions predicted by the quantal response equilibrium. (a) The empirical frequencies of players' responses (solid lines) are compared to the quantal response equilibrium (dash lines) predicting no changes in the response distribution of player 2 and for player 1 elevated response frequencies near the Nash equilibrium in the asymmetric high cost condition, and accordingly suppressed response frequencies around the Nash equilibrium in the asymmetric low cost condition. (b) Players' priors represented by histograms over initial positions in the first trials of each block of 40 trials. The solid lines represent simple analytic representations of the priors. (c) Quantal response equilibrium considering players' priors (dashed lines) compared to the same data (solid lines). (Online version in colour.)

cooperate–defect), it can be seen that the predominant Nash solution becomes more frequent in the asymmetric high cost condition, and less frequent in the asymmetric low cost condition compared to the classic symmetric game—compare figure 5c. The corresponding shifts for the response frequencies of player 1 reproduce the same pattern as observed in the human players. This suggests that reinforcement learning models based on Q-learning cannot only explain convergence to Nash equilibrium solutions [57], but more generally convergence to quantal response equilibria.

3. Discussion

In this study, we have investigated the concept of quantal response equilibria in human multi-agent interactions in a continuous sensorimotor version of the symmetric and asymmetric

Prisoner's Dilemma. In particular, we have tested the hypothesis that quantal response equilibria may provide a more accurate description of stable states of human interaction than the prevailing Nash solution concept. During the interactions, subjects were haptically coupled and learned to avoid the haptic coupling force opposing their forward motion, signifying the pay-off for the interaction. In previous studies, it was found that such haptic couplings between two different players in the Prisoner's Dilemma are compatible with the Nash solution, as most interaction endpoints laid in the same quadrant of the workspace than the Nash equilibrium [31]. Similar analyses have also advocated the adequacy of the Nash solution concept for describing sensorimotor interactions in more general scenarios, including mixed equilibrium games like matching pennies [57], coordination games with multiple Nash equilibria like the battle of sexes, chicken or stag hunt [32] as well as Bayesian games that require sensorimotor communication [34].
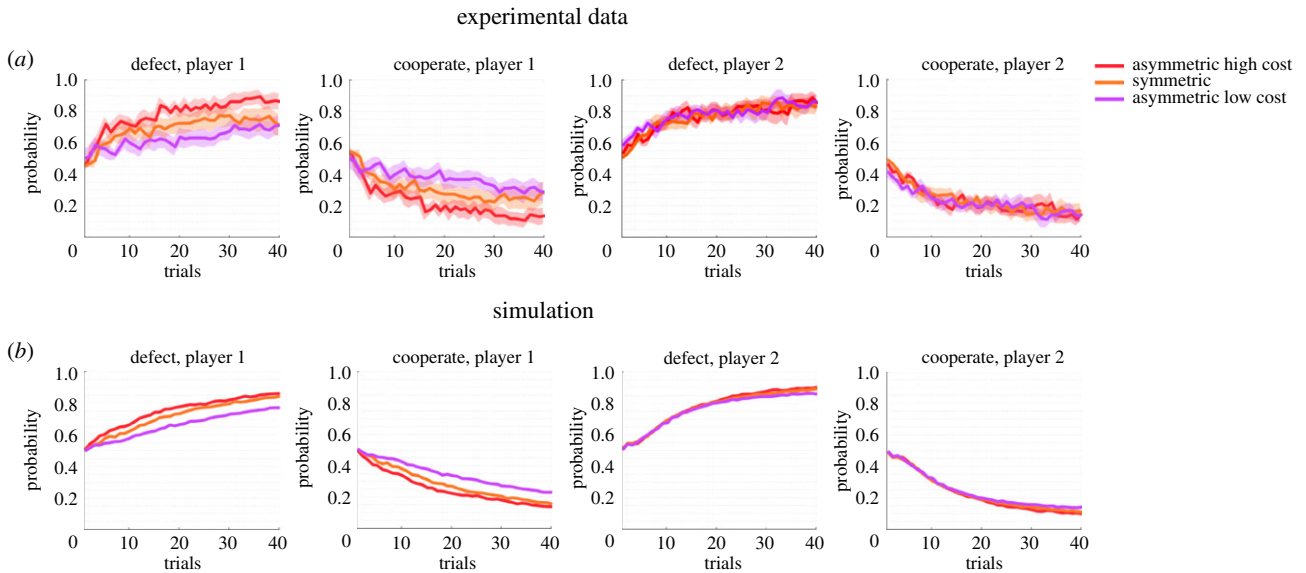
Figure 4. Temporal evolution of players' responses. (*a*) Evolution of each player's probability of defecting and cooperating across a set of 40 trials. The shading indicates one standard error of the mean across subjects. (*b*) Evolution of the probability of defecting and cooperating for two simulated Q-learning players.
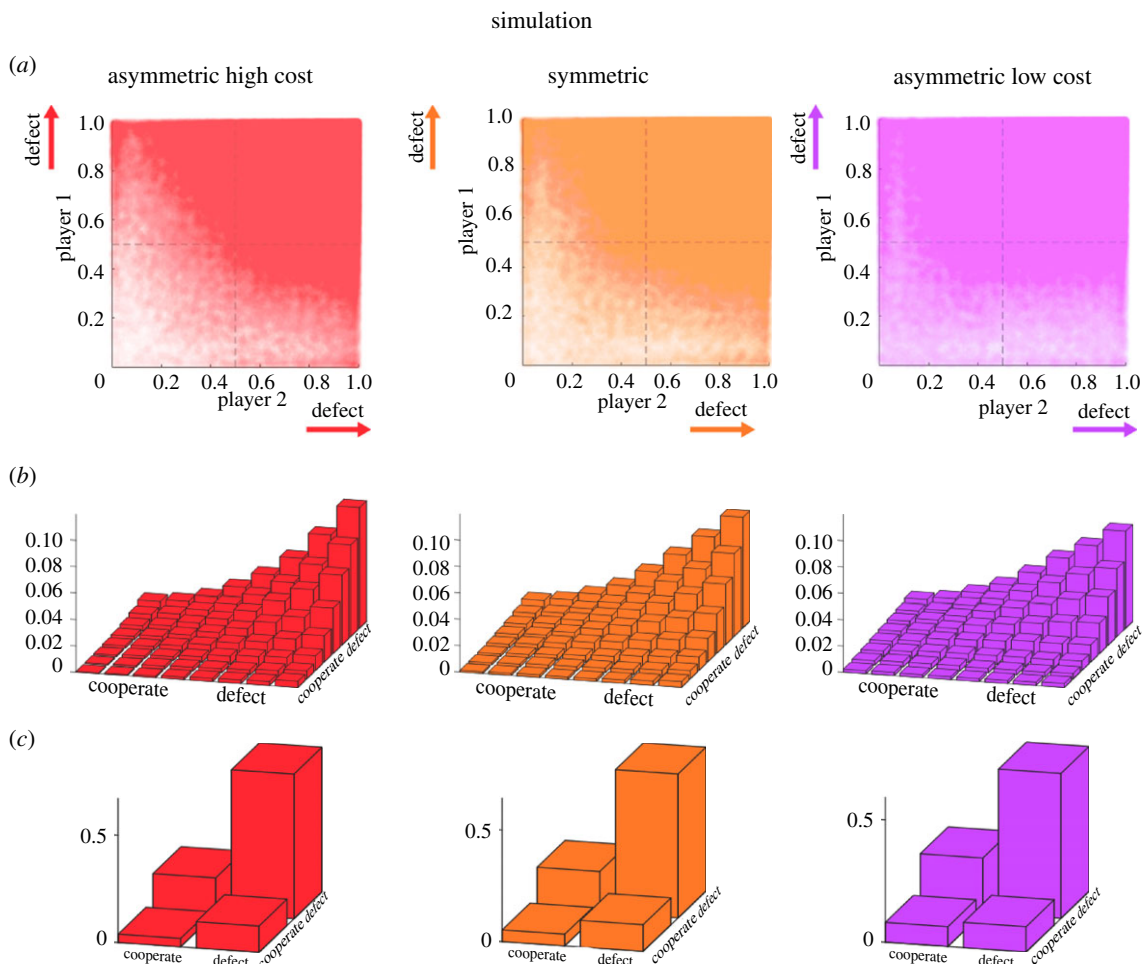

Figure 5. Simulated Q-learners. (*a*) Scatter plots of the final decisions in the x_1x_2-plane for the three games, where actions are expected to cluster around the single pure Nash equilibrium located in the top-right corner at position (1,1). (*b*) Two-dimensional histograms of the simulated scatter plots. (*c*) Final decisions in terms of binary categorizations. (Online version in colour.)

Importantly, none of the above studies could distinguish the Nash solution from the quantal response equilibrium, as the two solution concepts are often very close together and perfectly coincide in the absence of computational or precision limits. Accordingly, we have designed a sensorimotor interaction game based on three different 2 × 2-matrix games

corresponding to the classic symmetric form and two asymmetric forms of the Prisoner's Dilemma, thus, allowing for the prediction of a response shift for player 1 in case of the quantal response equilibrium and no such shift in case of the Nash equilibrium. Our results are clearly compatible with the predicted shift and incompatible with the no-shift

prediction of the Nash equilibrium. As the quantal response equilibrium can be seen as a generalization of the Nash equilibrium that contains the Nash solution as a special case in the limit of perfect rationality [58], our results suggest that the quantal response equilibrium should be seen as the more general concept to understand sensorimotor interaction, albeit it will often coincide with the corresponding Nash equilibrium solution.

The most common specification of the quantal response equilibrium model is based on softmax strategies with a single precision parameter [49]. The interpretation of softmax- or logit-strategies in terms of bounded rational choice rules with limited precision in terms of a trade-off between pay-off and entropy has also put the quantal response equilibrium at the heart of bounded rational game theory [58–64]. Extending this trade-off by including prior strategies, bounded rational choice can be described by Boltzmann-like distributions like equations (4) and (5) in the electronic supplementary material, where the individual precision parameter quantifies how much a player is able to deviate from their prior towards a utility maximizing strategy [46,65]. We found in our study that considering players' priors significantly improves the fit predicted by the quantal response equilibrium, especially near the boundaries. Importantly, this is not a result from overfitting by assuming arbitrary priors, but we extracted priors experimentally from the distributions over initial positions at the beginning of each block of trials. This also gives further credence to a growing body of literature that uses utility-information trade-offs to model bounded rational decision-making in the sensorimotor context, emphasizing the role of the prior in such trade-offs [47].

Like the Nash equilibrium, bounded rational response equilibria are defined as fixed points and do not detail any mechanism regarding the decision-making and learning processes that ultimately converge to these fixed points. In our study we have used a continuous Q-learning model with basis functions that was able to reproduce the predicted shifts and the convergence to the quantal response equilibrium. The Q-learner played all three games with the same parameter set. Since the Q-learner is also based on a softmax strategy, it naturally reproduces the predictions of the quantal response equilibrium, because the action probabilities are biased by more or less pay-off in the asymmetric conditions. To study the learning curves we focused on the change of final positions across trials, since we found that initial and final positions within trajectories were generally close. Specifically, we found that in more than 77% of the trials for the symmetric Prisoner's Dilemma, and more than 77% and 78% of the trials for the high and low asymmetric versions, respectively, the players' final decision laid within a 1.6 cm neighbourhood (10% of the workspace) of their initial position in each trial, and there was no systematic change over the block of trials. This suggests that adaptation processes during the trial only had a minor effect and could be neglected.

Our study is part of a broader family of studies that investigate differences between decision-making in sensorimotor tasks and cognitive tasks [66]. The asymmetric Prisoner's Dilemma has been previously investigated in a number of studies where subjects were told the pay-off matrices and they had to make deliberate decisions over the course of a set of repeated games [67–70]. Usually, the aim of the studies was to investigate the effect of the asymmetry on the propensity for cooperation. The results of the studies are difficult to compare due to substantial variations in experimental design, as there are many different ways of introducing asymmetry, for example, affecting all entries or only some entries in the pay-off matrix, or where one player has consistently higher payoffs than the other, or mixed designs, etc. Nevertheless, many of the studies suggest that asymmetry makes reasoning in the game more difficult, and report lower rates of cooperation in asymmetric pay-off conditions [67–70]. In contrast, in our sensorimotor games the frequency of the cooperative solution (cooperate–cooperate) is not modulated by the asymmetry— compare electronic supplementary material, figure S4. Instead, an increase or decrease in the prevalence of the Nash solution (defect–defect) is accompanied by a corresponding decrease or increase in exploitative (cooperate–defect) responses, where player 1 cooperates more or less depending on the asymmetry condition. This is not only in line with the quantal response equilibrium prediction, but also agrees with previous experimental results [31] that have shown that cooperation does not arise as a stable solution during haptic coupling, but only in cognitive versions of the Prisoner's Dilemma involving conscious deliberation. Our current study adds to this previous line of research by highlighting the importance of taking into account limited information processing capabilities due to bounded rationality and how to capture these using information constraints [45–48].

References

1. Sindik J, Vidak N. 2008 Application of game theory in describing efficacy of decision making in sportsman's tactical performance in team sports. *Interdiscip. Descrip. Complex Systems Sci. J.* **6**, 53–66.

2. Bockel A. 2015 *The golden rule in sports: investing in the conditions of cooperation for a mutual advantage in sports competitions*. Berlin, Germany: Springer VS.

3. Rico-Gonzalez M, Ortega JP, Nakamura F, Moura F, Arcos AL. 2020 Origin and modifications of the geometrical centre to assess team behaviour in team sports: a systematic review. *RICYDE. Revista*

Internacional de Ciencias del Deporte **16**, 318–329. (doi:10.5232/ricyde2020.06106)

4. Sebanz N, Knoblich G, Prinz W. 2003 Your task is my task: shared task representations in dyadic interactions. In _Proc. Annual Meeting of the Cognitive Science Society, 31 July - 2 August, Boston, MA_, vol. 25. Hove, UK: Psychology Press.

5. Sebanz N, Bekkering H, Knoblich G. 2006 Joint action: bodies and minds moving together. _Trends Cogn. Sci._ **10**, 70–76. (doi:10.1016/j.tics.2005.12.009)

6. Pezzulo G. 2011 Shared representations as coordination tools for interaction. _Rev. Philos. Psychol._ **2**, 303–333. (doi:10.1007/s13164-011-0060-5)

7. Pezzulo G, Donnarumma F, Dindo H. 2013 Human sensorimotor communication: a theory of signaling in online social interactions. _PLoS ONE_ **8**, 1–11. (doi:10.1371/journal.pone.0079876)

8. Reddish P, Fischer R, Bulbulia J. 2013 Let's dance together: synchrony, shared intentionality and cooperation. _PLoS ONE_ **8**, 1–13. (doi:10.1371/journal.pone.0071182)

9. Zahavi D, Glenda S. 2015 _Beyond the analytic-continental divide: varieties of shared intentionality: tomasello and classical phenomenology_. London, UK: Routledge.

10. Candidi M, Curioni A, Donnarumma F, Sacheli LM, Pezzulo G. 2015 Interactional leader-follower sensorimotor communication strategies during repetitive joint actions. _J. R Soc. Interface_ **12**, 20150644. (doi:10.1098/rsif.2015.0644)

11. Bolt NK, Poncelet EM, Schultz BG, Loehr JD. 2016 Mutual coordination strengthens the sense of joint agency in cooperative joint action. _Conscious Cogn._ **46**, 173–187. (doi:10.1016/j.concog.2016.10.001)

12. Heesen R, Genty E, Rossano F, Zuberbühler K, Bangerter A. 2017 Social play as joint action: a framework to study the evolution of shared intentionality as an interactional achievement. _Learn. Behav._ **45**, 390–405. (doi:10.3758/s13420-017-0287-9)

13. Vesper C _et al._ 2017 Joint action: mental representations, shared information and general mechanisms for coordinating with others. _Front. Psychol._ **7**, 2039. (doi:10.3389/fpsyg.2016.02039)

14. Lyre H. 2018 Socially extended cognition and shared intentionality. _Front. Psychol._ **9**, 831. (doi:10.3389/fpsyg.2018.00831)

15. Dotov DG, de Cock VC, Geny C, Ihalainen P, Moens B, Leman M, Bardy B, Bella SD. 2019 The role of interaction and predictability in the spontaneous entrainment of movement. _J. Exp. Psychol. Gen._ **148**, 1041–1057. (doi:10.1037/xge0000609)

16. Makowski PT. 2020 Shared intentionality and automatic imitation: the case of _La Ola_. _Philos. Soc. Sci._ **50**, 465–492. (doi:10.1177/0048393120918302)

17. Scharpf FW. 1990 Games real actors could play: the problem of mutual predictability. _Rational. Soc._ **2**, 471–494. (doi:10.1177/1043463190002004005)

18. Sebanz N, Knoblich G. 2021 Progress in joint-action research. _Curr. Dir. Psychol. Sci._ **30**, 138–143. (doi:10.1177/0963721420984425)

19. Ménoret M, Varnet L, Fargier R, Cheylus A, Curie A, des Portes V, Nazir TA, Paulignan Y. 2014 Neural correlates of non-verbal social interactions: a dual-EEG study. _Neuropsychologia_ **55**, 85–97. (doi:10.1016/j.neuropsychologia.2013.10.001)

20. Pezzulo G, Donnarumma F, Dindo H, D'Ausilio A, Konvalinka I, Castelfranchi C. 2019 The body talks: sensorimotor communication and its brain and kinematic signatures. _Phys. Life Rev._ **28**, 1–21. (doi:10.1016/j.plrev.2018.06.014)

21. Medina JR, Lee D, Hirche S. 2012 Risk-sensitive optimal feedback control for haptic assistance. In _2012 IEEE Int. Conf. on Robotics and Automation, 14–19 May, St Paul, MN_, pp. 1025–1031. Piscataway, NJ: IEEE.

22. Ros R, Baroni I, Demiris Y. 2014 Adaptive human-robot interaction in sensorimotor task instruction: from human to robot dance tutors. _Robot. Autonomous Syst._ **62**, 707–720. (doi:10.1016/j.robot.2014.03.005)

23. Sawers A, Ting LH. 2014 Perspectives on human-human sensorimotor interactions for the design of rehabilitation robots. _J. Neuroeng. Rehabil._ **11**, 142. (doi:10.1186/1743-0003-11-142)

24. Chackochan VT, Sanguineti V. 2017 Modelling collaborative strategies in physical human-human interaction. In _Converging clinical and engineering research on neurorehabilitation II_, vol. 15, pp. 253–258. Berlin, Germany: Springer.

25. Martin J. 1994 _Osborne and ariel rubinstein. A course in game theory, volume 1 of MIT press books_. Cambridge, MA: The MIT Press.

26. Başar T, Olsder GJ. 1998 _Dynamic noncooperative game theory_, vol. 160. Philadelphia, PA: SIAM.

27. Camerer CF. 2003 Behavioural studies of strategic thinking in games. _Trends Cogn. Sci._ **7**, 225–231. (doi:10.1016/S1364-6613(03)00094-9)

28. Griessinger T, Coricelli G. 2015 The neuroeconomics of strategic interaction. _Curr. Opin. Behav. Sci._ **3**, 73–79. (doi:10.1016/j.cobeha.2015.01.012)

29. Watkins Peter CJCH. 1992 Dayan. Q-learning. _Machine Learning_ **8**, 279–292. (doi:10.1007/BF00992698)

30. Sutton RS, Barto AG. 2018 _Reinforcement learning: an introduction_. Cambridge, MA: MIT press.

31. Braun DA, Ortega PA, Wolpert DM. 2009 Nash equilibria in multi-agent motor interactions. _PLoS Comput. Biol._ **5**, e1000468. (doi:10.1371/journal.pcbi.1000468)

32. Braun DA, Ortega PA, Wolpert DM. 2011 Motor coordination: when two have to act as one. _Exp. Brain Res._ **211**, 631–641. (doi:10.1007/s00221-011-2642-y)

33. Grau-Moya J, Hez E, Pezzulo G, Braun DA. 2013 The effect of model uncertainty on cooperation in sensorimotor interactions. _J. R. Soc. Interface_ **10**, 20130554. (doi:10.1098/rsif.2013.0554)

34. Leibfried F, Grau-Moya J, Braun DA. 2015 Signaling equilibria in sensorimotor interactions. _Cognition_ **141**, 73–86. (doi:10.1016/j.cognition.2015.03.008)

35. Schmid G, Braun DA. 2020 Human group coordination in a sensorimotor task with neuron-like decision-making. _Sci. Rep._ **10**, 8226. (doi:10.1038/s41598-020-64091-4)

36. Nash JF. 1950 Equilibrium points in n-person games. _Proc. Natl Acad. Sci. USA_ **36**, 48–49. (doi:10.1073/pnas.36.1.48)

37. Wu S-W, Delgado MR, Maloney LT. 2009 Economic decision-making compared with an equivalent motor task. _Proc. Natl Acad. Sci. USA_ **106**, 6088–6093. (doi:10.1073/pnas.0900102106)

38. Körding K. 2007 Decision theory: what 'should' the nervous system do? _Science (New York, N.Y.)_ **318**, 606–610. (doi:10.1126/science.1142998)

39. Jarvstad A, Hahn U, Rushton SK, Warren PA. 2013 Perceptuo-motor, cognitive, and description-based decision-making seem equally good. _Proc. Natl Acad. Sci. USA_ **110**, 16 271–16 276. (doi:10.1073/pnas.1300239110)

40. Goeree JK, Holt CA. 1999 Stochastic game theory: for playing games, not just for doing theory. _Proc. Natl Acad. Sci. USA_ **96**, 10 564–10 567. (doi:10.1073/pnas.96.19.10564)

41. Bonau S. 2017 A case for behavioural game theory. _J. Game Theory_ **6**, 7–14. (doi:10.5923/j.jgt.20170601.02)

42. Rubinstein A. 1998 _Modeling bounded rationality_. New York, NY: MIT Press.

43. Von Neumann J, Morgenstern O. 1944 _Theory of games and economic behavior_. Princeton, NJ: Princeton University Press.

44. Savage LJ. 1972 _The foundations of statistics_. Mineola, NY: Dover Publications Inc.

45. Tishby N, Polani D. 2011 Information theory of decisions and actions. In _Perception-action cycle: models, architectures, and hardware_ (eds Vassilis Cutsuridis, Amir Hussain, John G. Taylor), pp. 601–636. New York, NY: Springer.

46. Ortega PA, Braun DA. 2013 Thermodynamics as a theory of decision-making with information-processing costs. _Proc. R. Soc. A_ **469**, 20120683. (doi:10.1098/rspa.2012.0683)

47. Gottwald S, Braun DA. 2020 The two kinds of free energy and the bayesian revolution. _PLoS Comput. Biol._ **16**, 1–32. (doi:10.1371/journal.pcbi.1008420)

48. Bhui R, Lai L, Gershman S. 2021 Resource-rational decision making. _Curr. Opin. Behav. Sci._ **41**, 15–21. (doi:10.1016/j.cobeha.2021.02.015)

49. McKelvey RD, Palfrey TR. 1995 Quantal response equilibria for normal form games. _Games Econ. Behav._ **10**, 6–38. (doi:10.1006/game.1995.1023)

50. Mckelvey RD, Palfrey TR. 1998 Quantal response equilibria for extensive form games. _Exp. Econ._ **1**, 9–41. (doi:10.1023/A:1009905800005)

51. Goeree JK, Holt CA, Palfrey TR. 2003 Risk averse behavior in generalized matching pennies games. _Games Econ. Behav._ **45**, 97–113. (doi:10.1016/S0899-8256(03)00052-6)

52. Goeree JK, Holt CA, Palfrey TR. 2005 Regular quantal response equilibrium. _Exp. Econ._ **8**, 347–367. (doi:10.1007/s10683-005-5374-7)

53. Goeree JK, Palfrey TR, Holt CA. 2010 _Behavioral and experimental economics: quantal response equilibria_. London, UK: Palgrave Macmillan.

54. Rapoport A, Chammah AM, Orwant CJ. 1965 *Prisoner's dilemma: a study in conflict and cooperation*, vol. 165. London, UK: University of Michigan Press.

55. Robinson DR, Goforth DJ. 2004 Alibi games: the asymmetric prisoner' s dilemmas. In *Meetings of the Canadian Economics Association, 4–6 June, Toronto, Canada*. Oxford, UK: Blackwell.

56. Robinson D, Goforth D. 2005 *The topology of the 2(2 games: a new periodic table*, vol. 3. London, UK: Routledge.

57. Lindig-León C, Schmid G, Braun DA. 2021 Nash equilibria in human sensorimotor interactions explained by Q-learning with intrinsic costs. *Sci. Rep.* **11**, 20779. (doi:10.1038/s41598-021-99428-0)

58. Wolpert DH. 2006 *Information theory—The bridge connecting bounded rational game theory and statistical physics*, pp. 262–290. Berlin, Germany: Springer.

59. Simon HA. 1956 Rational choice and the structure of the environment. *Psychol. Rev.* **63**, 129–138. (doi:10.1037/h0042769)

60. Simon H. 1972 *Decision and organization: theories of bounded rationality*. Amsterdam, The Netherlands: North-Holland.

61. Simon HA *et al.* 1984 *Models of bounded rationality, volume 1: economic analysis and public policy.* Cambridge, MA: MIT Press Books.

62. Horvitz E. 1988 Reasoning under varying and uncertain resource constraints. In *Proc. of the Seventh National Conf. on Artificial Intelligence, Minneapolis, MN, August*, pp. 111–116. San Mateo, CA: Morgan Kaufmann Publishers Inc. See http://www.aaai.org/Papers/AAAI/1988/AAAI88-020.pdf.

63. Horvitz EJ, Cooper GF, Heckerman DE. 1989 Reflection and action under scarce resources: theoretical principles and empirical study. In *Proc. Eleventh Int. Joint Conf. on Artificial Intelligence, 20–25 August, Detroit, MI*, pp. 1121–1127. San Mateo, CA: Morgan Kaufmann Publishers Inc.

64. Horvitz E, Zilberstein S. 2001 Computational tradeoffs under bounded resources. *Artif. Intell.* **126**, 1–4. (doi:10.1016/S0004-3702(01)00051-0)

65. Genewein T, Leibfried F, Grau-Moya J, Braun DA. 2015 Bounded rationality, abstraction, and hierarchical decision-making: an information-theoretic optimality principle. *Front. Robot. AI* **2**, 27. (doi:10.3389/frobt.2015.00027)

66. Wu S-W, Delgado MR, Maloney LT. 2015 Motor decision-making. In *Brain mapping: an encyclopedic reference*, vol. 3 (ed. AW Toga), pp. 417–427. New York, NY: Academic Press.

67. Murnighan JK, King TR, Schoumaker F. 1990 The dynamics of cooperation in asymmetric dilemmas. *Res. Pap. Econ.* **7**, 179–202.

68. Murnighan J. 1991 Cooperating when you know your outcomes will differ. *Simul. Gaming* **22**, 463–475. (doi:10.1177/1046878191224003)

69. Murnighan JK, King TR. 1992 The effects of leverage and payoffs on cooperative behavior in asymmetric dilemmas. In *Social dilemmas: theoretical issues and research findings* (eds WBG Liebrand, DM Messick, HAM Wilke), pp. 163–182. Oxford, UK: Pergamon Press.

70. Beckenkamp M, Hennig-Schmidt H, Maier-Rigaud FP. 2006 *Cooperation in symmetric and asymmetric Prisoner's Dilemma games*. Bonn: Preprints of the Max Planck Institute for Research on Collective Goods. (http://hdl.handle.net/10419/26909).

71. Lindig-León C, Schmid G, Braun DA. 2021 Bounded rational response equilibria in human sensorimotor interactions. Figshare.

9

royalsocietypublishing.org/journal/rspb *Proc. R. Soc. B* **288**: 20212094