Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

The three R's of trust

P Read Montague[1,2,3], Terry Lohrenz[1] and Peter Dayan[4]

Trust is fundamental to all human exchange. It is therefore necessary to understand the behavioural logic and neural underpinnings of how it is established, maintained, broken, and repaired. There are open evolutionary questions about how trust mechanisms should and do evolve; however, here we focus on more proximate psychological and neural mechanisms that mediate trust. Our perspective singles out three basic features: (1) reaping; (2) regarding; and (3) recursive modelling. Reaping involves mechanisms that respond to punishment or reward. Regarding comprises other-regarding or pro-social mechanisms. Recursive modelling encompasses hierarchical cognitive modelling of others through exchanges. Modern work has begun to reveal how these features are realized as modifications to conventional decision-making systems and how they malfunction in disease and injury.

Addresses
[1] Virginia Tech Carilion Research Institute, Virginia Tech, Roanoke, VA 24016, USA
[2] Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA
[3] Wellcome Centre for Imaging Neuroscience, 12 Queen Square, London WC1N 3BG, UK
[4] Gatsby Computational Neuroscience Unit, Alexandra House, 17 Queen Square, London WC1N 3AR, UK

Corresponding author: Read Montague, P (read@vt.edu)

Current Opinion in Behavioral Sciences 2015, **3**:102–106

This review comes from a themed issue on **Social behavior**

Edited by **Molly J Crockett** and **Amy Cuddy**

For a complete overview see the Issue and the Editorial

Available online 10th March 2015

http://dx.doi.org/10.1016/j.cobeha.2015.02.009

Introduction

In one basic sense, trust is all about statistical reliability. We can observe that many people have safely crossed a bridge over a river and, in confident belief in nature's blind indifference to our cares and concerns, trust the bridge and cross ourselves. However, trust in the bridge can also be based on interpersonal considerations — albeit realized via a kind of structural relationship involving governments, regulations, and the law. A huge benefit is often available when we can trust in this way. The interpersonal notion of trust changes character further when one interacts in real-time with other immediately connected people — particularly in circumstances in which they have the incentive to defect. This kind of trust is essential for all sorts of productive exchange, but the rapid-fire kinetics, and the variability across potential partners, make the subject slippery and inchoate. When interpersonal trust breaks down, for instance in psychiatric conditions such as borderline personality disorder, significant distress and damage can ensue.

Trust raises both distal and proximal problems. The distal one is why it might have evolved in the first place. Even though trust has substantial benefits if everyone is an angel, the devil is reliably known to get the best tunes. Thus, what is the Darwinian route by which untrustworthy defectors are prevented from successfully invading an halcyon trusting group? Work on efficient evolutionary stable strategies (ESSs) that we review very briefly below, has largely focused on the nature of the policies of agents and the properties of environments that lead behaviour associated with trustworthiness to be robust in populations. However, we primarily focus instead on the proximal psychological and neural mechanisms that presumably realize the critical features of these policies, and yet leave their traces even in short-term interactions involving many fewer players.

Evolutionarily stable trust

Cooperation among self-interested organisms remains a deep difficulty for evolutionary biologists — in terms of how and why classes of cooperation emerge in interacting populations. Over the last 30 years, some of the central problems of cooperation and cheating (defecting) have been modelled as games, most notably the Prisoner's Dilemma [1•]. In the 'iterated' version of this game, two players can choose one of two actions at each round — to defect or cooperate — and the payoffs for the joint choices across the two players structure the style of play. This game has been used as a kind of canonical model for understanding how trustworthy behaviour can arise when defection is so seductive that defectors might be expected to be able to invade a population of trustworthy individuals [2–5]. Many subtle issues arise that are still of active interest — for instance there has been recent work on the benefits of being an extortionist, at least in small populations [6•]. Efforts have also been made to connect the distal and proximal levels mentioned above [7,8•].

Trust as a rationally irrational decision

Evolution might indeed have inked trustworthiness into the very fabric of decision-making. However, this then poses what we referred to above as the proximal question which concerns the individual or cultural mechanisms that underlie its realization. That is, in the relevant cases, for me to trust you is to expect you to behave in a way that might seem to me to be irrational for you. On what bases can I hope that you will act in a way that appears to be against your short-term self-interest? Of course, what is sauce for the goose is sauce for the gander too — so I will have to worry about how you will reciprocally interpret my actions and gestures.

Reaping

One reason to trust you is the possibility of punishment (or, conversely, reward) — that is, reaping what you sow. Punishment could either be exacted by me directly, or by third parties or institutions [9,10], and could either be reflective or reflexive. In any case, since you know that defection comes at a potentially high price, particularly when that price might involve a ruined reputation, the apparent benefit of your doing so evaporates — a fact of which I am also aware.

Regarding

Alternatively, even without the possibility of punishment or reward, your utility function might directly include terms that depend on social factors — that is, you might feel guilty if you defect on me [11,12] be altruistically minded [13•], or indeed you might be pleased to see a defector laid low [14]. In this case, you have a taste for outcomes that would be seen as promoting trustworthiness, and so what seems irrational would in fact be perfectly rational. More complicated, though, is the question of how I could know that you have a guilty constitution — in general, this might come back to statistical reliability (e.g. most people I have encountered in my 'in-group' do), perhaps generalizing on the basis of what I see or read about you [15], or require me to have observed or interacted with you in the past [16,17], something that can even experimentally lead to the establishment of lucrative mutual understanding.

Recursive modelling

Most other-regarding utility functions focus only on the degree of unfairness of the net outcome. However, the source of the unfairness — in particular, the intention of the defector — is also important [18•]. I am more likely to be generous towards you if you have been generous to me by choice or by rational calculation, rather than by fiat. This implies that I must model aspects of the determiners of your actions. This reaches its apogee in repeated interactions between players, especially when more benefit can be extracted from the environment over multiple rounds if players can coordinate their behaviour, but with payoffs encouraging short-term defection. The danger in such circumstances is the unravelling of coordination, leaving all players worse off. In this case, if I think (or can safely or inexpensively probe to find out) that you're smart enough to realize this fact, and, crucially, smart enough to realize that I'm smart enough to realize this too, then we can form what amounts to a temporary alliance — even if our individual utility functions lack appropriate social features. For this to work, the mechanisms to support such an alliance must at least potentially pre-exist in both my and your minds before we interact, including the image of me in your mind, and the image of this image in mine, and so on. Some aspects of the punishments and rewards associated with reaping can arise through recursive modelling — but typically not those associated with larger-scale, impersonal, societal mechanisms.

Trust is built and sustained by generous gestures, broken by defections, and repaired by coaxing [19•,20]. To understand the semantics of any of these gestures requires something akin to these images, that is, sufficiently complex mutual modelling. Unfortunately, normative versions of such calculations over multiple steps, involving reasoning about the initially incompletely known capacities, goals and intentions of ones partners, are radically computationally intractable. There is quite some evidence for our ability to model other people, and some of the neural substrates concerned have been identified [21–26]. However, people are, as expected, reported to be severely limited in their capacity for recursive modelling [27,28]. Methods from machine learning [8•,26,29•,30,31] are able to make some further headway to examine behaviour that is appropriate for a given level, and to explore the effects of partner modelling. However, potentially deleterious short-cuts and heuristics are rapidly essential for either man or machine, perhaps in the form of factors such as social contagion [32] — which we might think of as Pavlovian behaviour or misbehaviour [33] in a social context.

It is only really in the case of other-regarding utility functions that trustworthy behaviour is a mark of true reputational trustworthiness that will generalize between tasks and times. Partners who can only be trusted because of the prospect of punishment or from careful recursive calculations could readily defect when the conditions are different. Unfortunately, despite some formal attempts [34] it may not always be obvious which of the three R's is operative in any given interaction, so uncertainty about the underlying reputations inevitably abounds.

Exploring trust

These drivers of, and behaviours associated with, trust have been most completely worked out in game theoretic aspects of theoretical, experimental and behavioural economics. In this, formal theories include Nash equilibria; or, in cases when players can be uncertain about each other, and treat this uncertainty in a statistically appropriate manner, Bayes-Nash equilibria [35], which admit learning. In particular the work of Camerer *et al.* [36•] on experiential weighted attraction (EWA) bridges these concepts. Most games exploring trust make defection sufficiently attractive that one-shot equilibrium solutions are trivial, and long-run coordination unravels. That this is not what is found empirically, has led to such suggestions as social-regarding utility functions [11], and both more explicit and more limited forms of reasoning about other players [18•], in settings of both complete [27] and incomplete [27,29•,30,36•] information.

Empirically, many several games involve and explore aspects of trust. Although some of the elements that we identified above have been studied in situations in which trust does not play a key role — for instance cognitive hierarchy [28], which is a form of recursive modelling, is revealed in behaviour in p-beauty games [37,38]; and social-regarding aspects of utility [11] in dictator games [39–41], which allow one subject complete control over how a stake is divided — trust has itself often been a central concern. Cases include the (iterated) prisoner's dilemma [3], public goods games [42,43]; which have examined many aspects of the effect of explicit punishment, ultimatum games [9], which show how partner modelling and the prospect of effective punishment lead to more cooperative behaviour, and trust games themselves [44,45].

The trust game [44] is perhaps the canonical testbed for exploring the various issues above. The game involves two players in well-defined (but asymmetric) roles and is played for one or more rounds. On each round, one player (the investor) is endowed with money and can choose to send any fraction (including 0) of this endowment to the partner (the trustee) in which case the amount sent earns a return (typically 3-fold). Control passes to the trustee who (like the dictator in a dictator game) can send back any fraction of the tripled amount to the investor (including 0). When money is sent, the trust displayed by the investor can be repaid by the trustee who, by virtue of the profit earned (here a tripling), can return the gesture *and* make some money in the process. This is how the potential economic efficiency of coordination is operationalized. In practice, investors send 40–60% of their endowment even when virtually nothing specific is known about the trustee [11,27,46], and when the structure the game allows the trustee to defect completely. However, the substantial trust shown by the investors is indeed reciprocated by the trustee [27], showing the power of mechanisms such as the ones discussed above.

Neural substrates

Recently, there has been an explosion of work investigating the neural underpinnings of the trust game and related social exchange games, illustrating all aspects of the three R's [47•,48•]. Here we highlight a few key points.

Speaking to the first 'R', Li *et al.* [49] investigated the neural underpinnings of the threat of sanctions in a modified multi-round trust game. Investors were given the opportunity to impose a monetary sanction on the trustee if the trustee did not meet the investor's repayment request. Activity in the ventral medial prefrontal cortex (vmPFC) at the time the sanction decision was revealed tracked the future repayment in both the sanction and non-sanction cases. However, the intercept was larger in the sanctions condition, mirroring the fact that repayments were *lower* in the sanctions condition.

Tricomi and colleagues [12] examined the neural underpinnings of the second 'R', regarding. Pairs of subjects were endowed with unequal amounts of money, and then asked in the fMRI scanner to rate further experimenter transfers of money to themselves or the partner. They found that brain regions known to respond to reward, ventral medial prefontal cortex (vmPFC) and ventral striatum (VS), differentially responded to the transfers depending on the induced state of inequality in a way consistent with other regarding utility.

In King-Casas [19•] the repeated trust game was used to probe the gestures that underpin the communication inherent to trust and that are implicit in the third 'R', recursive modelling. They concentrated on rather direct forms of reciprocity — how either player responds to generous or miserly offers. Behaviourally, investor reciprocity predicted changes in trust in the trustee. Neurally, benevolent reciprocity elicited signals in the caudate, reminiscent of those associated with reward, Further, in this region there was an 'intention to trust' signal, that moved back in time as reputation developed, like the TD error in machine learning. These results provide evidence that one of the proximate mechanisms of trust operates in the human brain.

More recently, work has appeared examining the neural mechanisms underlying the detailed recursive models of social interaction described above. Several recent papers have addressed the neural substrates of these proposed mechanisms [34,50–52]. For example, Xiang *et al.* [34] examined neural signatures of prediction errors calculated from a model of behaviour incorporating the modelling ideas described above and using the other-regarding [11] Fehr-Schmidt utility function. They reported two orders of prediction errors that would arise in me if I am playing you: the first measures the difference between your actual and expected moves; while the second measures the difference between my move and the move that your model of me would predict.

Trust in computational psychiatry

One growing application area for the mechanisms probed by trust games and their congeners is psychiatry and neurology [53]. The idea is that many forms of perturbed mental function may reflect changes in the capacity to model others and respond to signals of reciprocation. Certainly this point of view stands at the heart of modern ideas about developmental disorders like Autism Spectrum Disorder. However, the capacity to model other humans is sophisticated and may consequently be commonly subject to damage from frank injury or even changes with age.

King-Casas and colleagues [20] used a multi-round trust game to expose an apparent and selective social modelling deficit in subjects diagnosed with Borderline Personality Disorder (BPD). The data suggest that subjects in this group over-react to incoming gestures from their partner, which leads them to break trust, and are then more reluctant than healthy controls to repair cooperation. Koshelev *et al.* [54] used model-based clustering of healthy investor behaviour in the trust game in to identify clusters that corresponded well to the groups of *trustee* subjects with mental disorders that included BPD, major depressive disorder (MDD), and autism spectrum disorder (ASD). Similarly, Yoshida *et al.* [52] used a related game (the stag hunt) to probe heterogeneity of the abilities of subjects to understand strategic sophistication in others in ASD. Finally, Dayan *et al.* [33], found that when healthy investors played the trust game against BPD trustees, their effective depth of recursive thought was lower than (other) healthy investors played against healthy trustees. These are early findings, but they suggest a potential use of computational models of social exchange problems to understand and characterize mental illnesses.

Summary

Given all its potential fragilities and seeming irrationality, trust between humans appears remarkably sturdy. A variety of microeconomic games — notably the so-called trust game — has been used to explore how it gets that way, leading to substantial psychological and neural insights. Aside from possibilities for punishment, trust can be underpinned by social-regarding utility functions, particularly in conjunction with the capacity for partners to engage in recursive modelling of each other's goals and intentions. The latter poses a huge computational challenge in the face of inevitable uncertainty. Thus, it is a pressing next step to understand the heuristics and shortcuts that underpin the robustness of trust, and to understand how it breaks down in mental illness.

Conflict of interest statement

Nothing declared.

Acknowledgments

PRM thank Wellcome Trust Principal Research Fellowship. TL would like to thank Virginia Tech. PD would like to thank the Gatsby Charitable Foundation.

References

1. • Rapoport A, Chammah AM: *Prisoner's Dilemma*. The University of Michigan Press; 1965.

The prisoner's dilemma has become one of the most important theoretical and experimental tools for thinking about cooperation and competition; and, in its repeated or iterated form, the establishment and breakdown of trust. Computer-based tournaments between strategies have been particularly revealing.

2. Trivers RL: **The evolution of reciprocal altruism**. *Q Rev Biol* 1971, **46**:35-57.
3. Axelrod R, Hamilton WD: **The evolution of cooperation**. *Science* 1981, **211**:1390-1396.
4. Smith JM: *Evolution and the Theory of Games*. Cambridge University Press; 1982.
5. Nowak MA: *Evolutionary Dynamics*. Harvard University Press; 2006.
6. • Press WH, Dyson FJ: **Iterated Prisoners' Dilemma contains strategies that dominate any evolutionary opponent**. *Proc Natl Acad Sci U S A* 2012, **109**:10409-10413.

Uses elementary matrix calculations to show that strategies exist, so-called zero determinant strategies, that allow a player to unilaterally set the long-term score of the other player, or extract an 'extortionate' share of the payoffs.

7. Keltner D, Kogan A, Piff PK, Saturn SR: **The sociocultural appraisals, values, and emotions (SAVE) framework of prosociality: core processes from gene to meme**. *Annu Rev Psychol* 2014, **65**:425-460.
8. • Devaine M, Hollard G, Daunizeau J: **Theory of mind: did evolution fool us?** *PLOS ONE* 2014, **9**:e87619.

Uses a recursive cognitive hierarchy model (that estimates lower level type) to simulate play in a competitive game (hide and seek) and a cooperative game (battle of the sexes). They show that having more sophisticated theory of mind (ToM) pays off in competitive games but not quite so much in cooperative games. Similarly, in an evolutionary game theory setting, competition favors high theory of mind (ToM) agents, but cooperation favors a mixture of lower ToM types.

9. Güth W, Schmittberger R, Schwarze B: **An experimental analysis of ultimatum bargaining**. *J Econ Behav Organ* 1982, **3**:367-388.
10. Fehr E, Gachter S: **Altruistic punishment in humans**. *Nature* 2002, **415**:137-140.
11. Fehr E, Schmidt KM: **A theory of fairness, competition, and cooperation**. *Q J Econ* 1999, **114**:817-868.
12. Tricomi E, Rangel A, Camerer CF, O'Doherty JP: **Neural evidence for inequality-averse social preferences**. *Nature* 2010, **463**:1089-1091.
13. • Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, Dolan RJ: **Harm to others outweighs harm to self in moral decision making**. *Proc Natl Acad Sci U S A* 2014, **111**:17320-17325.

Considers why some circumstances foster narrower self-interest, whereas other foster more regard for fairness. Introduces self-centred inequity aversion in the form of an other-regarding utility function.

14. Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD: **Empathic neural responses are modulated by the perceived fairness of others**. *Nature* 2006, **439**:466-469.
15. Delgado MR, Frank RH, Phelps EA: **Perceptions of moral character modulate the neural systems of reward during the trust game**. *Nat Neurosci* 2005, **8**:1611-1618.
16. Phan KL, Sripada CS, Angstadt M, McCabe K: **Reputation for reciprocity engages the brain reward center**. *Proc Natl Acad Sci U S A* 2010, **107**:13099-13104.
17. Fareri DS, Chang LJ, Delgado MR: **Effects of direct social experience on trust decisions and neural reward circuitry**. *Front Neurosci* 2012, **6**:148.
18. • McCabe KA, Rigdon ML, Smith VL: **Positive reciprocity and intentions in trust games**. *J Econ Behav Organ* 2003, **52**:267-275.

Includes some clever sequential interaction games that favour the 'trust and reciprocity' view that the intention of players matters for fairness, along with the equity of the outcomes.

19. • King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR: **Getting to know you: reputation and trust in a two-person economic exchange**. *Science* 2005, **308**:78-83.

Reveals in the caudate an 'intention to trust' signal that transfers in time in a way reminiscent of temporal prediction errors as reputation is established in the game.

20. King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR: **The rupture and repair of cooperation in borderline personality disorder**. *Science* 2008, **321**:806-810.

21. Amodio DM, Frith CD: **Meeting of minds: the medial frontal cortex and social cognition**. *Nat Rev Neurosci* 2006, **7**:268.

22. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS: **Associative learning of social value**. *Nature* 2008, **456**:245-249.

23. Frith CD, Frith U: **The neural basis of mentalizing**. *Neuron* 2006, **50**:531-534.

24. Hampton A, Bossaerts P, O'Doherty J: **Neural correlates of mentalizing-related computations during strategic interactions in humans**. *Proc Nat Acad USA* 2008, **105**:6741-6746.

25. Saxe R, Moran JM, Scholz J, Gabrieli J: **Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects**. *Soc Cogn Affect Neurosci* 2006, **1**:229-234.

26. Diaconescu AO, Mathys C, Weber LA, Daunizeau J, Kasper L, Lomakina EI, Fehr E, Stephan KE: **Inferring on the intentions of others by hierarchical Bayesian learning**. *PLoS Comput Biol* 2014, **10**:e1003810.

27. Camerer CF: *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton University Press; 2003.

28. Camerer CF, Ho TH, Chong JK: **A cognitive hierarchy model of games**. *Q J Econ* 2004, **119**:861-898.

29. • Gmytrasiewicz PJ, Doshi P: **A framework for sequential planning in multi-agent settings**. *J Artif Intell Res* 2005, **24**:49-79.
Models repeated interactions between players with imperfect information about each other's characteristics or types, and a formally limited capacity for recursive modelling, as what they call an interactive partially observable Markov decision process. This provided the formal framework for much subsequent modelling of the trust game.

30. Wunder M, Kaisers M, Yaros JR, Littman M: **Using iterated reasoning to predict opponent strategies**. In *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011).* Edited by Tumer, Yolum, Sonenberg, Stone. International Foundation for Autonomous Agents and Multiagent Systems; 2011:593-600.

31. Ray D, King-Casas B, Montague PR, Dayan P: **Bayesian model of behavior in economic games**. *NIPS* 2008.

32. Goetz JL, Keltner D, Simon-Thomas E: **Compassion: an evolutionary analysis and empirical review**. *Psychol Bull* 2010, **136**:351-374.

33. Dayan P, Niv Y, Seymour B, Daw ND: **The misbehavior of value and the discipline of the will**. *Neural Netw* 2006, **19**:1153-1160.

34. Xiang T, Ray D, Lohrenz T, Dayan P, Montague PR: **Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought**. *PLoS Comput Biol* 2012, **8**:e1002841.

35. Harsanyi JC: **Games with incomplete information played by Bayesian players. 1**. *Manage Sci* 1967, **14**:159-182.

36. • Camerer CF, Ho TH, Chong JK: **Sophisticated experience-weighted attraction learning and strategic teaching in repeated games**. *J Econ Theory* 2002, **104**:137-188.
Offers and tests a way of treating a formally limited capacity for other-modeling in terms of what it calls a cognitive hierarchy (I model your model of my model of you, etc.). Of particular note is the severe constraint in normal populations — the mean number of levels in the hierarchy appears to be less than two.

37. Keynes JM: *The General Theory of Employment, Interest, and Money*. Macmillan; 1936.

38. Nagel R: **Unraveling in guessing games: an experimental study**. *Am Econ Rev* 1995, **85**:1313-1326 1313.

39. Camerer C, Thaler RH: **Anomalies: ultimatums, dictators and manners**. *J Econ Perspect* 1995, **9**:209-219.

40. Kahneman D, Knetsch JL, Thaler RH: **Fairness and the assumptions of economics**. *J Bus* 1986, **59**:S285-S300.

41. Hoffman E, McCabe KA, Smith VL: **On expectations and the monetary stakes in ultimatum games**. *Int J Game Theory* 1996, **25**:289-301.

42. Ostrom E, Walker J, Gardner R: **Covenants with and without a sword: self-governance is possible**. *Am Polit Sci Rev* 1992, **86**:404-417.

43. Fehr E, Gachter S: **Fairness and retaliation: the economics of reciprocity**. *J Econ Perspect* 2000, **14**:159-181.

44. Berg J, Dickhaut J, McCabe K: **Trust, reciprocity, and social history**. *Games Econ Behav* 1995, **10**:122-142.

45. Camerer C, Weigelt K: **Experimental tests of a sequential equilibrium reputation model**. *Econometrica* 1988, **56**:1-36.

46. Camerer CF, Fehr E: **When does "economic man" dominate social behavior?** *Science* 2006, **311**:47-52.

47. • Tzieropoulos H: **The trust game in neuroscience: a short review**. *Soc Neurosci* 2013, **8**:407-416.
Reviews a wealth of work on the cognitive neuroscience of the trust game, including many factors, such as the role of the hormone oxytocin that we were not able to cover.

48. • Ruff CC, Fehr E: **The neurobiology of rewards and values in social decision making**. *Nat Rev Neurosci* 2014, **15**:549-562.
Pays particular attention to the integration of social-influences and non-social influences on value structures associated with choice.

49. Li J, Xiao E, Houser D, Montague PR: **Neural responses to sanction threats in two-party economic exchange**. *Proc Natl Acad Sci U S A* 2009, **106**:16835-16840.

50. Yoshida W, Dolan RJ, Friston KJ: **Game theory of mind**. *PLoS Comput Biol* 2008, **4**:e1000254.

51. Yoshida W, Seymour B, Friston KJ, Dolan RJ: **Neural mechanisms of belief inference during cooperative games**. *J Neurosci* 2010, **30**:10744-10751.

52. Yoshida W, Dziobek I, Kliemann D, Heekeren HR, Friston KJ, Dolan RJ: **Cooperation and heterogeneity of the autistic mind**. *J Neurosci* 2010, **30**:8815-8818.

53. Kishida KT, King-Casas B, Montague PR: **Neuroeconomic approaches to mental disorders**. *Neuron* 2010, **67**:543-554.

54. Koshelev M, Lohrenz T, Vannucci M, Montague PR: **Biosensor approach to psychopathology classification**. *PLoS Comput Biol* 2010, **6**:e1000966.