

Trust Perception

P Mende-Siedlecki and A Todorov, Princeton University, Princeton, NJ, USA

© 2015 Elsevier Inc. All rights reserved.

In the opening scene of Shakespeare's *All's Well That Ends Well*, the countess imparts the following immortal words to her son: "Love all, trust a few, do wrong to none." In other words, while general prosociality may be granted indiscriminately, trust should hold a more privileged status in interpersonal interactions. The overwhelming abundance of neuroimaging research on trust perception over the past 15 years seems to have confirmed this hierarchy.

When determining whether or not another person is trustworthy, the most relevant pieces of evidence are their appearance and their behavior. As such, the vast majority of neuroscientific research on trust perception has focused on these two sources of information. More recent work has also begun to address how multiple sources of trust information are integrated together. Finally, researchers have attempted to gain insight into neuroendocrine influences on trust perception, through a focus on specific hormones (e.g., oxytocin and testosterone) that have been demonstrated to enhance and inhibit interpersonal trust.

Perceiving Trustworthiness in Faces

Traditionally, neuroimaging studies on social perception have associated the evaluation of facial trustworthiness with the amygdala (Todorov, Mende-Siedlecki, & Dotsch, 2013). Some of the earliest evidence for this link comes from neuropsychological research conducted by Adolphs, Tranel, and Damasio (1998), which demonstrated clear dissociations between patients with bilateral amygdala damage and normal controls with respect to evaluations of facial trustworthiness. Specifically, the bilateral amygdala patients rated untrustworthy-looking faces as being more trustworthy and approachable than did controls, brain-damaged controls, or unilateral amygdala lesion patients. These authors suggest that this difference can be attributed to deficits in extracting relevant social information from the face, as opposed to a more general decrement in face processing (Adolphs et al., 1998). Indeed, these results can be contrasted with more recent work suggesting that both acquired and developmental prosopagnosics provide standard trustworthiness evaluations (Quadflieg, Todorov, Lagusse, & Rossion, 2012; Todorov & Duchaine, 2008).

In the intervening years, numerous fMRI investigations have corroborated these initial findings, as have subsequent meta-analyses (Bzdok et al., 2011; Mende-Siedlecki, Said, & Todorov, 2013), which affirm the amygdala's contributions to the social evaluation of faces across numerous behavioral tasks and stimuli sets. That being said, this additional work has added new wrinkles to the narrative. For example, while the earliest of these neuroimaging studies observed a negative relationship between facial trustworthiness and amygdala activity (Engell, Haxby, & Todorov, 2007; Winston, Strange, O'Doherty, & Dolan, 2002), such that the amygdala responded most strongly to

untrustworthy-looking faces, more recent evidence points to a nonlinear relationship between amygdala activity and facial trustworthiness. Across a growing number of studies, researchers have observed enhanced amygdala activity to both untrustworthy- and trustworthy-looking faces, as compared to the more neutral faces towards the midpoint of the trustworthiness dimension (Mattavelli, Andrews, Asghar, Towler, & Young, 2012; Said, Baron, & Todorov, 2009; Said, Dotsch, & Todorov, 2010; Todorov, Baron, & Oosterhof, 2008; Todorov, Said, Oosterhof, & Engell, 2011).

A parallel line of research has sought to address the temporal dynamics of facial trustworthiness evaluation by measuring event-related potentials (ERPs; Dzhelyova, Perrett, & Jentsch, 2012; Kovács-Bálint, Stefanics, Trunk, & Hernádi, 2014; Marzi, Righi, Ottonello, Cincotta, & Viggiano, 2014; Rudoy & Paller, 2009; Yang, Qi, Ding, & Song, 2011). Though the initial findings of these studies are somewhat mixed, taken together, they suggest that facial trustworthiness exerts an early influence on visual processing (Dzhelyova et al., 2012; Marzi et al., 2014), dovetailing with behavioral work suggesting that evaluations of facial trustworthiness are made after only 100 milliseconds of exposure to faces (Willis & Todorov, 2006). Moreover, these studies demonstrate enhanced processing associated with untrustworthy faces, which persists throughout later latencies, potentially reflecting subsequent motivational consequences triggered by trustworthiness evaluation (Marzi et al., 2014; Yang et al., 2011). Finally, these authors posit a similarity between ERP responses to facial trustworthiness and previously observed responses associated with processing of facial expressions (Dzhelyova et al., 2012; Marzi et al., 2014; Yang et al., 2011). This finding has been interpreted as further evidence for the emotion overgeneralization hypothesis (Marzi et al., 2014; Yang et al., 2011), which suggests that the social evaluation of faces has emerged as an extension of the processes supporting the recognition and comprehension of emotional facial expressions (Montepare & Dobish, 2003; Oosterhof & Todorov, 2008; Said, Sebe, & Todorov, 2009; Todorov, 2008; Zebrowitz & Montepare, 2008). In fact, computational modeling of face trustworthiness shows that whereas untrustworthy-looking faces resemble angry faces, trustworthy-looking faces resemble happy faces (Oosterhof & Todorov, 2008).

Finally, we note that incidental cues associated with faces can influence evaluations of facial trustworthiness, as well. For example, peripheral distractor faces in a visual search paradigm (Frischen, Ferrey, Burt, Pistchik, & Fenske, 2012; Raymond, Fenske, & Westoby, 2005), faces associated with unreliable eye gaze in a gaze-cueing task (Bayliss & Tipper, 2006), and faces associated with response inhibition in a go/no-go task (Fenske, Raymond, Kessler, Westoby, & Tipper, 2005) are all rated as looking less trustworthy than corresponding control faces. The neural bases of the latter effect were recently uncovered by Doallo et al. (2012), who observed that not only were no-go faces perceived to be less trustworthy but also they

elicited enhanced activity in the amygdala during a subsequent recognition task. Moreover, both behavioral and neural indices of this devaluation were predicted by lateral orbitofrontal cortex (OFC) activity, such that greater recruitment of this inhibitory control mechanism during the go/no-go task was associated with both less trustworthy evaluations of and greater amygdala activity in response to no-go faces.

Perceiving Trustworthiness in Behavior

While momentary, incidental associations can certainly influence our perceptions of trustworthiness, intentional behavior has an arguably more profound impact. The majority of neuroimaging work within this domain has examined the brain's response to variations in interpersonal trust based on either behavior experienced in the context of economic games or behavior learned through observation or description.

Throughout the past decade, researchers in the field of neuroeconomics have tried to identify the neural underpinnings of a different form of interpersonal trust – trust that is built and lost during exchange-based interactions (Rilling, King-Casas, & Sanfey, 2008; Rilling & Sanfey, 2011). This line of work has overwhelmingly highlighted the role of the striatum in trust-based social decision-making (Delgado, Frank, & Phelps, 2005; King-Casas et al., 2005; Rilling et al., 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004a; Stanley et al., 2012; Wardle et al., 2013). Specifically, the caudate nucleus is engaged by reciprocated cooperation during economic games like the trust game and the prisoner's dilemma game (Delgado et al., 2005; King-Casas et al., 2005; Rilling et al., 2002), and moreover, activity in the caudate may come to represent a decision partner's reputational trust over time (King-Casas et al., 2005; Stanley et al., 2012; Wardle et al., 2013).

Additional work in this vein has highlighted the contributions of other brain areas involved in trust-based social decision-making. For example, several studies have noted that the anterior insula responds to violations of trust signaled by unreciprocated cooperation (Rilling et al., 2008; Sanfey et al., 2003). Moreover, neuropsychological work has shed further light on other brain regions necessary for normal trust-related behavior. For example, individuals with unilateral damage to the amygdala do not show decreases in trust in response to betrayals in the trust game (Koscik & Tranel, 2011), while patients with damage to the ventromedial prefrontal cortex (vmPFC) behave less trustworthily across a battery of economic games (Krajbich, Adolphs, Tranel, Denburg, & Camerer, 2009). In addition to the amygdala (Baumgartner, Heinrichs, Vöhlken, Fischbacher, & Fehr, 2008; Bhatt, Lohrenz, Camerer, & Montague, 2012; Stanley et al., 2012), other subcortical structures have been implicated in processes underlying trust decisions, including the parahippocampal gyrus (Bhatt et al., 2012) and ventral tegmental area (Krueger et al., 2007). Similarly, in addition to the vmPFC, other prefrontal regions have been implicated, including the dorsomedial prefrontal cortex (dmPFC; Krueger et al., 2007; Rilling et al., 2004b) and OFC (Rilling et al., 2002, 2004a). Finally, the downstream consequences of trustworthy and untrustworthy behavior are observable neurally as well, with distinct brain systems showing enhanced responses to faces of defectors and cooperators

(Singer, Kiebel, Winston, Dolan, & Frith, 2004) and individuals who either exceeded or fell short of expectations (Chang & Sanfey, 2009).

Perceptions of trustworthiness are highly related to evaluations of morality (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Cottrell, Neuberg, & Li, 2007; Goodwin, Piazza, & Rozin, 2014; Leach, Ellemers, & Barreto, 2007), a process investigated by a separate but related area of neuroimaging research. Typically, in these studies, participants learn about the moral character of other individuals through brief, textual descriptions of their behavior. Interestingly, making moral judgments based on behavior is associated with several brain regions also implicated in studies on trust-based social decision-making. For example, the vmPFC and OFC display increased activity in response to positive, moral behaviors (Cloutier, Ambady, Meagher, & Gabrieli, 2012; Tsukiura & Cabeza, 2011; Zahn et al., 2009), while the anterior insula responds more strongly to negative, immoral behaviors (Tsukiura & Cabeza, 2011; Zahn et al., 2009).

Integrating Across Multiple Sources of Trust Information

Recently, researchers have begun to investigate the neural dynamics supporting the integration of multiple pieces of trust-related information, which may originate from different sources. In one of the earliest such studies, Delgado et al. (2005) observed that reading brief but evocative vignettes regarding trust game partners' moral character prior to interacting with them influenced both behavioral and neural markers of interpersonal trust during the game. Specifically, although payoff rates were ultimately equal across reputationally 'good,' 'bad,' and 'neutral' partners, participants made more risky investments in the 'good' partner. Moreover, these authors observed diminished caudate responses associated with feedback from both the 'good' and 'bad' partners, compared with the neutral partner, suggesting that having strong social and/or moral priors reduces one's dependence on more immediate feedback (Delgado et al., 2005, though see Wardle et al., 2013). Subsequent work has explored similar biases in trust-based social decision-making triggered by initial impressions stemming from facial trustworthiness (Chang, Doll, van't Wout, Frank, & Sanfey, 2010; Kim, Choi, & Jang, 2012; Van't Wout & Sanfey, 2008), race (Stanley et al., 2012), previous interactions (Fareri, Chang, & Delgado, 2012), and personality-related reputational priors (Fouragnan et al., 2013).

A number of other studies have examined how the brain integrates learned behaviors with other sources of trust information. One of the first such inquiries was an EEG experiment conducted by Rudoy and Paller (2009), where participants formed impressions of trustworthy and untrustworthy faces paired with positive and negative personality attributes. The authors observed that the neural correlates of appearance-based and behavior-based processing were dissociable and, further, that the appearance-based influences preceded the influence of behavior-based information (Rudoy & Paller, 2009). A subsequent fMRI study tackled a similar topic – how the brain updates initial appearance-based impressions based

on subsequently learned behavioral information (Baron, Gobini, Engell, & Todorov, 2011). Baron and colleagues noted that dmPFC activity in response to behavioral information correlated with a subsequent measure of impression change but, more interestingly, that the strength of this learning was modulated by the strength of the initial amygdala response to the faces prior to learning. Specifically, participants who showed the strongest initial amygdala responses to facial trustworthiness also showed the weakest correlations between dmPFC activity and behavioral learning effects.

Finally, a more recent program of research has examined the neural dynamics that underlie the updating of behavior-based impressions regarding trustworthiness in response to new, inconsistent behavioral information (Mende-Siedlecki, Cai, & Todorov, 2013; Mende-Siedlecki, Baron, & Todorov, 2013). After an initial investigation identified an extended network of regions involved in updating (Mende-Siedlecki, Cai, & Todorov, 2013), subsequent work has tried to characterize the general principles guiding activity in this network (Mende-Siedlecki, Baron, & Todorov, 2013). Behavioral research suggests that certain behaviors impact our evaluations of people more heavily than others (Skowronski & Carlston, 1987; Wojciszke, Brycz, & Borke-nau, 1993). For example, immoral or competent behaviors are more heavily weighted when forming impressions than moral or incompetent behaviors. Mende-Siedlecki and colleagues demonstrated that these asymmetries in diagnosticity stem from differences in the perceived frequency of behaviors. Specifically, behaviors that are perceived to be rare (including untrustworthy and immoral behaviors) are likely to be more diagnostic, to provoke larger updates in our impressions of other people, and to drive activity in diagnosticity-sensitive regions (the left ventrolateral PFC, left inferior frontal gyrus, and left superior temporal sulcus) more strongly than their nondiagnostic counterparts (Mende-Siedlecki, Baron, & Todorov, 2013).

Neuroendocrine Perspectives on Trust Perception

As the first three sections suggest, the neural bases of trust perception have been well documented over the past 15 years. A more recent pursuit has been to assess how these markers respond to the influence of specific hormones related to trust behavior. Much of this work has been concerned with the role of oxytocin, while subsequent research has addressed the impact of testosterone, as well.

Oxytocin is a neuropeptide known to play a critical role in attachment and affiliative behaviors in both humans and non-human mammals (Bartz & Hollander, 2006; Heinrichs & Domes, 2008). Following initial indications that oxytocin is released in response to both trusting and trustworthy behaviors (Zak, Kurzban, & Matzner, 2005), Kosfeld and colleagues sought to determine the explicit impact of intranasal administration of oxytocin on interpersonal trust. In the context of a trust game, participants displayed greater trust in their exchange partners, as quantified by significantly larger amounts of money transferred to those partners on average, as compared with participants in the placebo condition. Notably, these authors suggest that this result was not merely due to

an increased tolerance for risky behavior in general, but an increased willingness to accept the social risks associated with interacting with another person (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005).

These effects have been extended beyond the domain of monetary exchange, with recent evidence suggesting that oxytocin increases trust when confidential information is being shared (Kéri & Kiss, 2011; Mikolajczak, Pinon, Lane, de Timary, & Luminet, 2010), as well as perceptions of facial trustworthiness (Theodoridou, Rowe, Penton-Voak, & Rogers, 2009). Other work has managed to shed light on the mechanism of oxytocin's influence on trust, observing that oxytocin promotes more non-personal (i.e., situational) interpretations of untrustworthy behaviors (Klackl, Pfundmair, Agroskin, & Jonas, 2013). Subsequent studies have sought to clarify the neural effects of oxytocin on trust behavior, as well. For example, Baumgartner and colleagues (2008) observed that participants who received intranasal administration of oxytocin prior to playing a trust game showed less change in trust following several incidences of betrayal by their partners (as compared to the placebo group). Critically, this behavioral difference in the oxytocin group was linked to reduced activity in the amygdala and dorsal striatum – two brain structures heavily implicated in trust-related social decision-making (Baumgartner et al., 2008).

This is not to suggest that oxytocin's trust-enhancing influence persists over and above the boundaries of contextual factors and individual differences. A recent review by Bartz and colleagues (2011) details numerous empirical demonstrations in which the positive effects of oxytocin on interpersonal trust are moderated by group membership, familiarity, or social knowledge. For example, oxytocin's trust-facilitating effects do not seem to persist when contextual information suggests that a partner is likely to be unreliable (Mikolajczak et al., 2010). In fact, in individuals with borderline personality disorder, oxytocin has actually been demonstrated to decrease trust (Bartz et al., 2011).

Finally, a parallel line of research has assessed the role of testosterone in interpersonal trust. By and large, this work suggests that testosterone inhibits trust – especially in individuals who are particularly prone to trusting behavior – as a result of an increase in social vigilance (Bos, Terburg, & Van Honk, 2010). In this particular study, female participants who were administered testosterone rated unfamiliar faces as being significantly less trustworthy than participants who received a placebo. In a related neuroimaging experiment, the trust-reducing effect of testosterone was associated with enhanced amygdala activity to neutral faces, along with a decoupling of functional connectivity between the amygdala and OFC. Intriguingly, additional behavioral work suggests that though testosterone may inhibit trust, it also seems to facilitate reciprocity (Boksem et al., 2013), suggesting that the hormone may enhance vigilant (and potentially antisocial) behaviors in the face of threats but that under less threatening conditions, testosterone may promote more prosocial activity.

Summary

The neural bases of interpersonal trust have been studied in a variety of contexts, through the use of multiple methodologies

and based upon insights from diverse theoretical backgrounds. The amygdala is strongly implicated throughout this variegated body of work, from the evaluation of facial trustworthiness to trust-based social decision-making. However, support from other brain structures seems to vary based on the design employed and the domain in which trustworthiness is manipulated. Future work should continue to explore how multiple pieces of trust-related information are integrated (especially when they conflict), how representations of trust are dynamically updated over time, and how these representations are modulated by top-down and bottom-up factors.

See also: INTRODUCTION TO SOCIAL COGNITIVE NEUROSCIENCE: A Neural Network for Moral Decision Making; Cooperation and Fairness; Face Perception: Extracting Social Information from Faces: The Role of Static and Dynamic Face Information; Person Knowledge and Attribution; Social Decision Making; The Amygdala and Social Perception; The Neural Correlates of Social Cognition and Social Interaction; INTRODUCTION TO SYSTEMS: Face Perception.

References

- Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.
- Baron, S. G., Gobbini, M. I., Engell, A. D., & Todorov, A. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, *6*, 572–581.
- Bartz, J. A., & Hollander, E. (2006). The neuroscience of affiliation: Forging links between basic and clinical research on neuropeptides and social behavior. *Hormones and Behavior*, *50*, 518–528.
- Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011a). Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences*, *15*, 301–309.
- Bartz, J. A., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., et al. (2011b). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *6*, 556–563.
- Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, *58*, 639–650.
- Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments: Should eye trust you? *Psychological Science*, *17*, 514–520.
- Bhatt, M. A., Lohrenz, T., Camerer, C. F., & Montague, P. R. (2012). Distinct contributions of the amygdala and parahippocampal gyrus to suspicion in a repeated bargaining game. *Proceedings of the National Academy of Sciences*, *109*, 8728–8733.
- Boksem, M. A., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., et al. (2013). Testosterone inhibits trust but promotes reciprocity. *Psychological Science*, *24*, 2306–2314.
- Bos, P. A., Terburg, D., & Van Honk, J. (2010). Testosterone decreases trust in socially naive humans. *Proceedings of the National Academy of Sciences*, *107*, 9991–9995.
- Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135–143.
- Bzdok, D., Langner, R., Caspers, S., Kurth, F., Habel, U., Zilles, K., et al. (2011). ALE meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure and Function*, *215*, 209–223.
- Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*, 87–105.
- Chang, L. J., & Sanfey, A. G. (2009). Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic exchange. *Frontiers in Behavioral Neuroscience*, *3*, 1–12.
- Cloutier, J., Ambady, N., Meagher, T., & Gabrieli, J. D. E. (2012). The neural substrates of person perception: Spontaneous use of financial and moral status knowledge. *Neuropsychologia*, *50*, 2371–2376.
- Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology*, *92*, 208–231.
- Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611–1618.
- Doallo, S., Raymond, J. E., Shapiro, K. L., Kiss, M., Eimer, M., & Nobre, A. C. (2012). Response inhibition results in the emotional devaluation of faces: Neural correlates as revealed by fMRI. *Social Cognitive and Affective Neuroscience*, *7*, 649–659.
- Dzhelyova, M., Perrett, D. I., & Jentzsch, I. (2012). Temporal dynamics of trustworthiness perception. *Brain Research*, *1435*, 81–90.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. *Journal of Cognitive Neuroscience*, *19*, 1508–1519.
- Fareri, D. S., Chang, L. J., & Delgado, M. R. (2012). Effects of direct social experience on trust decisions and neural reward circuitry. *Frontiers in Neuroscience*, *6*.
- Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., & Tipper, S. P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychological Science*, *16*, 753–758.
- Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., & Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *The Journal of Neuroscience*, *33*, 3602–3611.
- Frischen, A., Ferrey, A. E., Burt, D. H., Pistchik, M., & Fenske, M. J. (2012). The affective consequences of cognitive inhibition: Devaluation or neutralization? *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 169.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*(1), 148–168. <http://dx.doi.org/10.1037/a0034726>.
- Heinrichs, M., & Domes, G. (2008). Neuropeptides and social behaviour: Effects of oxytocin and vasopressin in humans. *Progress in Brain Research*, *170*, 337–350.
- Kéri, S., & Kiss, I. (2011). Oxytocin response in a trust game and habituation of arousal. *Physiology and Behavior*, *102*, 221–224.
- Kim, H., Choi, M. J., & Jang, I. J. (2012). Lateral OFC activity predicts decision bias due to first impressions during ultimatum games. *Journal of Cognitive Neuroscience*, *24*(2), 428–439.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*, 78–83.
- Klackl, J., Pfundmair, M., Agroskin, D., & Jonas, E. (2013). Who is to blame? Oxytocin promotes nonpersonalistic attributions in response to a trust betrayal. *Biological Psychology*, *92*, 387–394.
- Kosciak, T. R., & Tranel, D. (2011). The human amygdala is necessary for developing and expressing normal interpersonal trust. *Neuropsychologia*, *49*, 602–611.
- Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*, 673–676.
- Kovács-Bálint, Z., Stefanics, G., Trunk, A., & Hernádi, I. (2014). Automatic detection of trustworthiness of the face: A visual mismatch negativity study. *Acta Biologica Hungarica*, *65*, 1–12.
- Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *The Journal of Neuroscience*, *29*, 2188–2192.
- Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., et al. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, *104*, 20084–20089.
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*, 234–249.
- Marzi, T., Righi, S., Ottonello, S., Cincotta, M., & Viggiano, M. P. (2014). Trust at first sight: Evidence from ERPs. *Social Cognitive and Affective Neuroscience*, *9*, 63–72.
- Mattavelli, G., Andrews, T. J., Asghar, A. U., Towler, J. R., & Young, A. W. (2012). Response of face-selective brain regions to trustworthiness and gender of faces. *Neuropsychologia*, *50*, 2205–2211.
- Mende-Siedlecki, P., Baron, S. G., & Todorov, A. (2013). Diagnostic value underlies asymmetric updating of impressions in the morality and ability domains. *The Journal of Neuroscience*, *33*, 19406–19415.
- Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, *8*, 623–631.
- Mende-Siedlecki, P., Said, C. P., & Todorov, A. (2013). The social evaluation of faces: A meta-analysis of functional neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *8*, 285–299.
- Mikolajczak, M., Pinon, N., Lane, A., de Timary, P., & Luminet, O. (2010). Oxytocin not only increases trust when money is at stake, but also when confidential information is in the balance. *Biological Psychology*, *85*, 182–184.

- Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*, 237–254.
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*, 11087–11092.
- Quadflieg, S., Todorov, A., Laguesse, R., & Rossion, B. (2012). Normal face-based judgements of social characteristics despite severely impaired holistic face processing. *Visual Cognition, 20*, 865–882.
- Raymond, J. E., Fenske, M. J., & Westoby, N. (2005). Emotional devaluation of distracting patterns and faces: A consequence of attentional inhibition during visual search? *Journal of Experimental Psychology: Human Perception and Performance, 31*, 1404.
- Rilling, J. K., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron, 35*, 395–405.
- Rilling, J. K., King-Casas, B., & Sanfey, A. G. (2008). The neurobiology of social decision-making. *Current Opinion in Neurobiology, 18*, 159–165.
- Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23–48.
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004a). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *NeuroReport, 15*, 2539–2543.
- Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004b). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage, 22*, 1694–1703.
- Rudoy, J. D., & Paller, K. A. (2009). Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. *Frontiers in Human Neuroscience, 3*.
- Said, C. P., Baron, S. G., & Todorov, A. (2009). Nonlinear amygdala response to face trustworthiness: Contributions of high and low spatial frequency information. *Journal of Cognitive Neuroscience, 21*, 519–528.
- Said, C. P., Dotsch, R., & Todorov, A. (2010). The amygdala and FFA track both social and non-social face dimensions. *Neuropsychologia, 48*, 3596–3605.
- Said, C. P., Sebe, N., & Todorov, A. (2009). Structural resemblance to emotional expressions predicts evaluation of emotionally neutral faces. *Emotion, 9*, 260.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*, 1755–1758.
- Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41*, 653–662.
- Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*, 689–699.
- Stanley, D., Sokol-Hessner, P., Fareri, D., Perino, M., Degladi, M., Banaji, M., et al. (2012). Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 367*, 744–753.
- Theodoridou, A., Rowe, A. C., Penton-Voak, I. S., & Rogers, P. J. (2009). Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior, 56*, 128–132.
- Todorov, A. (2008). Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In A. Kingstone & M. Miller (Eds.), *The year in cognitive neuroscience 2008*, Annals of the New York Academy of Sciences, 1124, pp. 208–224. Blackwell Publishing Inc.
- Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: A model based approach. *Social Cognitive and Affective Neuroscience, 3*, 119–127.
- Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology, 25*, 395–410.
- Todorov, A., Mende-Siedlecki, P., & Dotsch, R. (2013). Social judgments from faces. *Current Opinion in Neurobiology, 23*, 373–380.
- Todorov, A., Said, C. P., Oosterhof, N. N., & Engell, A. D. (2011). Task-invariant brain responses to the social value of faces. *Journal of Cognitive Neuroscience, 23*, 2766–2781.
- Tsukiura, T., & Cabeza, R. (2011). Shared brain activity for aesthetic and moral judgments: Implications for the Beauty-is-Good stereotype. *Social Cognitive and Affective Neuroscience, 6*, 138–148.
- Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*, 796–803.
- Wardle, M. C., Fitzgerald, D. A., Angstadt, M., Sripada, C. S., McCabe, K., & Phan, K. L. (2013). The caudate signals bad reputation during trust decisions. *PLoS One, 8*, e68884.
- Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological Science, 17*, 592–598.
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*, 277–283.
- Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology, 64*, 327–335.
- Yang, D., Qi, S., Ding, C., & Song, Y. (2011). An ERP study on the time course of facial trustworthiness appraisal. *Neuroscience Letters, 496*, 147–151.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., et al. (2009). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex, 19*, 276–283.
- Zak, P. J., Kurzban, R., & Matzner, W. T. (2005). Oxytocin is associated with human trustworthiness. *Hormones and Behavior, 48*, 522–527.
- Zebrowitz, L. A., & Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass, 2*, 1497–1517.