Title

Cognitive neuroscience of honesty and deception: a signaling framework

Permalink

https://escholarship.org/uc/item/2j06p6qx

Authors

Jenkins, Adrianna C
Zhu, Lusha
Hsu, Ming

Publication Date

2016-10-01

DOI

10.1016/j.cobeha.2016.09.005

Published in final edited form as:
Curr Opin Behav Sci. 2016 October ; 11: 130–137. doi:10.1016/j.cobeha.2016.09.005.

Cognitive neuroscience of honesty and deception: A signaling framework

Adrianna Jenkins[1], **Lusha Zhu**[2,*], and **Ming Hsu**[1,*]

[1]Haas School of Business and Helen Wills Neuroscience Institute, University of California, Berkeley

[2]PKU-IDG/McGovern Institute For Brain Research, School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Peking-Tsinghua Center for Life Sciences, Peking University, China

Abstract

Understanding the neural basis of human honesty and deception has enormous potential scientific and practical value. However, past approaches, largely developed out of studies with forensic applications in mind, are increasingly recognized as having serious methodological and conceptual shortcomings. Here we propose to address these challenges by drawing on so-called signaling games widely used in game theory and ethology to study behavioral and evolutionary consequences of information transmission and distortion. In particular, by separating and capturing distinct adaptive problems facing signal senders and receivers, signaling games provide a framework to organize the complex set of cognitive processes associated with honest and deceptive behavior. Furthermore, this framework provides novel insights into feasibility and practical challenges of neuroimaging-based lie detection.

Introduction

Questions regarding honesty and deception have been at the heart of many iconic episodes of human history. The Watergate hearings, for example, revolved around the now-famous question of willful deception on part of the Nixon White House: “What did the President know and when did he know it?” At the neural level, early interest in (dis)honesty stemmed from attempts to develop methods of distinguishing the truth from lies in forensic and legal settings [1,2]. These include, among others, efforts to improve criminal justice (e.g., by identifying perpetrators) and intelligence analysis (e.g., by predicting terrorist activity).

Guided by these goals, early studies, typically using the Comparison Question Test (CQT) or Guilty Knowledge Test (GKT), sought to identify a set of physiological (e.g., arousal) or neurocognitive (e.g., anxiety, guilt) processes that could serve as objective markers of

[*]Correspondence to lushazhu@pku.edu.cn or mhsu@haas.berkeley.edu.

deception [3–6]. In the GKT, participants view pieces of information that are either relevant or irrelevant to the target incident, and physiological responses are compared during exposure to items that only someone with knowledge of the target incident should recognize as relevant versus to other items [7].

Comparing lie versus truth-telling conditions, these studies repeatedly found differential responses in regions of the prefrontal cortex previously associated with cognitive control and higher-order cognition [8–14]. Recent meta-analyses have further shown that lying was associated with greater activation in regions including dorsolateral and ventrolateral prefrontal cortex, anterior insula and superior frontal lobule (Figure 1) [1,15,16]. Strikingly, no brain region has been consistently found to respond more during truth-telling than to lying, a finding often interpreted as supporting the notion that deception is a more cognitively demanding than truth-telling.

Despite their popularity, however, there is growing dissatisfaction with past studies inspired by the CQT and GKT approaches [1,17]. The most commonly cited criticism centers on the fact that participants were often instructed by the experimenters to deceive or withhold information [1,17]. This, however, raises important questions regarding the external validity of studies using “instructed deception”. In particular, by removing from participants their ability to choose to deceive, instructed deception paradigms make it challenging if not impossible for experimenters to study the involvement of motivational and decision-making processes in deception.

Honesty and Deception in Ethology and Economics

Behaviorally, studies of honesty and deception have their roots in ethology and economics, owing to the importance of honesty and deception to problems of mate selection, predator avoidance, and economic exchange, among others. At the heart of both literatures is the idea that honesty and deception are properties of the communicative signals that organisms send to one another in the service of some economic or evolutionary function, which can be captured via so-called signaling games [18–20].

In particular, signaling games capture the antecedents and consequences of both honesty and deception across a diverse set of decision scenarios, thereby providing a stylized environment to take into account the adaptive problems faced by the signal senders and receivers. The signals may involve the use of language, as in the case of humans, or physiological and morphological characteristics more commonly associated with animal signaling [18,21]. In either case, the signals themselves have no immediate causal effect or direct payoff consequences for the sender and receiver—an observation echoed in the common refrain that “talk is cheap” (Figure 1). Instead, it is how the signal is interpreted by the signal recipient that ultimately carries consequences. For example, it is possible that the vocal calls of two species may have identical auditory characteristics but have completely different meaning, where one is used for predator alerts and the other for mate attraction. Similarly, a bargaining offer from one party to another can have very different interpretations depending on whether the two parties are historical rivals or partners [22,23].

A signaling framework of honesty and deception

Here we offer a cognitive framework of honesty and deception by describing and organizing processes underlying such behavior in the language of economic games of signaling [18,24,25]. As with the broader signaling literature in evolutionary biology and economics, the value of such an approach lie not so much in explaining new facts about processes underlying honesty and deception, but rather in thinking about them in the context of instrumental interaction between goal-directed agents.

Our framework first distinguishes between cognitive processes belonging to senders and receivers, respectively. On the part of the sender, this entails processes associated with (i) constructing a representation of the signaling problem, including identifying characteristics of the players, actions, and outcomes under consideration and (ii) forming appropriate beliefs about the relationships among possible actions and possible outcomes. This second set may involve the engagement of theory of mind or mentalizing processes to construct beliefs about the receiver's mental states (e.g., beliefs, desires, intentions) and future actions [26,27], as in the case of "intentional" deception. Alternatively, explicit belief construction may not be involved, with associations being generated through more basic processes such as associative learning or even genetically specified reflexes, which typically fall under the heading of "functional" deception in the nonhuman animal literature.

Based on these expectations, the sender also engages processes associated with (iii) assigning value representations to different actions under the sender's consideration. These values can contain not only the direct fitness or payoff implications, but also psychological values, such as internal costs associated with an aversion to deception or inequity. (iv) Additionally, the sender must select among these actions in a way that balances competing motives, for example the motive to pursue one's own self-interest versus the motive to be honest, typically involving processes such as cognitive control [8,28]. With some notable exceptions, studies to date have focused almost exclusively at the level of action selection [1,8,17,29,30], with only a few studies systematically manipulating experimental variables relating to valuation, outcome, or contextual characteristics [31–34].

For the receiver, a similar set of processes unfolds following receipt of a signal. These include: (i) identifying and evaluating potential (honest and deceptive) signals, (ii) anticipating outcomes resulting from said signals, (iv) constructing the values of different actions, and (iii) selecting appropriate action(s) based on the anticipated outcomes (Figure 2B).

A signaling framework therefore clarifies two of the major criticisms invoked against GKT and CQT. First, by mandating deception, these paradigms remove motivational components that are critical in honest or deceptive signaling. Second, and less appreciated, is the fact that participants in the GKT and CQT are asked to act as both signal receiver as well as sender. In particular, because both tasks involve decoding a signal from the interrogator and then sending a signal back to that interrogator, it can be challenging to isolate processes associated with any particular stage in the generation of honest or deceptive behavior.

Neural and computational mechanisms

Beyond clarifying existing criticisms, a signaling framework has the potential to organize existing data and suggest novel ways to address existing debates regarding the neurobiological substrates underlying honesty and deception. This is particularly important given the growing number of studies that take a signaling approach methodologically, either implicitly or explicitly, such that participants are allowed to freely decide between honest and deceptive action [34–37]. For example, at the signal selection stage, neuroimaging studies have frequently found evidence that brain regions associated with cognitive control are involved in decisions regarding honesty and deception [15,16], but their inherently correlational nature leaves open questions about the extent to which cognitive control is needed in order to be honest or in order to be deceptive. By distinguishing between different levels of processing, our signaling framework highlights the presence of a third possibility, namely that cognitive control may be needed or for both honesty and deception, but at different stages.

Recently, by pairing signaling games with a lesion approach, it was shown that damage to regions of the human lateral prefrontal cortex, previously implicated in neuroimaging studies of honesty and deception, was associated with lower levels of honesty (Figure 3). That is, senders with damage to the lateral prefrontal cortex, an area associated with cognitive control, were more willing to send deceptive signals in order to earn more money, suggesting that cognitive control is necessary for producing honest signals when it is in one's own interest to lie. This case illustrates how the use of behavioral measures derived from signaling games can be decisive in allowing researchers to directly test mechanistic hypotheses regarding the relationship between brain and behavior.

Similarly, at the outcome anticipation and valuation stages, existing studies have suggested that these processes overlap in important ways to those implicated in previous studies of goal-directed behavior [39,40]. For example, Bhatt et al. [32] studied theory of mind processes using a bargaining task involving a series of one-shot bargains over a single unit of some good. The buyer, who values the good at v, suggests a selling price s to a seller. Upon receiving s, the seller submits a selling price p. If $p > v$, no deal occurs, otherwise the transaction is executed with seller receiving p and buyer $v - p$. In this case, therefore, the buyer has an incentive to underreport the true v, a fact that the seller should take into consideration when determining the selling price p.

Using this task, it was found that the buyers' stated value of the underlying good was associated with activity in the buyers' right temporoparietal junction (rTPJ), a region previously implicated in theory of mind processes, especially representing others' beliefs [41,42]. Interestingly, this was only the case for individuals who were strategic in their reports of private value v, suggesting a role for processes supported by the rTPJ in "intentional" deception (Figure 4). In another study, Baumgartner et al. [33] used a variant of the Trust game [24] with an antecedent "Promise" stage in which participants indicated how much they would share of the total amount of money they received. The researchers found that, during the initial promise and anticipation stages of the task, activity in several regions, including the insula, predicted whether or not participants would ultimately break those promises, supporting a possible role for the insula in the anticipation of norm violation [43,44].

Implications for Neuroimaging-Based Lie Detection

This framework also sheds light on some important issues in current debates on neuroimaging-based lie detection. For example, an influential review by Sip et al. [17] raised two broad challenges regarding the use of neuroimaging for lie detection: (1) the difficulty of inferring deception based on activity in brain regions associated with emotion, mentalizing, and risk taking, as they are involved in many other cognitive and behavioral processes, and (2) the lack of experimental paradigms that capture real-world deception.

As Haynes [45] points out, however, one does not need a complete characterization of the underlying neurocognitive mechanisms to develop diagnostics of deception. Any cognitive process that is involved in deception can, in principle, be used for lie detection purposes if it is sufficiently selective and specific [45]. A signaling framework could in principle allow researchers to decompose lie detection into a set of subcomponents, each of which could be investigated independently.

In the example given in Figure 2, this corresponds to asking whether the stock broker lied by asking whether the patterns of neural responses can be found that predicts whether broker (i) had knowledge of the stock being bad, (ii) anticipating consequences of having the lie exposed, (iii) encoding disutility of lying or cost of being caught, or (iv) response conflict associated with lying. Whereas past neuroimaging studies have largely been concerned with lie detection at the action selection level [1,7,47], future studies can begin to compare and contrast the relative strengths and weakness of lie detection at each level.

At the same time, because lie detection itself constitutes a signaling game between the interrogators and the interrogated, practical applications must be robust to attempts to game the system—for example by regulating one’s cognitive or affective response to a stimulus. Here, the fact that signaling games enable studies of realistic deception within a laboratory setting represents a particular advantage for testing the robustness of potential lie detection measures across contexts, in addition to their selectivity and specificity within a specific context. Although no less daunting, addressing these questions will begin to inform and improve current lie detection efforts by providing more rigorous and cumulative scientific approach [1,2].

Future Directions

In one sense, signaling games are just economic games of incomplete information, where players can use the actions of their counterparts to make inferences about hidden information. Yet, despite its apparent simplicity and success in explaining a number of key empirical regularities regarding honest and deceptive behavior, ranging from mating selection to negotiation, genuine puzzles remain at nearly all levels of analysis. For example, we still lack a satisfactory theory explaining how specific signals have evolved to acquire meaning. Most existing empirical studies have taken the set of signals available to the sender to be given, and leave open the question of how senders and receivers came to agree on a set of signals, and the mapping of signal to the "true" state, or for that matter whether such an agreement is necessary at all. A better understanding of these processes would require modeling the complex dynamics involved in sender receiver interactions. As noted in earlier, however, there has been little attention on how receivers process and respond to signals at the neural level beyond the basic sensory and perceptual properties.

Moreover, any realistic account of honesty and deception in humans must be able to account for the fact that information sharing is dominated by unstructured communication involving natural language and a diverse collection of nonverbal cues. However, no study to our knowledge has studied the neural mechanisms of honesty and deception in the context of unstructured communication. In behavioral studies, unstructured communication was found to substantially increased truth-telling and cooperation compared to structured communication where messages were preselected by the experimenters [19]. Although intuitive, none of our existing theories are able to explain why this is so, and what specific features of unstructured communication are responsible for the observed differences. Advances in this area will likely require additional consideration of the contribution of language processes, for example by incorporating game theoretic models of pragmatics recently developed in linguistics [48].

Acknowledgments

The authors would like to thank support from National Institutes of Health grant MH098023 to MH.

References

1**. Farah MJ, Hutchinson JB, Phelps EA, Wagner AD. Functional MRI-based lie detection: scientific and societal challenges. Nat Rev Neurosci. 2014; 15:123–131. Discussion of the scientific state of fMRI-based lie detection, the challenges associated with its application in forensic settings, as well as the broader ethical and societal implications. [PubMed: 24588019]

2. Langleben DD, Moriarty JC. Using Brain Imaging for Lie Detection: Where Science, Law and Research Policy Collide. Psychol Public Policy Law. 2013; 19:222–234. [PubMed: 23772173]

3. Seymour TL, Seifert CM, Shafto MG, Mosmann AL. Using response time measures to assess "guilty knowledge". J Appl Psychol. 2000; 85:30–37. [PubMed: 10740954]

4. Farwell LA, Donchin E. The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. Psychophysiology. 1991; 28:531–547. [PubMed: 1758929]

5. Spence SA, Hunter MD, Farrow TFD, Green RD, Leung DH, Hughes CJ, Ganesan V. A cognitive neurobiological account of deception: evidence from functional neuroimaging. Philos Trans R Soc London Ser B, Biol Sci. 2004; 359:1755–1762. [PubMed: 15590616]

6. Abe N, Suzuki M, Tsukiura T, Mori E, Yamaguchi K, Itoh M, Fujii T. Dissociable roles of prefrontal and anterior cingulate cortices in deception. Cereb Cortex. 2006; 16:192–199. [PubMed: 15858160]

7. Stern, PC. The Polygraph and Lie Detection. The National Academies Press; 2003.

8*. Spence SA, Farrow TFD, Herford AE, Wilkinson ID, Zheng Y, Woodruff PWR. Behavioural and functional anatomical correlates of deception in humans. Neuroreport. 2001; 12:2849–2853. The first functional neuroimaging study of human deception using an instructed lie paradigm involving whether presented everyday acts have been performed or not correctly and incorrectly. [PubMed: 11588589]

9. Abe N, Okuda J, Suzuki M, Sasaki H, Matsuda T, Mori E, Tsukada M, Fujii T. Neural correlates of true memory, false memory, and deception. Cereb Cortex. 2008; 18:2811–2819. [PubMed: 18372290]

10. Nuñez JM, Casey BJ, Egner T, Hare T, Hirsch J. Intentional false responding shares neural substrates with response conflict and cognitive control. Neuroimage. 2005; 25:267–277. [PubMed: 15734361]

11. Phan KL, Magalhaes A, Ziemlewicz TJ, Fitzgerald DA, Green C, Smith W. Neural correlates of telling lies: A functional magnetic resonance imaging study at 4 Tesla. Acad Radiol. 2005; 12:164–172. [PubMed: 15721593]

12. Kozel FA, Johnson KA, Mu Q, Grenesko EL, Laken SJ, George MS. Detecting deception using functional magnetic resonance imaging. Biol Psychiatry. 2005; 58:605–613. [PubMed: 16185668]

13. Ganis G, Kosslyn SM, Stose S, Thompson WL, Yurgelun-Todd DA. Neural correlates of different types of deception: An fMRI investigation. Cereb Cortex. 2003; 13:830–836. [PubMed: 12853369]

14. Langleben DD, Schroeder L, Maldjian JA, Gur RC, McDonald S, Ragland JD, O'Brien CP, Childress AR. Brain activity during simulated deception: an event-related functional magnetic resonance study. Neuroimage. 2002; 15:727–732. [PubMed: 11848716]

15*. Lisofsky N, Kazzer P, Heekeren HR, Prehn K. Investigating socio-cognitive processes in deception: a quantitative meta-analysis of neuroimaging studies. Neuropsychologia. 2014; 61:113–122. Meta-analysis of functional neuroimaging studies of honesty and deception. In addition to finding that prefrontal regions were more actived in production of deceptive responses compared to truthful responses, it was found that there was increased activation in the dorsal ACC, the right temporo-parietal junction (TPJ)/angular gyrus, and the bilateral temporal pole (TP) during social interactive. Most tasks in the meta-analysis involved instructed lie paradigms. [PubMed: 24929201]

16. Christ SE, Van Essen DC, Watson JM, Brubaker LE, McDermott KB. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. Cereb Cortex. 2009; 19:1557–1566. [PubMed: 18980948]

17*. Sip KE, Roepstorff A, McGregor W, Frith CD. Detecting deception: the scope and limits. Trends Cogn Sci. 2008; 12:48–53. Influential perspective paper on challenges facing neuroscientific study of deception. Highlights the methodological shortcomings of studies at the time which overwhelmingly rely on instructed deception, as well as need to clarify the complex set of cognitive processes underlying honesty and deceptive behavior. [PubMed: 18178516]

18**. Searcy WA, Nowicki S. The Evolution of Animal Communication: Reliability and Deception in Signaling Systems. 2010 An excellent overview of theoretical and empirical literature on honesty and deception in animal communication. Many of the examples have clear parallel in human deception, including those involving mating, resource contests, and even parental care.

19. Crawford V. A survey of experiments on communication via cheap talk. J Econ Theory. 1998; 78:286–298.

20*. Gneezy U. Deception: The role of consequences. Am Econ Rev. 2005; 95:384–394. One of the first papers to show using behavioral economic games that human participants display an aversion to dishonesty.

21. Hauser, MD. The evolution of communication. MIT press; 1997.

22. Carazo P, Font E. "Communication breakdown": the evolution of signal unreliability and deception. Anim Behav. 2014; 87:17–22.

23. Dakin R, Montgomerie R. Deceptive Copulation Calls Attract Female Visitors to Peacock Leks. Am Nat. 2014; 183:558–564. [PubMed: 24642499]

24. Camerer, CF. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press; 2003.

25. Lee D. Game theory and neural basis of social decision making. Nat Neurosci. 2008; 11:404–409. [PubMed: 18368047]

26. Sodian B. The development of deception in young children. Br J Dev Psychol. 1991; 9:173–188.

27. Vasek ME. Lying as a skill: The development of deception in children. Decept Perspect Hum Nonhum deceit. 1986 [no volume].

28. Greene J, Sommerville R, Nystrom LE, Darley JM, Cohen J. An fMRI Investigation of Emotional Engagement in Moral Judgment. Science (80-.). 2001; 293:2105–2108.

29. Greene JD, Paxton JM. Patterns of neural activity associated with honest and dishonest moral decisions. Proc Natl Acad Sci U S A. 2009; 106:12506–12511. [PubMed: 19622733]

30. Abe N. The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. Curr Opin Neurol. 2009; 22:594–600. [PubMed: 19786872]

31**. Zhu L, Jenkins AC, Set E, Scabini D, Knight RT, Chiu PH, King-Casas B, Hsu M. Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. Nat Neurosci. 2014; 17:1319–1321. First study to combine economic games and lesion method to address causal questions left unaddressed in existing neuroimaging studies of honesty and deception. [PubMed: 25174003]

32. Bhatt MA, Lohrenz T, Camerer CF, Montague PR. Neural signatures of strategic types in a two-person bargaining game. Proc Natl Acad Sci USA. 2010; 107:19720–19725. [PubMed: 21041646]

33. Baumgartner T, Fischbacher U, Feierabend A, Lutz K, Fehr E. The neural circuitry of a broken promise. Neuron. 2009; 64:756–770. [PubMed: 20005830]

34. Sun D, Chan CCH, Hu Y, Wang Z, Lee TMC. Neural correlates of outcome processing post dishonest choice: An fMRI and ERP study. Neuropsychologia. 2015; 68:148–157. [PubMed: 25582407]

35. Abe N, Greene JD. Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. J Neurosci. 2014; 34:10564–10572. [PubMed: 25100590]

36. Volz KG, Vogeley K, Tittgemeyer M, Von Cramon DY, Sutter M. The neural basis of deception in strategic interactions. Front Behav Neurosci. 2015; 9

37. Panasiti MS, Pavone EF, Mancini A, Merla A, Grisoni L, Aglioti SM. The motor cost of telling lies: Electrocortical signatures and personality foundations of spontaneous deception. Soc Neurosci. 2014; 9:573–589. [PubMed: 24979665]

38. Mazar N, Amir O, Ariely D. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. J Mark Res. 2008; 45:633–644.

39. Behrens TEJ, Hunt LT, Rushworth MFS. The computation of social behavior. Sci (New York, NY). 2009; 324:1160–1164.

40. Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci. 2008; 9:545–56. [PubMed: 18545266]

41. Saxe R, Kanwisher N. People thinking about thinking people The role of the temporo-parietal junction in “theory of mind”. Neuroimage. 2003; 19:1835–1842. [PubMed: 12948738]

42. Jenkins AC, Mitchell JP. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cereb Cortex. 2010; 20:404–410. [PubMed: 19478034]

43. Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the Ultimatum Game. Science. 2003; 300:1755–1758. [PubMed: 12805551]

44. Hsu M, Anen C, Quartz SR. The right and the good: distributive justice and neural encoding of equity and efficiency. Science. 2008; 320:1092–1095. [PubMed: 18467558]

45*. Haynes JD. Detecting deception from neuroimaging signals - a data-driven perspective. Trends Cogn Sci. 2008; 12:126–127. Offers an alternative approach to lie detection by focusing on neural correlates of deception, as opposed to understanding mechanistic details of the functional contribution of these regions. In many ways, a modern neuroscience update on the polygraph. [PubMed: 18308617]

46*. Schacter DL, Loftus EF. Memory and law: what can cognitive neuroscience contribute? Nat Neurosci. 2013; 16:119–123. Provides useful discussion of opportunities and challenges in focusing on memory processes, as opposed to cognitive control processes, to provide a neural marker of deception. [PubMed: 23354384]

47. Yang Z, Huang Z, Gonzalez-Castillo J, Dai R, Northoff G, Bandettini P. Using fMRI to decode true thoughts independent of intention to conceal. Neuroimage. 2014; 99:80–92. [PubMed: 24844742]

48. Benz A, Jäger G, Rooij van R. Game Theory and Pragmatics. 2005

Highlights

5 Past neural studies of honesty and deception have been criticized on methodological and conceptual grounds.

6 Studies from economics and ethology have used signaling games to study honesty and deception at a different level of analysis.

7 Signaling games clarify motivational processes underlying honest and deception actions.

8 These games provide a useful theoretical foundation for future neuroscientific investigations.

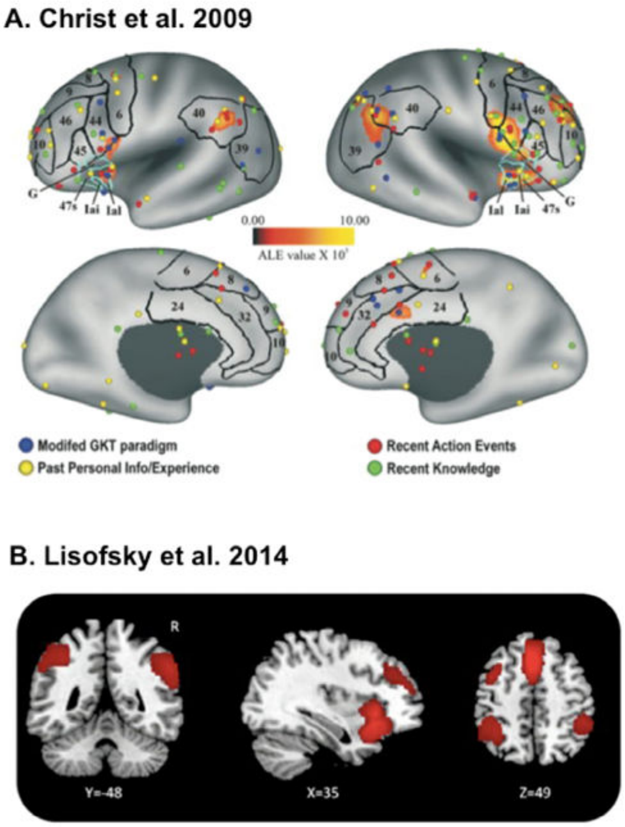

Figure 1. Neural correlates of honesty and deception

(A–B) Previous meta-analysis results of neural correlates of truth versus lie telling adapted from [15,16] implicating greater engagement of lateral prefrontal cortex, medial prefrontal cortex, insula, and lateral parietal cortex under deceptive compared to honest actions.

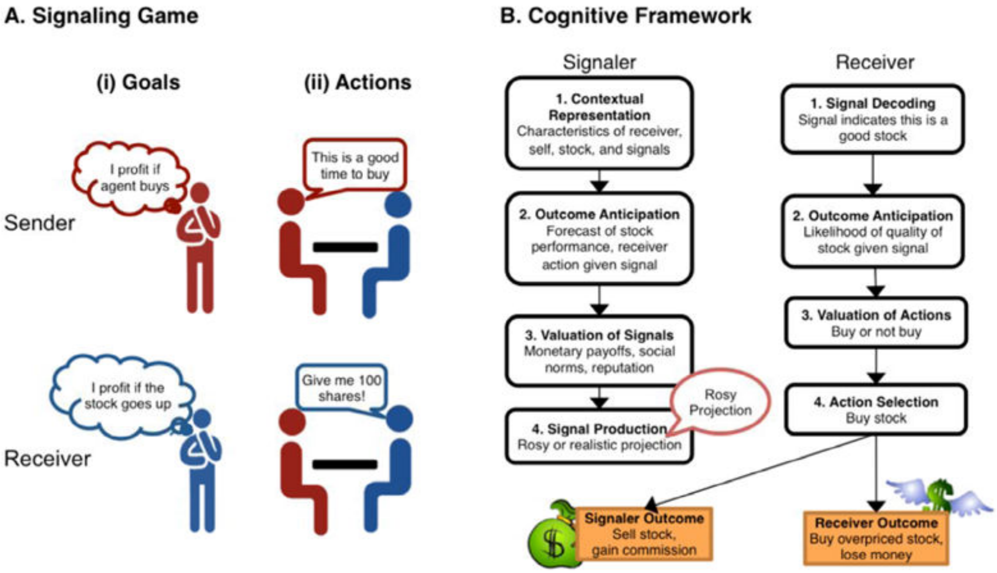
A. Signaling Game
(i) Goals
(ii) Actions
Sender
I profit if agent buys
This is a good time to buy
Receiver
I profit if the stock goes up
Give me 100 shares!
B. Cognitive Framework
Signaler
Receiver
1. Contextual Representation Characteristics of receiver, self, stock, and signals
2. Outcome Anticipation Forecast of stock performance, receiver action given signal
3. Valuation of Signals Monetary payoffs, social norms, reputation
4. Signal Production Rosy or realistic projection
Rosy Projection
1. Signal Decoding Signal indicates this is a good stock
2. Outcome Anticipation Likelihood of quality of stock given signal
3. Valuation of Actions Buy or not buy
4. Action Selection Buy stock
Signaler Outcome Sell stock, gain commission
Receiver Outcome Buy overpriced stock, lose money

Figure 2. Signaling games as cognitive probes

(A) An important feature of signaling games is in explicitly capturing the link between the goals on the part of the sender and the receiver, and the instrumental actions that are taken to reach the goals. For example, in an interaction involving a broker (sender) and an investor (receiver), whereas the goal of the investor is to choose a stock that will appreciate in value, the goal of the broker is to simply sell the stock. **(B)** Cognitive processes involved in a signaling interaction can be decomposed first into those involving the sender and the receiver, and then into a series of processes within each.

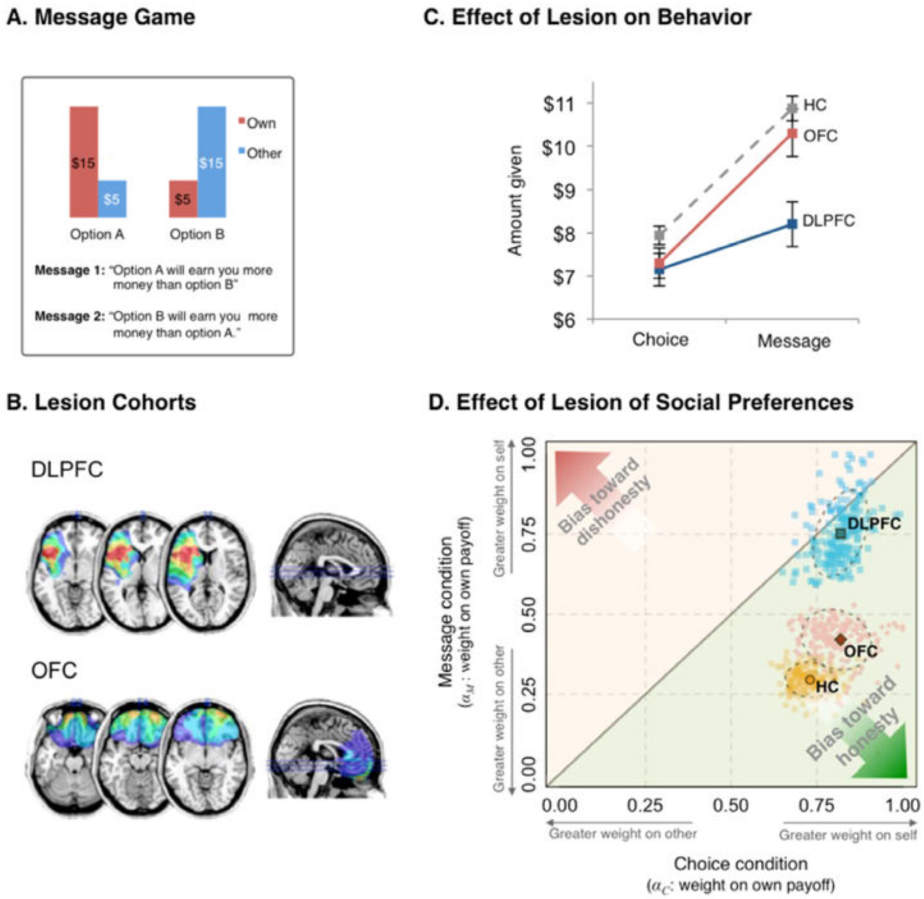

Figure 3.

(A) The message game involves the participant in the role of the signaler, who is presented with two options, A and B, associated with different monetary consequences. The signaler has furthermore two actions available to the participant in the form of two statements describing the monetary consequences of the options to the recipient. Specifically, the participants must choose between sending a truthful message (Message 2) that sacrifices economic self-interest in favor of honesty, or a false message (Message 1) that satisfies self-interest at the expense of being honest. **(B)** Lesion reconstruction overlay of patients with damage to DLPFC and OFC, respectively. **(C)** Compared to a Choice condition with identical monetary consequences but without the inclusion of honesty concerns, HC participants increased giving by $2.94±.44. In contrast, DLPFC cohort's giving increased by less than half this amount ($1.05±.43), which was significantly lower than those of the HC cohort. OFC participants were nearly identical to HCs, and were significantly different from DLPFC participants. **(D)** Computational modeling of preferences for (dis)honesty show significant weight on honesty for both OFC and HC, but not DLPFC, participants. Note however that, controlling for preferences over outcomes, none of the groups showed a bias toward deception. Adapted from Zhu et al. [31].

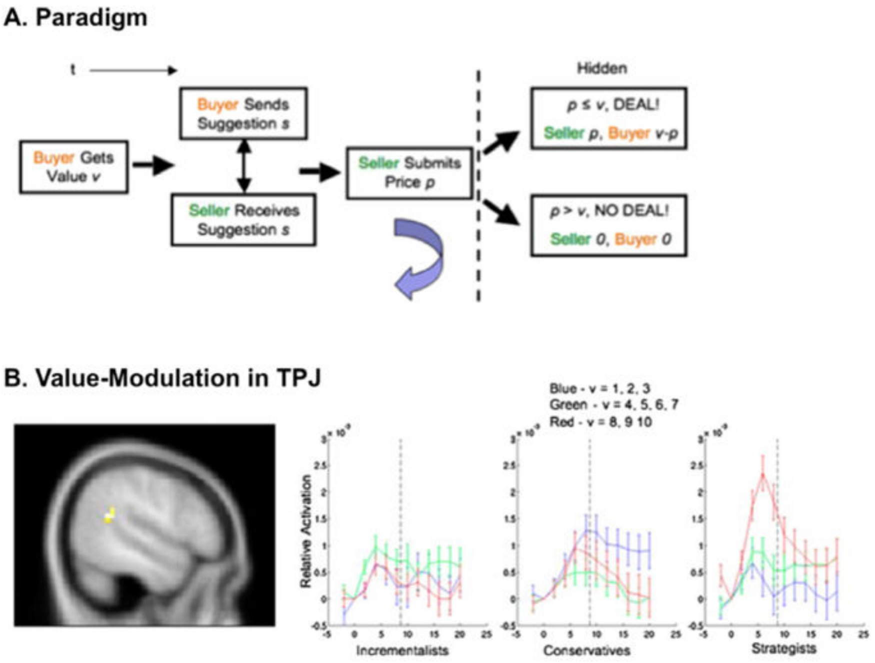

Figure 4.

(A) The bilateral bargaining task consists of a series of one-shot bargaining regarding a single unit of some good. The buyer, who values the good at *v*, suggests a selling price *s* to a seller. Upon receiving *s*, the seller submits a selling price *p*. If $p > v$, no deal occurs, otherwise the transaction is taken with seller receiving *p* and buyer $v - p$. Note in this case the buyer has an incentive to underreport the true *v*. **(B)** The value of the underlying good modulated activity in the right temporoparietal junction (rTPJ) previously implicated in theory of mind processes, but only for individuals who were strategic in behavioral reports of *v*. Adapted from Bhatt et al. [32].