Understanding identification-based trust in light of attachment: Meta-analytic neuroimaging evidence

Yan Wu[1,2], Apoorva Veerareddy[3], Mary R. Lee, Gabriele Bellucci[5], Julia A. Camilleri[6,7], Simon B. Eickhoff[6,7], Frank Krueger[3,8*]

[1] Department of Psychology, College of Education, Hangzhou Normal University, Hangzhou 311121, China

[2] Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou Normal University, Hangzhou 311121, China

[3] School of Systems Biology, George Mason University, Fairfax, VA, USA

[4] US Department of Veteran Affairs, Washington, DC, USA

[5] Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

[6] Institute for Neuroscience and Medicine (INM-7), Research Center Jülich, Germany

[7] Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University Düsseldorf, Germany

[8] Department of Psychology, University of Mannheim, Mannheim, Germany

***Correspondence:**

Frank Krueger

fkrueger@gmu.edu (F. Krueger)

Manuscript Information:

Number of words in abstract: 167

Number of words in main text: 9795

Number of tables/ figures: 3/2

Abstract

Trust is vital for establishing social relationships and is a crucial precursor for attachment. Recent investigations explored separately the neuropsychological basis of trust in strangers (measured by the trust game) and attachment (measured through attachment figures). However, a direct empirical support for the coming underlying neural mechanisms between trust and attachment is missing. Here, we conducted a coordinate-based meta-analysis on functional magnetic resonance imaging studies on trust and attachment using the activation likelihood estimation approach. Our results demonstrated that decisions to trust strangers in repeated interactions (i.e., identification-based trust) engaged the ventral striatum (vSTR, part of the mesolimbic dopaminergic pathway) likely signaling the expected rewarding outcomes following trust decisions. Further, both feedbacks about repaid trust in repeated interactions and attachment engaged the dorsal striatum (dSTR, part of the nigrostriatal dopaminergic pathway) likely encoding learning dynamics. Our findings suggest that identification-based trust can be understood in the light of attachment, involving the mesocorticolimbic "reward" pathway (vSTR) and nigrostriatal "habit formation" pathway (dSTR) in building and sustaining social relationships.

Keywords: interpersonal trust; attachment; identification-based trust; mesolimbic pathway; nigrostriatal pathway, mesocortical pathway

1. Introduction

Interpersonal trust is essential to many human interactions; it encompasses a willingness to accept *vulnerability or uncertainty* based on positive expectations regarding another person's behavior (Rousseau et al., 1998). Over the past decades, trust researchers have identified precursors and determinants of trust behavior, while lately shifting the focus to the neuropsychological mechanisms of trust employing functional magnetic resonance imaging (fMRI) (Balliet & Van Lange, 2013; Fehr et al., 2005; Johnson & Mislin, 2011; Riedl & Javor, 2012). Several single and meta-analytic studies have examined the neural substrates of trust (Bellucci et al., 2017, 2018; Delgado et al., 2005; Krueger et al., 2007; McCabe et al., 2001). Recently, a *neuropsychoeconomic (NPE) model of trust* was proposed that synthesizes the neural findings into three types of trust —*calculus-, knowledge-, and identification-based trust* (Lewicki & Bunker, 1995)— with the presumption that the neural basis for identification-based trust overlaps with the human *attachment system* (Krueger & Meyer-Lindenberg, 2019).

Attachment, a strong and lasting psychological connectedness between two people, which initially represents parent-child relationships, also portrays other forms of affectional bonds beyond infancy and close relationships (Ainsworth, 1989). Attachment theory suggests that close social bonds are characterized by seeking proximity and feeling of security to the attachment figures, which provides affect-regulatory benefits (Bowlby, 1973, 1979; Laurita et al., 2019; Mikulincer & Shaver, 2003). Early social encounters with significant others contribute to attachment formation and form different internal working models, which influence how social relationships are initiated and sustained over the lifespan (Long et al., 2020; Ziv, 2005). An increasing number of neuroscience studies has explored the underlying neurobiology of attachment, and neuropsychological models of human attachment have been developed (Antonucci et al., 2018; Bosmans et al., 2020; Long et al., 2020; Numan & Young, 2016; Ran & Zhang, 2018; Strathearn et al., 2019). It remains, however, unclear whether the same brain regions underlying attachment behaviors are also engaged by interpersonal trust —specially identification-based trust.

1.1 Behavioral measures and neuropsychological mechanisms of interpersonal trust

One of the most common research approaches to measure trust behavior in human interactions is through the employment of the *trust game* (TG) (Berg et al., 1995; Camerer, 2003). The TG is a two-player economic exchange game, with one player designated as the *trustor* and the other one as the *trustee*. In the standard game version, both players are given an initial monetary endowment, and players complete three sequential stages. During the *trust stage*, the trustor decides whether or not to transfer any portion of her endowment to the trustee —with any transferred amount multiplied (usually tripled) by the experimenter. During the *reciprocity stage*, the trustee decides whether or not to return some portion of the received money back to the trustor. Finally, during the *feedback stage*, both players are informed about their payoffs, so the trustor is informed about the trustee's decision. The amount transferred by the trustor measures trust, whereas the amount by the trustee reciprocity (a proxy of the trustee's trustworthiness). This game can be played over single or multiple iterations allowing to measure distinctive trust types (Cochard et al., 2004). On the one hand, the *one-round TG* allows measuring trust propensity toward an anonymous partner —capturing calculus-based trust (i.e., performing calculations about the costs and benefits). On the other hand, the *multi-round TG* enables gauging trust-based dynamics —capturing knowledge-based trust (i.e., gaining knowledge about the trustee to advance the trust relationship) and identification-based trust (i.e., developing a rewarding identification with the trustee) (Lewicki & Bunker, 1995).

Based on a previous coordinate-based meta-analysis of TG fMRI studies (Bellucci et al., 2017), trust decisions in the multi-round TG (i.e., identification-based trust) consistently activate the ventral striatum (vSTR, nucleus accumbens). The vSTR is a vital region of the *mesolimbic dopaminergic pathway* probably implicated in predictions of a partner's most likely behavior based on the partner's reputation (Delgado et al., 2005; Fareri et al., 2012; King-Casas et al., 2005). Whereas trust decisions in the one-round TG (i.e., calculus-based trust) consistently activate the right anterior insula (AI) —a core brain region of the *salience network* likely encoding aversive feelings such as a sense of aversion evoked by the risk of betrayal (Aimone et al., 2014; Bohnet et al., 2008). In contrast, the dorsal striatum (dSTR, caudate nucleus) —a central brain region of the *nigrostriatal dopaminergic pathway*— was found consistently recruited during the feedback stage of the multi-round TG, presumably related to learning dynamics associated with belief updates about the partner's behavior (Bellucci et al., 2017). The trustor may therefore form beliefs about the behavior of the trustee through repeated interactions and base his/her decisions on the partner's trustworthiness. The initial uncertainty about the outcome of a trust decision is reduced, as the behavior of the partner becomes predictable on the basis of more stable beliefs about the partner's character (i.e., her trustworthiness).

The NPE model assumes five trust components (treachery, reward, uncertainty, strategy, and trustworthiness) associated with three psychological processes (i.e., motivation, affect, and cognition) involving core brain regions rooted in domain-general large-scale brain networks (Krueger & Meyer-Lindenberg, 2019). The expectation of *reward* (motivation) companied with the risk of *treachery* (affect) causes *uncertainty*, which leads to the vulnerability of trusting another person. Two different kinds of *bounded rationality* (cognition) can be used to minimize uncertainty. In the instance of *economic rationality*, trustors with *extrinsic incentives* (i.e., self-regarding interest) can adopt a context-based *strategy* to gain personal benefits, thus removing uncertainty by transforming economic risk of treachery to economically positive expectations of reciprocity. In the instance of *social rationality*, trustors with *intrinsic incentives* can evaluate the relationship-based *trustworthiness* to contribute to the relationship's success, hence removing uncertainty by transforming social risk of treachery to socially positive expectations of reciprocity.

The NPE model assumes that trust evolves through repeated interactions. The evolution of trust starts with *calculus-based trust*, driven mainly by the *salience network,* including key regions such as AI, amygdala, and dorsal anterior cingulate cortex (dACC). It transfers then to knowledge-based trust, which is driven mainly by the *central-executive network* (comprising the dorsolateral PFC (dlPFC) and the ventrolateral PFC (vlPFC) to adopt a context-based strategy) or the *default-mode network* (incorporating essential brain regions such as the temporoparietal junction (TPJ) and dorsomedial PFC (dmPFC) to evaluate trustworthiness). Finally, *identification-based trust* evolves, which is driven mainly by a *dopaminergic reward pathway system* comprising the *mesolimbic pathway* (connecting the ventral tegmentum area (VTA) in the midbrain to the nucleus accumbens and olfactory tubercle in the vSTR), the *mesocortical pathway* (linking the VTA to the PFC including the ventromedial PFC (vmPFC)), and the *nigrostriatal pathway* (coupling the substantia nigra (SN) in the midbrain to the caudate nucleus and putamen in the dSTR) (Krueger & Meyer-Lindenberg, 2019).

While the NPE model of trust suggests that repeated, positive interactions lead to identification-based trust by establishing shared identity and affective connection between the interacting partners (Krueger & Meyer-Lindenberg, 2019), direct empirical support for the overlap of identification-based trust (as measured with the multi-round TG) with attachment is still missing. As trust building among strangers has become constitutive of modern society, it is important to develop a better understanding about the trust-building processes. Exploring the commonalities between identification-based trust and attachment provides opportunities to gather knowledge about trust dynamics and has the potential to create central implications for social behavior research.

1.2 Behavioral measures and neuropsychological mechanisms of attachment

Over the past few decades, compelling evidence on the neuropsychological underpinnings of attachment emerged, utilizing experimental paradigms that require to process facial images (Atzil et al., 2011; Barrett et al., 2012), videos (Strathearn et al., 2008; Wan et al., 2014), or names (Ortigue et al., 2007) of attachment figures (e.g., mothers viewing their babies (Hoekzema et al., 2017; Michalska et al., 2014) or adults their romantic partners (Aron et al., 2005; Bartels & Zeki, 2000)). Both qualitative and quantitative fMRI meta-analysis studies revealed coherent key brain regions involved in attachment — highlighting the involvement of the dopaminergic reward-related brain system in processing attachment figures.

A previous qualitative meta-analysis, for example, uncovered a common dopaminergic reward-related brain system that involves both dopamine and oxytocin receptors (e.g., VTA and caudate nucleus) independently of maternal or romantic attachment (Ortigue et al., 2010). Another coordinate-based meta-analysis on maternal attachment demonstrated the involvement of regions underlying reward and maternal motivation and promoting behavioral approach, including the STR (e.g., putamen and caudate) as well as the lateral and medial globus pallidus (Rigo et al., 2019). Finally, a multimodal voxel-based meta-analysis showed consistent activation patterns in the bilateral anterior insula extending to the thalamus, putamen, caudate, and amygdala in mothers processing stimuli from their own vs. unknown infants (Rocchetti et al., 2014).

Based on the existing neurobiological evidence, Strathearn (2011; 2019) proposed a neurobiological model involving three systems underlying human attachment: (i) the *oxytocinergic "affiliation" pathway system*, (ii) the *dopamineergic "reward/ reinforcement"* (mesocorticolimbic) and "*habit formation"* (nigrostriatal) *pathway system*, and (iii) the *glucocorticoid stress response system*. The *oxytocinergic pathway system* reflects an evolutionarily ancient system mediating human affiliation, which plays an essential role in attachment formation, stress regulation, and affiliative enhancement with attachment figures (Feldman et al., 2016; Panksepp, 2004). Moreover, oxytocin plays a crucial role in regulating social cognition and forming social affiliative behavior (Ferguson et al., 2002) as well as in highlighting the salience, reinforcing the value of social cues and upregulating in-group cooperation (De Dreu & Kret, 2016). Oxytocinergic neurons project centrally to brain regions important for social and maternal behaviors, including the medial preoptic area, stria terminal bed nucleus, vSTR, and VTA (Numan, 2006).

The *dopaminergic (mesolimbic, mesocortical, and nigrostriatal) pathway system* functions as essential elements in the brain reward system that modulate goal-directed behavior (Arias-Carrión et al., 2010). Overall, these pathways are involved in reinforcement stimulus–reward learning, decision-making based on future predicted reward (McClure et al., 2003), and long-term conditioning of preference for attachment figures (Strathearn, 2011). Dopaminergic neurons generally originate in midbrain VTA and SN and project into several brain regions, incorporating vSTR, dSTR, and ACC (P. R. Montague et al., 2004). The *mesolimbic pathway* originates in SN/VTA and projects to vSTR to mediate reward-related adaptation and learning (Pessiglione et al., 2006). It encodes a teaching signal associated with the prediction of reward outcomes (Schultz et al., 1997) and is imperative for memory and motivational behaviors (Bromberg-Martin et al., 2010). The *mesocortical pathway* originates in VTA and projects to PFC, where it mediates emotional and motivational responses (Lapish et al., 2007) and transmits signals about reward or salience (Lavin, 2005). The *nigrostriatal pathway* connects the SN in the midbrain with the dSTR (i.e., caudate nucleus and putamen) in the forebrain (Bourdy et al., 2014). This pathway also participates in reward functions such as the reinforcement of memory consolidation (Wise, 2009) and the guidance for behavioral choice (Howard et al., 2017).

The *glucocorticoid stress response system* plays a significant role in regulating the immune and stress response (Whirledge & Cidlowski, 2013). Glucocorticoids are steroid hormones that contribute to the physiological stress response, providing negative feedback inhibition of the hypothalamus-pituitary-adrenal (HPA) system and restoring homeostasis (Strathearn, 2011). In humans, deficits in early caregiving have been associated with exaggerated increases in cortisol in response to stress (Albers et al., 2008), whereas secure attachment has been linked with lower cortisol levels in response to stress (Kuo et al., 2019).

The oxytocinergic, dopaminergic, and glucocorticoid systems underlying human attachment are also crucial for trust behavior. For example, intranasal administration of oxytocin leads to increased trust toward strangers (Kosfeld et al., 2005) and hinders changes in trust behavior after a betrayal (Baumgartner et al., 2008). The latter study also revealed that participants receiving oxytocin showed decreases in responses in the amygdala (salience network) and caudate nucleus (nigrostriatal dopaminergic pathway) —suggesting that oxytocin impairs learning mechanisms after experiencing breach (via the amygdala) or repayment of trust (via the caudate) (Baumgartner et al., 2008). Furthermore, meta-analysis evidence exists showing consistent activation in the vSTR (mesolimbic pathway) and dSTR (nigrostriatal pathway) for during multi-round TG interactions (i.e., identification-based trust) (Bellucci et al., 2017). Moreover, higher degrees of interpersonal trust are associated with lower levels of cortisol elevation to social stress suggesting that trust might function as a social buffering mechanism (Takahashi et al., 2005), and plasma cortisol levels are associated with trust in patients with juvenile myoclonic epilepsy (Javor et al., 2020).

Despite trust behavior and attachment involving oxytocin and dopamine pathway systems, a direct comparison between these two lines of research has been lacking, and no empirical evidence exists to date to comprehend the neural basis of trust behavior in light of attachment. Trust is a core component in an attachment relationship (i.e., family or lovers), i.e., the formation of attachment involves the development of interdependent trust (Campbell & Stanton, 2019). The neurobiology model of human attachment may serve as a useful perspective in understanding the neural basis of identification-based trust build from strangers.

1.3 The present study – A request for meta-analytic neuroimaging evidence

In this study, we conducted a *coordinate-based meta-analysis* of fMRI studies on trust and attachment —using the *activation likelihood estimation (ALE) algorithm* (Eickhoff et al., 2009)— to identify brain regions that are commonly activated by trust and attachment. We performed the meta-analysis due to the limitations associated with single fMRI studies (e.g., small samples, variations in experimental paradigm, and flexibility in analysis methods) and the absence of neural studies combined trust behavior and attachment measurement.

First, we investigated consistent brain activations associated with trust measured with the TG, including one-round TGs quantifying calculus-based trust and multi-round TGs identification-based trust. Next, we explored how these neural patterns overlap with those elicited by attachment measured with responses of mothers and romantic partners while processing attachment figures-derived stimuli. Based on the existing meta-analytic literature, we hypothesized shared activation in the dopaminergic mesolimbic (vSTR, reward/reinforcement) and the nigrostriatal (dSTR, habit formation) pathway systems between identification-based trust (Bellucci et al., 2017) and attachment (Rigo et al., 2019; Rocchetti et al., 2014). By highlighting the recruitment of neural systems supporting attachment in identification-based trust, the current study attempts to understand trust dynamics within an attachment framework.

2. Materials and methods

2.1 Literature search and selection

A systematic online database search on PubMed, ISI Web of Science, and Google Scholar (until the end of July 2020) was conducted to identify candidate articles for our coordinate-based meta-analysis using different combinations of relevant keywords: “trust,” “trust game,” “investment game,” “trustor,” “investor,” “fMRI,” “neuroimaging,” “attachment,” “affiliation,” “maternal love,” “romantic love,” “bonding,” and “beloved.” We have looked at a variety of other outlets, including the BrainMap database (http:/brainmap.org), studies referenced in review papers, and direct searches of frequently occurring authors' names.

2.2 Inclusion and exclusion criteria

The following inclusion criteria had to be met to be included as an article for the present meta-analysis: (1) Articles involving healthy participants free from psychological and neurological diagnosis or drug consumption (e.g., neuropharmaceuticals). Articles involving patients were also chosen if findings from a control group were reported separately. (2) Articles incorporating one- or multi-round TG versions, in which participants played the role of the trustor. (3) Articles with experimental contrasts involving a cooperative partner or positive feedback (i.e., reciprocated trust) in the multi-round TG, while contrasts reporting uncooperative partner or negative feedback (i.e., trusting behavior had been taken advantage of) were excluded. (4) Articles reporting whole-brain fMRI analyses, while articles describing results derived from ROIs or small-volume correction analyses were excluded to fulfill assumptions of the underlying ALE algorithm. (5) Articles for which fMRI results were obtained based on binary contrast or parametric analysis from a general linear model. (6) Articles with brain activations reported in a standardized stereotaxic space (i.e., Talairach, Montreal Neurological Institute, MNI). For the current meta-analysis, MNI space was taken as a reference structured space; coordinates recorded in Talairach space were translated to MNI space using icbm2tal supplied with GingerALE software (http://www.brainmap.org/ale/).

2.3 Paper and experimental contrast selection

Our systematic online literature research yielded 29 articles employing the TG (**Fig. 1A**) and 20 articles (**Fig. 1B**) examining attachment that met our inclusion requirements. For the TG, nine articles were one-round, and 20 were multi-round, 28 reported activations for the trust stage, nine for the feedback stage, and eight for both stages. For the attachment, twelve articles studied maternal love, and eight romantic love. The final dataset included 90 experimental contrasts and 668 foci collected from 2,419 participants (**Tab. 1**). Of these 90 contrasts, 48 were reported for the trust stage, among which 19 experimental contrasts for the trust stage in the one-round TG (77 foci across 545 subjects) and 29 in the multi-round TG (140 foci across 815 subjects), 20 for the feedback stage in the multi-round RG (108 foci across 623 subjects), and 22 for attachment (324 foci across 436 subjects, 13 for maternal love and 9 for romantic love).

2.4 ALE algorithm and main effect analyses

In-house MATLAB scripts implementing the ALE algorithm for coordinate-based meta-analyses of neuroimaging findings were used for the current meta-analysis (Eickhoff et al., 2009, 2012). To characterize across-study consistent activations for trust in both TG versions, feedback learning in the multi-round TG, and attachment, ALE maps for main effects were computed separately. These ALE maps were generated as modeled activation maps by computing the union of activation probabilities for each voxel. Voxelwise ALE scores describing the convergence of results at each unique location in the brain were produced by this algorithm. ALE scores were evaluated against a null-distribution of random spatial association between studies, allowing for random effects inference, to distinguish "true" convergence from "noise". The P values of the “true” ALE corresponded to the proportion of equal or higher values obtained under the null-distribution. The ALE maps were then thresholded at a cluster-level family-wise error (FWE) corrected threshold of $P<0.05$ with a conservative cluster forming threshold of $P<0.001$ using 10,000 permutations (Eickhoff et al., 2012).

Also, given the low number of experiments in our meta-analysis (particularly for the one-round TG trust stage), we considered only clusters as significant if: (1) their contributions came from at least two articles to prevent our findings from being driven only by a single study and (2) the most dominant experiment (MDE) contributed to the significant cluster on average less than 50% and the two MDEs (2MDEs) on average less than 80% to meet criteria of robust, unbiased results as suggested by a prior simulation study (Eickhoff et al., 2016). The fraction of the ALE value accounted for by each experiment contributing to the cluster was computed to determine the contributions of the experiments. This average non-linear experimental contribution to the ALE value was determined from the ratio of the ALE values with and without the experiment at the cluster site (Eickhoff et al., 2016). No single experiment contributed more than 50% to any significant cluster (maximum contribution of the MDE=32.95%), and the sum of the contribution of the 2MDEs was overall under 80% (maximum contribution of the 2MDEs=62.72%) (**Tab. 2**).

2.5 Conjunction analyses

Conjunction analyses were also carried out to reveal convergent activations between the thresholded ALE maps of each TG stage and attachment by computing minimum conjunction with an FWE cluster-level corrected threshold of $P<0.05$ and a cluster forming threshold of $P<0.001$ (Nichols et al., 2005). The meta-analysis results were overlaid onto a normalized anatomical ch2better.nii template provided with the MRIcron software (https://people.cas.sc.edu/rorden/mricron/index.html). The assignment of anatomical labels was based on the SPM Anatomy toolbox (www.fz-juelich.de/ime/spm_anatomy_ toolbox, v.2.2b) and the MRIcron software for Brodmann areas.

3. Results

3.1 ALE main effects

Trust stage. To investigate consistent activation maxima for calculus-based trust and identification-based trust, analyses of the across-studies main effect for the trust stage of the one-round and multi-round TG were performed, respectively. The meta-analysis for calculus-based trust revealed consistent activations in right dorsal AI (dAI) (5 contributing contrasts, i.e., 26.3% of total experiments, MDE=29.24%, 2MDEs=54.17 %) (**Fig. 2A, Tab. 2-3**), while the meta-analysis for identification-based trust showed consistent activations in the left vSTR (6 contributing contrasts, i.e., 20.0% of total experiments, MDE=29.18%, 2MDEs=54.96%) (**Fig. 2B, Tab. 2-3**).

Feedback stage. To identify brain regions consistently activated when participants learn the outcome, the main effect across studies of the positive feedback of the multi-round TG was explored. Consistent activation maxima were observed in the left putamen (8 contributing contrasts, i.e., 40% of total experiments, MDE=15.07%, 2MDEs=29.18%), and right caudate (including 153 voxels of the caudate head and 11 voxels of the caudate body; 14 contributing contrasts, i.e., 70% of total experiments, MDE=11.69%, 2MDEs=20.04%) (**Fig. 2C, Tab. 2-3**).

Attachment. To identify brain regions consistently activated when participants were exposed to attachment figure stimuli (e.g., name or picture of one's own child or romantic partner), the main effect across studies of maternal love and romantic love was explored. Consistent activation maxima were observed in the left vmPFC (4 contributing contrasts for maternal love and 2 for romantic love, i.e., 27.27% of total experiments, MDE =29.77 %, 2MDEs = 45.85 %), and the right caudate (including 131 voxels of the caudate head and 138 voxels of the caudate body; 6 contributing contrasts of maternal love and 4 contributing contrasts of romantic love, i.e., 45.45% of total experiments, MDE =17.58%, 2MDEs = 35.06%) (**Fig. 2D, Tab. 2-3**). Note that there was not enough power to contrast maternal with romantic love due to an insufficient number of experiments.

3.2 ALE conjunction analyses

Finally, to test for convergent activations between TG stages and attachment, three different conjunction analyses were performed: between attachment and (1) the trust stage in the one-round TG (calculus-based trust); (2) the trust stage in the multi-round TG (identification-based trust); and (3) the feedback stage in the multi-round TG. The conjunction analyses revealed an overlap between the feedback stage and attachment in the right caudate (including 76 voxels of the caudate head and 11 voxels of the caudate body) (**Fig. 2E, Tab. 3**). No suprathreshold clusters were found for the other two analyses.

4. Discussion

Trust is crucial for establishing relationships between strangers to sustain a secure and flourish society. Does trust-building among strangers through repeated interactions share common neural underpinnings with attachment? Here, we investigated whether identification-based trust (measured by trust in the multi-round TG) but not calculus-based trust (measured by trust in the one-round TG) engages brain mechanisms associated with attachment. Applying a coordinate-based meta-analysis approach that implements the ALE method, we identified consistently activated brain regions for calculus-based trust (dAI, salience network), identification-based trust (vSTR, mesolimbic pathway), feedback in the multi-round TG (dSTR, nigrostriatal pathway), and attachment (dSTR (nigrostriatal pathway) and vmPFC (mesocortical pathway)). Further, we demonstrated that the positive feedback in the multi-round TG and attachment shared a partly overlapping brain region being part of the nigrostriatal pathway (i.e., dSTR, caudate head).

Our results suggest that strangers form dyadic trust through repeated interactions in the TG involving the mesolimbic (vSTR) and nigrostriatal (dSTR) pathways, which are also associated with attachment (Strathearn et al., 2019). Higher levels of dyadic trust suggest assurance that a partner will act pro-relationally in the future (Campbell & Stanton, 2019). In perceptions of partners' trustworthiness (i.e., positive feedback), players form beliefs regarding the future dependability of the relationship and seek a social approach toward them (i.e., nigrostriatal pathway), and feel secure when trusting them (i.e., mesolimbic pathway). The overlapping in the nigrostriatal pathway between positive feedback in the multi-round TG and attachment suggests the formation of a trust relationship resembled the responses of secure attachment.

4.1 The psychological processes of brain regions underlying trust

The NPE model proposes the underlying neuropsychological mechanism on how interpersonal trust evolves through repeated interactions (Krueger & Meyer-Lindenberg, 2019): from calculus-based trust, over knowledge-based trust, to identification-based trust (Lewicki & Bunker, 1995). A trust relationship begins with calculus-based trust, driven primarily by the salience network (risk of treachery). A shift from the salience network (dAI) to the reward network (vSTR) can be observed when transitioning from one-round to multi-round TG interactions —probably reflecting a shift from calculus-based trust (guided by uncertainty about the risk of treachery) to identification-based trust (guided by certainty about the trustworthiness of the trustee). Through repeated interactions, the trust relationship matures to identification-based trust, driven predominantly by the reward network (anticipation of the reward). As a result, trustors develop a rewarding identification with trustees and trust them confidently.

First, in our ALE analysis, trust decisions in the one-round TG (as a measure of calculus-based trust) consistently activated the right dAI as a key region of the salience network —confirming the results of previous meta-analysis studies (Bellucci et al., 2017, 2018). In the one-round TG, the dAI may encode the risk of betrayal for trusting an unknown trustee. Previous parcellation studies suggest that the AI can be subdivided into a dAI region (associated with a cognitive network) and a ventral AI (vAI) region (associated with an affective network) (Chang et al., 2013; Kelly et al., 2012). A recent study combining coordinate-based fMRI meta-analyses with task-based and task-free connectivity analyses showed that the right AI was the common brain region consistently activated in both the TG and the ultimatum game (UG) (Bellucci et al., 2018). Activations resulting from trust decisions and rejection of unfair offers clustered in the right dAI, while activations resulting from reciprocating decisions and rejection of unfair offers clustered in the right vAl. These findings suggest that the dAI mediates cognitive processes associated with an expectation of conformity with social norms. In contrast, the vAI mediates affective processes associated with aversion to the injustice that leads to enforcement of social norms (Krueger et al., 2020).

Second, trust decisions in the multi-round TG (as a measure of identification-based trust) consistently activated the left vSTR, supporting the results of a previous meta-analysis study with our current bigger sample size (Bellucci et al., 2017). As a key region of the mesolimbic dopamine pathway, the vSTR is reliably associated with reward anticipation (Knutson & Cooper, 2005) and net benefit encoding of a choice option (Wittmann et al., 2010). In the multi-round TG, the vSTR may be engaged in predictions about whether a future reward is likely to follow a trust decision. Through repeated interactions with rewarded trust —increasing the certainty about the trustees' trustworthiness— trustors have more confidence regarding the trustees' identity and are more inclined to trust them. When trustors made trust decisions under these circumstances, they are confident of receiving the reward being trusted —resulting in a highly predictable reward value as signaled by the vSTR (Schultz et al., 1997).

Finally, positive feedback (i.e., rewarded trust) in the multi-round TG consistently activated the dSTR (including the putamen and caudate head). The dSTR is a fundamental region of the nigrostriatal dopamine pathway —associated with reinforcement learning of action-outcome associations to regulate and optimize future actions leading to a reward (Delgado, 2007; Wise, 2009). The dSTR is further comprised of the subparts of the dorsolateral striatum (dlSTR, the putamen) and the dorsomedial striatum (dmSTR, the caudate). The dlSTR may encode stimulus-response associations that are learned incrementally through trial and error, while the dmSTR may perform computations of action values based on action-outcome representations and guide goal-directed action selection controlled by outcome expectations (Takahashi, 2008). Positive feedback in the multi-round TG activated both the dlSTR and the dmSTR, suggesting an involvement of both learning dynamics and action guidance. For the multi-round TG, the willingness to trust others is contingent upon the feedback on the consequences of previous choices. Through repeated interactions with rewarded trust, trustors become more confident about the trustworthiness of the trustee (learning signaled by the dlSTR) and more likely to trust that person in the future, and the dmSTR guides subsequent trust decisions that facilitate the social affiliation of the trust-trustee relationship.

4.2 The neural correlates of attachment

The neurobiology model of human attachment suggests the involvement of three pathway systems: oxytocinergic “affiliation” pathway system, dopamineergic mesocorticolimbic “reward/reinforcement’ and nigrostriatal “habit formation” pathways system, and the glucocorticoid “stress-response” system (Strathearn et al., 2019). The oxytocinergic “affiliation” pathway system is essential to the development of social and spatial memories, affiliation behavior, and emotional control (Baskerville & Douglas, 2010). Both animal and human research support the role of oxytocin in promoting social approach behavior and overcoming avoidance of proximity (Meyer-Lindenberg et al., 2011). The dopaminergic system —comprising the mesolimbic, mesocortical, and nigrostriatal pathways— is involved in stimulus-reward reinforcement learning and decision-making based on future predicted reward (Ferguson et al., 2002). Research evidenced that in contrast to insecure and rejecting mothers, mothers with secure attachment patterns demonstrated increased activation of mesocorticolimbic dopamine regions (including vSTR and vmPFC) and the oxytocin-associated hypothalamic/pituitary regions (Strathearn, 2011). The glucocorticoid stress-response system is critical both for human attachment (Strathearn et al., 2019) and trust behavior (Riedl & Javor, 2012). However, both previous and current meta-analyses failed to find consistent activation in the hypothalamus-pituitary regions. This may be due to the fact that the examined studies did not involve a stress context, therefore, stress-related responses could not be observed. More research is needed to clarify the role of cortisol in both trust and attachment neuroscience.

Consistent with the neurobiological model of human attachment (Strathearn et al., 2019), the current meta-analysis revealed that attachment figures consistently activated the left vmPFC (mesocortical pathway) and the right dmSTR (nigrostriatal pathway). The mesocortical pathway is engaged in adaptive behaviors and survival-oriented responses (Ivlieva, 2011) and supports the assumption that human tends to seek and maintain proximity to caring and supporting others to increase the likelihood of survival (Bowlby, 1982/1969). The activation in the vmPFC may signal a safety value (Eisenberger et al., 2011), entailing the function of maintaining an approach-oriented motivation to foster closeness with others (Long et al., 2020). The nigrostriatal pathway is considered essential for the formation of stimulus-response associations that underlie skilled movements and habitual actions and guide goal-directed behaviors that rely on action–outcome associations (Liljeholm & O'Doherty, 2012; Redgrave et al., 2010). In human attachment, the nigrostriatal pathway plays a crucial role in habit formation, which supports a strong approach motivation to attachment figures and formation of a secure attachment (Strathearn et al., 2019). The consistent activation in the dmSTR by attachment supports the role of this region in facilitating goal-directed social approach behavior. The lack of consistent activation in the dlSTR may be due to the fact that the paradigms used to examine attachment did not involve learning dynamics but merely elicit attachment responses by introducing stimuli relevant to the attachment figure.

Compared to the previous meta-analyses on maternal attachment, our current study combining maternal and romantic attachment failed to observe consistent activations in other areas such as globus pallidus, insula, thalamus, amygdala, and inferior frontal gyrus (Rigo et al., 2019; Rocchetti et al., 2014). This may be due to the implementation of different meta-analytic approaches such as the effect-size version of signed differential mapping applied in Rocchetti et al. (2014), the multi-level density analysis tool in Rigo et al. (2019), and the ALE method in our current study. To control for the multiple comparisons problem in our study, the ALE maps were thresholded at a cluster level FEW corrected threshold of $P<0.05$ with a conservative cluster-forming threshold of $P<0.001$ implementing 10,000 permutations (Eickhoff et al., 2012), which resulted in restricted and smaller foci because this approach has a lower false positives rate and therefore is more conservative than other meta-analytic neuroimaging methods (Radua et al., 2012).

Importantly, neither the previous meta-analyses of human attachment (Ortigue et al., 2010; Rigo et al., 2019; Rocchetti et al., 2014) nor our current one revealed a consistent activation of brain regions involving the mesolimbic dopamine pathway (e.g., vSTR) which has been implicated in other aspects of reward learning for attachment in humans (Strathearn et al., 2019) and for social bond formation in animals (Insel, 2003; Johnson & Young, 2015). One reason for such a lack might be the experimental paradigms (e.g., viewing facial images or videos of attachment figures) employed in most fMRI attachment studies, which only evoke the representation of attachment figures for a short duration and hence might fail to catch the attachment establishment process. Furthermore, the majority of the studies we examined implemented a first-person perspective in which participants typically observe individuals in isolation —i.e., viewing attachment figures when they are alone and not engaging in reciprocal interactions with them. These paradigm limitations may also account for the absence of activations in the oxytocinergic “affiliation” and glucocorticoid stress-response pathways. The examination of behavioral, psychological, and brain processes in at least two persons who communicate directly with each other will be a more ecologically valid approach to evaluating the biological and brain basis for attachment (Long et al., 2020). Moreover, the attachment figure engages in behaviors in a more ecological and interactive way may be more likely to directly activate the dopaminergic, oxytocinergic, and glucocorticoid systems. For instance, warm contacts (e.g., hugs) with attachment figure are associated with higher oxytocin levels (Light et al., 2005), and affiliative behaviors, particularly those involving touch, serve as stress-buffer engaging brain regions with dense oxytocin (i.e., paraventricular nucleus) and opioid neurons (i.e., arcuate nucleus) (Morrison, 2016).

4.3 The commonalities between trust and attachment

Both positive feedback in the multi-round TG and attachment engaged the nigrostriatal dopamine pathway (i.e., dmSTR) that is associated with habit formation (Strathearn et al., 2019). While during the early stages of life, the biological role of the behavioral attachment system is most important, Bowlby (1988) believed that the system is active over the entire life span and is expressed in thoughts and behaviors related to seeking proximity beyond attachment figures to appropriate others. It is reasonable to assume that part of the neural substrates underlying the establishment of human attachment during early life also performs a similar function of forming a social affiliation with others (even strangers) in later life. For the multi-round TG interactions, if trustees always reciprocate the trustors' trust, then they provide security to the trustor and reduce uncertainty. Hence, trustors can feel confident and safe that their trust will be rewarded and thus form a "trust habit" with the trustee, which are processes functionalized by the mesocorticolimbic and nigrostriatal dopamine pathways.

Another similarity between trust and human attachment is the involvement of the oxytocinergic "affiliation" pathway system, though no consistent activation has been found in previous and present meta-analysis results. Oxytocin is implicated in human mothering, fathering, co-parenting, close friendship, and romantic attachment (Love, 2014) and encourages social approach behavior (Meyer-Lindenberg et al., 2011). Integration of oxytocin and dopamine in striatum ignites bonding, instilling attachments with motivation and vigor (Feldman, 2017; Love, 2014). Oxytocin also promotes interpersonal trust, which could be taken as an indicator of social approach in humans —by inhibiting defensive behaviors and associating this inhibition with the activation of dopaminergic reinforcement circuits to increase the value of social experiences (Campbell, 2008). For example, our willingness to trust strangers can be potentiated by the exogenous administration of oxytocin (Kosfeld et al., 2005), though this effect may be context- (Ide et al., 2018) or individual-dependent (e.g., effective in individuals with a lower disposition to trust, see Declerck et al., 2020) (see for a critical review, Nave et al., 2015).

Furthermore, participants undergoing nasal administration of oxytocin showed no change in their trusting behavior after learning that their trust had been multiple times breached while participants receiving placebo decreased their trust (Baumgartner et al., 2008). This disparity in trust adaptation was associated with a decrease in the activation of the amygdala, the midbrain regions, and the dmSTR in participants receiving oxytocin compared to placebo —indicating that neural systems mediating fear processing (amygdala and midbrain regions) and behavioral adaptations to feedback information (dmSTR, nigrostriatal "habit formation" pathway) modulate the effect of oxytocin on trust. Previously, Xu and colleagues (2019) suggested that the main function of oxytocin is not to promote mutual trust per se but to adapt and learn from trustworthy individuals who are either in-group members and/or perceived experts. Together, the oxytocin and dopamine pathways permit social recognition, trust, and a range of socially affiliative behaviors (Skuse & Gallagher, 2009).

4.4 Limitations

The current meta-analysis has some limitations. First, due to a lack of fMRI studies, knowledge-based trust could not be analyzed in the current coordinate-based meta-analysis. To confirm the assumptions about the neural signatures of trust dynamics —ranging from calculus-, over knowledge-, to identification-based trust— future neuroimaging studies employing targeted experimental manipulations (e.g., providing contextual information) are needed to investigate knowledge-based trust (Fouragnan et al., 2013). Second, our meta-analytic findings are based dominantly on neural correlates of human social behavior in single/isolated individuals, often referred to as "first-person social neuroscience" (Long et al., 2020). Since both trust and attachment involve two-person interactions, a "second-person social neuroscience" approach (Schilbach et al., 2013) is desirable, in which a special focus is directed towards measuring brain activation in two interacting individuals and deriving a measure of inter-brain coherence implementing fMRI hyperscanning (Babiloni & Astolfi, 2014; Czeszumski et al., 2020; Koike et al., 2015; Montague, 2002). Finally, this study identified an overlapping brain region involved in trust and attachment; however, the direct relationship and interactions between these two processes remain to be illuminated. Future fMRI studies applying the TG with attachment assessments (Collins & Read, 1990; Simpson et al., 1996) are necessary to investigate the impact of attachment styles on trust behavior, which would provide a more comprehensive picture of human attachment. For example, behavioral studies demonstrated that for individuals high rather than low in attachment avoidance, oxytocin compared to placebo reduced betrayal aversion, and increased trust and cooperation (De Dreu, 2012), but for participants high in attachment anxiety and intimacy-avoidance, oxytocin impeded trust behavior (Bartz et al., 2011). Despite those limitations, our results demonstrated for the first time that identification-based trust and attachment share partially overlapping neural systems.

4.5 Conclusions

Taken together, our meta-analytic results revealed that calculus-based trust (identified with the one-round TG) and identification-based trust (identified with the multi-round TG) engaged distinct neural networks with dissociate functional roles. Calculus-based trust involved the right dAI as part of the salience network, presumably signaling the risk of treachery upon trust decisions, while identification-based trust encompassed the vSTR as part of the mesolimbic dopaminergic pathway, probably signaling the expected reward of being reciprocated. Attachment engaged the vmPFC as part of the mesocortical dopaminergic pathway, likely signaling the safety value associated with attachment figures and the dmSTR as part of the nigrostriatal dopaminergic pathway, probably facilitating social approach behaviors. An overlap between activation during positive feedback in the multi-round TG and attachment in the dmSTR as part of the nigrostriatal pathway was identified, possibly reinforcing the action-outcome associations and promoting social approach behavior —but no overlap in the mesolimbic and mesocortical pathways was observed. Our results should be validated by future studies directly investigating the neural responses of attachment and interpersonal trust. The next wave of studies on attachment and trust dynamics within social relationships (comprising not only strangers but also friends, lovers, and family) will likely yield noteworthy findings in how attachment and trust interact and change throughout life. By highlighting the commonalities of the psychoneurobiological basis between trust and attachment, our findings shed light on how trust established among strangers very much resembles crucial components for the formation of attachment. Our results also have clinical applications, especially for people with insecure/dismiss attachment styles who experienced adverse childhood events (e.g., maternal neglect or traumatic abuse) or for adolescents who go through a phase of attachment crisis characterized by the search for new attachment figures and who also show differing reward processing that has been associated, for instance, with enhanced risk-taking behaviors (Asscheman et al., 2020). Integrating the research of trust and attachment will not only advances our knowledge on the neurobiological underpinnings of social affiliation but also helps to develop prevention and intervention strategies for people at such risk.

Funding: FK was supported by a MINERVA grant (FA9550-18-1–0455). YW was supported by the National Natural Science Foundation of China (grant number 31471073).

Conflict of Interest: The authors are unaware of any conflicts of interest, financial or otherwise.

References

Ainsworth, M. S. (1989). Attachments beyond infancy. *American Psychologist*, *44*(4), 709.

Albers, E. M., Marianne Riksen-Walraven, J., Sweep, F. C. G. J., & Weerth, C. de. (2008). Maternal behavior predicts infant cortisol recovery from a mild everyday stressor. *Journal of Child Psychology and Psychiatry*, *49*(1), 97–103. https://doi.org/10.1111/j.1469-7610.2007.01818.x

Antonucci, L. A., Taurisano, P., Coppola, G., & Cassibba, R. (2018). Attachment style: The neurobiological substrate, interaction with genetics and role in neurodevelopmental disorders risk pathways. *Neuroscience & Biobehavioral Reviews*, *95*, 515–527. https://doi.org/10.1016/j.neubiorev.2018.11.002

Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., & Pöppel, E. (2010). Dopaminergic reward system: A short integrative review. *International Archives of Medicine*, *3*(1), 24. https://doi.org/10.1186/1755-7682-3-24

Aron, A., Fisher, H., Mashek, D. J., Strong, G., Li, H., & Brown, L. L. (2005). Reward, motivation, and emotion systems associated with early-stage intense romantic love. *Journal of Neurophysiology*, *94*(1), 327–337. https://doi.org/10.1152/jn.00838.2004

Asscheman, J. S., Deater-Deckard, K., Lauharatanahirun, N., van Lier, P. A. C., Koot, S., King-Casas, B., & Kim-Spoon, J. (2020). Associations between peer attachment and neural correlates of risk processing across adolescence. *Developmental Cognitive Neuroscience*, *42*, 100772–100772. PubMed. https://doi.org/10.1016/j.dcn.2020.100772

Atzil, S., Hendler, T., & Feldman, R. (2011). Specifying the neurobiological basis of human attachment: Brain, hormones, and bBehavior in synchronous and intrusive mothers. *Neuropsychopharmacology*, *36*(13), 2603–2615. https://doi.org/10.1038/npp.2011.172

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience & Biobehavioral Reviews*, *44*, 76–93. https://doi.org/10.1016/j.neubiorev.2012.07.006

Balliet, D., & Van Lange, P. A. M. (2013). Trust, punishment, and cooperation across 18 societies: A meta-analysis. *Perspectives on Psychological Science*, *8*(4), 363–379. https://doi.org/10.1177/1745691613488533

Barrett, J., Wonch, K. E., Gonzalez, A., Ali, N., Steiner, M., Hall, G. B., & Fleming, A. S. (2012). Maternal affect and quality of parenting experiences are related to amygdala response to infant faces. *Social Neuroscience*, *7*(3), 252–268. https://doi.org/10.1080/17470919.2011.609907

Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *NeuroReport*, *11*(17). https://journals.lww.com/neuroreport/Fulltext/2000/11270/The_neural_basis_of_romantic_love.46.aspx

Bartz, J., Simeon, D., Hamilton, H., Kim, S., Crystal, S., Braun, A., Vicens, V., & Hollander, E. (2011). Oxytocin can hinder trust and cooperation in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *6*(5), 556–563. https://doi.org/10.1093/scan/nsq085

Baskerville, T. A., & Douglas, A. J. (2010). Dopamine and oxytocin interactions underlying behaviors: Potential contributions to behavioral disorder. *CNS Neuroscience & Therapeutics*, *16*(3), e92–e123. https://doi.org/10.1111/j.1755-5949.2010.00154.x

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*(4), 639–650. https://doi.org/10.1016/j.neuron.2008.04.009

Bellucci, G., Chernyak, S. V., Goodyear, K., Eickhoff, S. B., & Krueger, F. (2017). Neural signatures of trust in reciprocity: A coordinate-based meta-analysis: Neural Signatures of Trust in Reciprocity. *Human Brain Mapping*, *38*(3), 1233–1248. https://doi.org/10.1002/hbm.23451

Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S. B., & Krueger, F. (2018). The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neuroscience & Biobehavioral Reviews*, *92*, 378–389. https://doi.org/10.1016/j.neubiorev.2018.06.024

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social History. *Games and Economic Behavior*, *10*(1), 122–142. https://doi.org/10.1006/game.1995.1027

Bosmans, G., Bakermans-Kranenburg, M. J., Vervliet, B., Verhees, M. W. F. T., & van IJzendoorn, M. H. (2020). A learning theory of attachment: Unraveling the black box of attachment development. *Neuroscience & Biobehavioral Reviews*, *113*, 287–298. https://doi.org/10.1016/j.neubiorev.2020.03.014

Bourdy, R., Sánchez-Catalán, M.-J., Kaufling, J., Balcita-Pedicino, J. J., Freund-Mercier, M.-J., Veinante, P., Sesack, S. R., Georges, F., & Barrot, M. (2014). Control of the nigrostriatal dopamine neuron activity and motor function by the tail of the ventral tegmental area. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *39*(12), 2788–2798. PubMed.

https://doi.org/10.1038/npp.2014.129

Bowlby, J. (1973). Attachment and loss: Volume II: Separation, anxiety and anger. In *Attachment and Loss: Volume II: Separation, Anxiety and Anger* (pp. 1–429). London: The Hogarth press and the institute of psycho-analysis.

Bowlby, J. (1979). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences*, *2*(4), 637–638. https://doi.org/10.1017/S0140525X00064955

Bowlby, J. (1988). *A secure base: Parent-child attachment and healthy human development.* (pp. xii, 205). Basic Books.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Camerer, C. F. (2003). PSYCHOLOGY AND ECONOMICS: Enhanced: Strategizing in the brain. *Science*, *300*(5626), 1673–1675. https://doi.org/10.1126/science.1086215

Campbell, A. (2008). Attachment, aggression and affiliation: The role of oxytocin in female social behavior. *Biological Psychology*, *77*(1), 1–10. https://doi.org/10.1016/j.biopsycho.2007.09.001

Campbell, L., & Stanton, S. C. (2019). Adult attachment and trust in romantic relationships. *Current Opinion in Psychology*, *25*, 148–151. https://doi.org/10.1016/j.copsyc.2018.08.004

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex*, *23*(3), 739–749. https://doi.org/10.1093/cercor/bhs065

Cochard, F., Nguyen Van, P., & Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior & Organization*, *55*(1), 31–44. https://doi.org/10.1016/j.jebo.2003.07.004

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models, and relationship quality in dating couples. *Journal of Personality and Social Psychology*, *58*(4), 644–663. https://doi.org/10.1037/0022-3514.58.4.644

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, *14*, 39. https://doi.org/10.3389/fnhum.2020.00039

De Dreu, C. K. W. (2012). Oxytocin modulates the link between adult attachment and cooperation through reduced betrayal aversion. *Psychoneuroendocrinology*, *37*(7), 871–880. https://doi.org/10.1016/j.psyneuen.2011.10.003

De Dreu, C. K. W., & Kret, M. E. (2016). Oxytocin conditions intergroup relations through upregulated in-group empathy, cooperation, conformity, and defense. *Biological Psychiatry*, *79*(3), 165–173. https://doi.org/10.1016/j.biopsych.2015.03.020

Declerck, C. H., Boone, C., Pauwels, L., Vogt, B., & Fehr, E. (2020). A registered replication study on oxytocin and trust. *Nature Human Behaviour*, *4*(6), 646–655. https://doi.org/10.1038/s41562-020-0878-x

Delgado, M. R. (2007). Reward-Related Responses in the Human Striatum. *Annals of the New York Academy of Sciences*, *1104*(1), 70–88. https://doi.org/10.1196/annals.1390.002

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*(11), 1611–1618. https://doi.org/10.1038/nn1575

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. https://doi.org/10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. https://doi.org/10.1002/hbm.20718

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70–85. https://doi.org/10.1016/j.neuroimage.2016.04.072

Eisenberger, N. I., Master, S. L., Inagaki, T. K., Taylor, S. E., Shirinyan, D., Lieberman, M. D., & Naliboff, B. D. (2011). Attachment figures activate a safety signal-related neural region and reduce pain experience. *Proceedings of the National Academy of Sciences*, *108*(28), 11721–11726. https://doi.org/10.1073/pnas.1108239108

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*. https://doi.org/10.3389/fnins.2012.00148

Fehr, E., Fischbacher, U., & Kosfeld, M. (2005). Neuroeconomic foundations of trust and social preferences: Initial evidence. *American Economic Review*, *95*(2), 346–351. https://doi.org/10.1257/000282805774669736

Feldman, R. (2017). The neurobiology of human attachments. *Trends in Cognitive Sciences*, *21*(2), 80–99. https://doi.org/10.1016/j.tics.2016.11.007

Feldman, R., Monakhov, M., Pratt, M., & Ebstein, R. P. (2016). Oxytocin pathway genes: Evolutionary ancient system impacting on human affiliation, sociality, and psychopathology. *Biological Psychiatry*, *79*(3), 174–184. https://doi.org/10.1016/j.biopsych.2015.08.008

Ferguson, J. N., Young, L. J., & Insel, T. R. (2002). The neuroendocrine basis of social recognition. *Frontiers in Neuroendocrinology*, *23*(2), 200–224. https://doi.org/10.1006/frne.2002.0229

Fouragnan, E., Chierchia, G., Greiner, S., Neveu, R., Avesani, P., & Coricelli, G. (2013). Reputational priors magnify striatal responses to violations of trust. *Journal of Neuroscience*, *33*(8), 3602–3611. https://doi.org/10.1523/JNEUROSCI.3086-12.2013

Hoekzema, E., Barba-Müller, E., Pozzobon, C., Picado, M., Lucco, F., García-García, D., Soliva, J. C., Tobeña, A., Desco, M., Crone, E. A., Ballesteros, A., Carmona, S., & Vilarroya, O. (2017). Pregnancy leads to long-lasting changes in human brain structure. *Nature Neuroscience*, *20*(2), 287–296. https://doi.org/10.1038/nn.4458

Howard, C. D., Li, H., Geddes, C. E., & Jin, X. (2017). Dynamic nigrostriatal dopamine biases Action Selection. *Neuron*, *93*(6), 1436-1450.e8. https://doi.org/10.1016/j.neuron.2017.02.029

Ide, J. S., Nedic, S., Wong, K. F., Strey, S. L., Lawson, E. A., Dickerson, B. C., Wald, L. L., La Camera, G., & Mujica-Parodi, L. R. (2018). Oxytocin attenuates trust as a subset of more general reinforcement learning, with altered reward circuit functional connectivity in males. *NeuroImage*, *174*, 35–43. https://doi.org/10.1016/j.neuroimage.2018.02.035

Insel, T. R. (2003). Is social attachment an addictive disorder? *Physiology & Behavior*, *79*(3), 351–357. https://doi.org/10.1016/S0031-9384(03)00148-3

Ivlieva, N. Yu. (2011). Involvement of the mesocorticolimbic dopaminergic system in adaptive behavior. *Neuroscience and Behavioral Physiology*, *41*(7), 715–729. https://doi.org/10.1007/s11055-011-9477-7

Javor, A., Zamarian, L., Ransmayr, G., Prieschl, M., Bergmann, M., Walser, G., Luef, G., Prokop, W., Delazer, M., & Unterberger, I. (2020). The role of cortisol in trust behavior: Results from an experimental study on healthy controls and patients with juvenile myoclonic epilepsy. *Epilepsy & Behavior*, *110*, 107138. https://doi.org/10.1016/j.yebeh.2020.107138

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Johnson, Z. V., & Young, L. J. (2015). Neurobiological mechanisms of social attachment and pair bonding. *Current Opinion in Behavioral Sciences*, *3*, 38–44. https://doi.org/10.1016/j.cobeha.2015.01.009

Kelly, C., Toro, R., Di Martino, A., Cox, C. L., Bellec, P., Castellanos, F. X., & Milham, M. P. (2012). A convergent functional architecture of the insula emerges across imaging modalities. *NeuroImage*, *61*(4), 1129–1142.

https://doi.org/10.1016/j.neuroimage.2012.03.021

King-Casas, B., Tomlin, D., Cedric Anen, Colin F. Camerer, Steven R. Quartz, & P. Read Montague. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*(5718), 78–83. https://doi.org/10.1126/science.1108062

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction: *Current Opinion in Neurology*, *18*(4), 411–417. https://doi.org/10.1097/01.wco.0000173463.24758.f6

Koike, T., Tanabe, H. C., & Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the “two-in-one” system in social interactions. *Neuroscience Research*, *90*, 25–32. https://doi.org/10.1016/j.neures.2014.11.006

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–676. https://doi.org/10.1038/nature03701

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., & Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, *104*(50), 20084–20089. https://doi.org/10.1073/pnas.0710103104

Krueger, Frank, Bellucci, G., Xu, P., & Feng, C. (2020). The critical role of the right dorsal and ventral anterior insula in reciprocity: Evidence from the trust and ultimatum games. *Frontiers in Human Neuroscience, 14,* 176. https://doi.org/10.3389/fnhum.2020.00176

Krueger, Frank, & Meyer-Lindenberg, A. (2019). Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends in Neurosciences*, *42*(2), 92–101. https://doi.org/10.1016/j.tins.2018.10.004

Kuo, P. X., Saini, E. K., Tengelitsch, E., & Volling, B. L. (2019). Is one secure attachment enough? Infant cortisol reactivity and the security of infant-mother and infant-father attachments at the end of the first year. *Attachment & Human Development*, *21*(5), 426–444. https://doi.org/10.1080/14616734.2019.1582595

Lapish, C. C., Kroener, S., Durstewitz, D., Lavin, A., & Seamans, J. K. (2007). The ability of the mesocortical dopamine system to operate in distinct temporal modes. *Psychopharmacology*, *191*(3), 609–625. PubMed. https://doi.org/10.1007/s00213-006-0527-8

Laurita, A. C., Hazan, C., & Spreng, R. N. (2019). An attachment theoretical perspective for the neural representation of close others. *Social Cognitive and Affective Neuroscience*, *14*(3), 237–251. https://doi.org/10.1093/scan/nsz010

Lavin, A. (2005). Mesocortical dopamine neurons operate in distinct temporal domains using multimodal signaling. *Journal of Neuroscience*, *25*(20), 5013–5023. https://doi.org/10.1523/JNEUROSCI.0557-05.2005

Lewicki, R. J., & Bunker, B. B. (1995). Trust in relationships: A model of development and decline. In *Conflict, cooperation, and justice: Essays inspired by the work of Morton Deutsch.* (pp. 133–173). Jossey-Bass/Wiley.

Light, K. C., Grewen, K. M., & Amico, J. A. (2005). More frequent partner hugs and higher oxytocin levels are linked to lower blood pressure and heart rate in premenopausal women. *Biological Psychology*, *69*(1), 5–21. https://doi.org/10.1016/j.biopsycho.2004.11.002

Liljeholm, M., & O’Doherty, J. P. (2012). Contributions of the striatum to learning, motivation, and performance: An associative account. *Trends in Cognitive Sciences*, *16*(9), 467–475. https://doi.org/10.1016/j.tics.2012.07.007

Long, M., Verbeke, W., Ein-Dor, T., & Vrtička, P. (2020). A functional neuro-anatomical model of human attachment (NAMA): Insights from first- and second-person social neuroscience. *Cortex*, *126*, 281–321. https://doi.org/10.1016/j.cortex.2020.01.010

Love, T. M. (2014). Oxytocin, motivation and the role of dopamine. *Pharmacology Biochemistry and Behavior, 119*, 49–60. https://doi.org/10.1016/j.pbb.2013.06.011

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

McClure, S. M., Daw, N. D., & Read Montague, P. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, *26*(8), 423–428. https://doi.org/10.1016/S0166-2236(03)00177-2

Meyer-Lindenberg, A., Domes, G., Kirsch, P., & Heinrichs, M. (2011). Oxytocin and vasopressin in the human brain: Social neuropeptides for translational medicine. *Nature Reviews Neuroscience*, *12*(9), 524–538. https://doi.org/10.1038/nrn3044

Michalska, K. J., Decety, J., Liu, C., Chen, Q., Martz, M. E., Jacob, S., Hipwell, A. E., Lee, S. S., Chronis-Tuscano, A., Waldman, I. D., & Lahey, B. B. (2014). Genetic imaging of the association of oxytocin receptor gene (OXTR) polymorphisms with positive maternal parenting. *Frontiers in Behavioral Neuroscience, 8*. https://doi.org/10.3389/fnbeh.2014.00021

Mikulincer, M., & Shaver, P. R. (2003). The attachment behavioral system in adulthood: Activation, psychodynamics, and interpersonal processes. In *Advances in Experimental Social Psychology* (Vol. 35, pp. 53–152). Elsevier. https://doi.org/10.1016/S0065-2601(03)01002-5

Montague, P. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions. *NeuroImage*, *16*(4), 1159–1164. https://doi.org/10.1006/nimg.2002.1150

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature*, *431*(7010), 760–767. https://doi.org/10.1038/nature03015

Morrison, I. (2016). Keep calm and cuddle on: Social touch as a stress buffer. *Adaptive Human Behavior and Physiology*, *2*(4), 344–362. https://doi.org/10.1007/s40750-016-0052-x

Nave, G., Camerer, C., & McCullough, M. (2015). Does oxytocin increase trust in humans? A critical review of research. *Perspectives on Psychological Science*, *10*(6), 772–789. https://doi.org/10.1177/1745691615600138

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Numan, M. (2006). Hypothalamic neural circuits regulating maternal responsiveness toward infants. *Behavioral and Cognitive Neuroscience Reviews*, *5*(4), 163–190. https://doi.org/10.1177/1534582306288790

Numan, M., & Young, L. J. (2016). Neural mechanisms of mother–infant bonding and pair bonding: Similarities, differences, and broader implications. *Hormones and Behavior*, *77*, 98–112. https://doi.org/10.1016/j.yhbeh.2015.05.015

Ortigue, S., Bianchi-Demicheli, F., de C. Hamilton, A. F., & Grafton, S. T. (2007). The neural basis of love as a subliminal prime: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *19*(7), 1218–1230. https://doi.org/10.1162/jocn.2007.19.7.1218

Ortigue, Stephanie, Bianchi-Demicheli, F., Patel, N., Frum, C., & Lewis, J. W. (2010). Neuroimaging of Love: FMRI Meta-Analysis Evidence toward New Perspectives in Sexual Medicine. *The Journal of Sexual Medicine*, *7*(11), 3541–3552. https://doi.org/10.1111/j.1743-6109.2010.01999.x

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press. https://books.google.de/books?id=qqcRGagyEuAC

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., & Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry*, *27*(8), 605–611. https://doi.org/10.1016/j.eurpsy.2011.04.001

Ran, G., & Zhang, Q. (2018). The neural correlates of attachment style during emotional processing: An activation likelihood estimation meta-analysis. *Attachment & Human Development*, *20*(6), 626–633. https://doi.org/10.1080/14616734.2018.1465105

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M. C., Lehericy, S., Bergman, H., Agid, Y., DeLong, M. R., & Obeso, J. A. (2010). Goal-directed and habitual control in the basal ganglia: Implications for Parkinson's disease. *Nature Reviews Neuroscience*, *11*(11), 760–772. https://doi.org/10.1038/nrn2915

Riedl, R., & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics*, *5*(2), 63–91. https://doi.org/10.1037/a0026318

Rigo, P., Kim, P., Esposito, G., Putnick, D. L., Venuti, P., & Bornstein, M. H. (2019). Specific maternal brain responses to their own child's face: An fMRI meta-analysis. *Developmental Review*, *51*, 58–69. https://doi.org/10.1016/j.dr.2018.12.001

Rocchetti, M., Radua, J., Paloyelis, Y., Xenaki, L., Frascarelli, M., Caverzasi, E., Politi, P., & Fusar-Poli, P. (2014). Neurofunctional maps of the 'maternal brain' and the effects of oxytocin: A multimodal voxel-based meta-analysis. *Psychiatry and Clinical Neurosciences*, *68*(10), 733–751. https://doi.org/10.1111/pcn.12185

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Simpson, J. A., Rholes, W. S., & Phillips, D. (1996). Conflict in close relationships: An attachment perspective. *Journal of Personality and Social Psychology*, *71*(5), 899–914. https://doi.org/10.1037/0022-3514.71.5.899

Skuse, D. H., & Gallagher, L. (2009). Dopaminergic-neuropeptide interactions in the social brain. *Trends in Cognitive Sciences*, *13*(1), 27–35. https://doi.org/10.1016/j.tics.2008.09.007

Strathearn, L. (2011). Maternal neglect: Oxytocin, dopamine and the neurobiology of attachment. *Journal of Neuroendocrinology*, *23*(11), 1054–1065. https://doi.org/10.1111/j.1365-2826.2011.02228.x

Strathearn, Lane, Li, J., Fonagy, P., & Montague, P. R. (2008). What's in a smile? Maternal brain responses to infant facial cues. *Pediatrics*, *122*(1), 40–51. PubMed. https://doi.org/10.1542/peds.2007-1566

Strathearn, Lane, Mertens, C. E., Mayes, L., Rutherford, H., Rajhans, P., Xu, G., Potenza, M. N., & Kim, S. (2019). Pathways relating the neurobiology of attachment to drug addiction. *Frontiers in Psychiatry*, *10*, 737. https://doi.org/10.3389/fpsyt.2019.00737

Takahashi, T., Ikeda, K., Ishikawa, M., Kitamura, N., Tsukasaki, T., Nakama, D., & Kameda, T. (2005). Interpersonal trust and social stress-induced cortisol elevation: *NeuroReport*, *16*(2), 197–199. https://doi.org/10.1097/00001756-200502080-00027

Takahashi, Y. (2008). Silencing the critics: Understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. *Frontiers in Neuroscience*, *2*(1), 86–99. https://doi.org/10.3389/neuro.01.014.2008

Wan, M. W., Downey, D., Strachan, H., Elliott, R., Williams, S. R., & Abel, K. M. (2014).

The neural basis of maternal bonding. *PLoS ONE*, *9*(3), e88436. https://doi.org/10.1371/journal.pone.0088436

Whirledge, S., & Cidlowski, J. A. (2013). A role for glucocorticoids in stress-impaired reproduction: Beyond the hypothalamus and pituitary. *Endocrinology*, *154*(12), 4450–4468. https://doi.org/10.1210/en.2013-1652

Wise, R. A. (2009). Roles for nigrostriatal—Not just mesocorticolimbic—Dopamine in reward and addiction. *Trends in Neurosciences*, *32*(10), 517–524. PubMed. https://doi.org/10.1016/j.tins.2009.06.004

Wittmann, M., Lovero, K. L., Lane, S. D., & Paulus, M. P. (2010). Now or later? Striatum and insula activation to immediate versus delayed rewards. *Journal of Neuroscience, Psychology, and Economics*, *3*(1), 15–26. https://doi.org/10.1037/a0017252

Xu, L., Becker, B., & Kendrick, K. M. (2019). Oxytocin facilitates social learning by promoting conformity to trusted individuals. *Frontiers in Neuroscience*, *13*, 56. https://doi.org/10.3389/fnins.2019.00056

Ziv, Y. (2005). Attachment-based intervention programs: Implications for attachment theory and research. In *Enhancing early attachments: Theory, research, intervention, and policy.* (pp. 61–78). Guilford Press.

Figure Captions

Figure 1. **PRISMA flow diagram for the functional magnetic resonance imaging article selection process**. (A) Trust and (B) Attachment.

N, number of studies found eligible in each step of the selection process.

Figure 2. **Meta-analytic results**. Results of the main effect analyses showing regions that were consistently activated during (A) trust in the one-round trust game (TG) (i.e., calculus-based trust); (B) trust in the multi-round TG (i.e., identification-based trust); (C) feedback learning in the multi-round TG; and (D) attachment. (E) Overlap between feedback in the multi-round TG and attachment.

dAI, dorsal anterior insula; NAc, nucleus accumbens; Cau, caudate; vmPFC, ventral medial prefrontal cortex; Put, putamen; R, right; L, left; $\cap$, conjunction

Table 1. Descriptions of the experiments in the selected studies.

Paper	N	Experiment	F
		Trust in One-round TG	
Aimone et al. (2014)	30	Human > Lottery, Average	3
	30	Trust > No Trust	2
	30	Human > Lottery, Trust	2
	8	Human Betrayal-averse > Human Not Betrayal-averse	1
Lauharatanahirun et al. (2012)	30	Trust Game > Gamble Game	9
Stanley et al. (2012)	40	Black > White (Regression)	8
	40	Black > White, Amount sent (Parametric Effect)	2
	40	Black > White (Group Contrast)	2
	40	Black > White, Amount sent (Group Contrast)	3
Kang et al. (2011)	16	Warm > Neutral Temperature	4
	16	Cold > Neutral Temperature	6
	16	Cold > Warm Temperature	9
McCabe et al. (2001)	6	Human > Computer, Trust	1
Fairley et al. (2019)	20	Investment Level (Parametric)	2
	20	TG > Lottery	2
Wei el al. (2019)	27	Conformity > Non-conformity	10
Prochazkova et at. (2018)	34	Trust Level (Parametric)	1
Engelmann et al. (2019)	51	No threat > Threat	8
	51	Threat > No threat	2
		Trust in multi-round TG	

Fett et al. (2019)	47	Investment > Control, Cooperative partner	10
Lemmers-Jansen et al. (2017)	43	Cooperative Investment > Unfair Investment, Increasing with Age	3
	43	Cooperative Investment, Investment, Increasing with Age	3
Ide et al. (2018)	16	Investment > Repayment (Parametric Effect)	4
Riedl et al. (2015)	18	Trust > Control	8
Gromann et al. (2014)	33	Healthy Controls > Patient Siblings	3
Bereczkei et al. (2013)	12	Investor > Control, High Machiavellians > Low Mchiavellians	14
Fourgnan et al. (2013)	18	Trust > No Trust	4
	18	Prior > No Prior	2
	18	No Prior > Prior	2
	18	Parametric Regression of Choice	4
Fourgnan (2013)	18	Trust > No Trust	8
Wardle et al. (2013)	29	Main Effect of Reputation during Trust Decision	7
Fareri et al. (2012)	18	Trust (Main Effect)	3
	18	Trust (Effect of Condition)	14
Fett et al. (2012)	45	BOLD Increase with Age, Cooperative	7
	45	Higher Activation with Age, Cooperative	2
Xiang et al. (2012)	44	High Prediction > Low Prediction	2
Sripada et al. (2009)	26	Human > Computer	9
Baumgartner et al. (2008)	24	Oxytocin > Placebo (Pre-feedback)	1
	24	Placebo > Oxytocin (Pre-feedback)	2

	24	Placebo > Oxytocin (Post-feedback)	10
Krueger et al. (2008)	44	Trust > Reciprocity	3
Krueger et al. (2007)	44	Trust > Control	2
	22	Trust > Reciprocity (Non-defectors)	1
	22	Trust > Reciprocity (Defectors)	1
Delgado et al. (2005)	12	Trust > Distrust	8
King-casas et al. (2005)	48	Trust > Reciprocity	1
Smith-Collins et al. (2013)	24	Reinforcement > No Reinforcement (Expected Cooperation)	2
		Feedback in Multi-round TG	
Fairley et al. (2019)	20	Outcome Phase (Parametric Effect)	4
Fareri et al. (2015)	26	Friend Reciprocated > All other conditions	4
	26	Correlation with Prediction Error Signal	6
Gromann et al. (2014)	33	Healthy Controls > Patient Siblings	3
Fouragnan et al. (2013)	18	No Prior > Prior	2
	18	Consistent > Inconsistent	2
Smith-Collins et al. (2013)	24	Expected > Unexpected	1
	24	Reinforcement > No Reinforcement (for Expected Cooperation)	2
	24	Cooperation > Betrayal (Unexpected)	6
	24	Cooperation > Betrayal (Unexpected, Increase)	5
Fareri et al. (2012)	18	Positive > Negative	19
	18	Main Effect of Condition	3
	18	Correlation with Prediction Error signal (all	6

		conditions)	
	18	Correlation with Prediction Error signal (social condition only)	15
	18	Correlation with Learning Rates	7
Xiang et al. (2012)	102	High Prediction > Low Prediction (Level 0 Investors)	4
	102	Level 0 Investors > Level 1 Investors	1
	44	Level 2 Investors > Level 0 Investors	1
Phan et al. (2010)	36	Reciprocated > Betrayed	6
Delgado et al. (2005)	12	Positive > Negative Feedback	11
		Attachment	
Atzil et al. (2011)	23	Own Infant > Unfamiliar Infant	19
Barrett et al. (2012)	22	Own Infant > Unfamiliar Infant, Positive	7
	22	Own Infant > Unfamiliar Infant, Negative	7
Bartel et al. (2004)	19	Own Child > Acquainted Child	28
Hoekzema et al. (2017)	20	Own Child > Other Child	20
Leibenluft et al. (2004)	7	Own Child > Familiar Child	36
Michalska et al. (2014)	34	Own Child > Other Child	35
Nitschke et al. (2004)	6	Own Child > Other Child	6
Noriuchi et al. (2008)	13	Own Child > Other Child	14
Ranote et al. (2004)	10	Own Infant > Unknown Infant	3
Schechter et al. (2012)	9	Separation > Play, Own Child	5
Strathearn et al. (2008)	26	Own Child > Unknown Child	10
Wan et al. (2014)	20	Own Child > Unknown Child	26

Bartels and Zeki (2000)	17	Romantic Partner > Friends	13
Ortigue et al. (2007)	36	Romantic Partner > Friends	27
Maier et al. (2019)	35	Intimate > Control	7
Hamilton et al. (2017)	20	Romantic Stimuli > Neutral	8
Vila et al. (2019)	21	Lover Faces > Neutral Faces	18
Yin et al. (2018)	32	High Romance > Low Romance	14
Zeki et al. (2010)	24	Loved > Neutral	11
Kim et al. (2009)	10	Romantic-love related stimuli at the early stage within 100 days	13
	10	Romantic-love related stimuli in 6 months	16

N, number of participants; F, number of foci; TG, trust game.

Table 2. Experiments' average contribution to each significant cluster in percentage

Cluster #	Paper	Experimental contrast	Average contribution (%)
Calculus-based trust (Trust decisions in the one-round TG)			
1	Aimone et al. (2014)	Human > Lottery, average	24.93
	Aimone et al. (2014)	Trust > Distrust	8.23
	Aimone et al. (2014)	Interaction between Partner and Choice	29.24
	Aimone et al. (2014)	Human_Betrayal_Averse > Human_Not_Betrayal_Averse	24.10
	Stanley et al. (2012)	Partner_Black > Partner_White	13.48
Identification-based trust (Trust decisions in the multi-round TG)			
1	Bereczkei et al. (2013)	Investor > Control	1.62
	Fouragnan et al. (2013)	Prior > No-Prior	25.78
	Fareri et al. (2012)	Trust Decision (Effect of Condition)	19.84
	Sripada et al. (2009)	Human > Computer	0.55
	Krueger et al. (2007)	Trust > Control	29.18
	Krueger et al. (2007)	Trust > Reciprocity (non defectors)	23.02
Feedback in the multi-round TG			
1	Fairley et al. (2019)	Investment level parametric	12.56
	Fareri et al. (2015)	Friend Reciprocity > All Other Conditions	15.07
	Fareri et al. (2015)	Tracking Prediction Error Signal	10.30
	Gromann et al. (2014)	Healthy Controls > Patient Siblings	12.01
	Fareri et al. (2012)	Positive > Negative Feedback	14.11
	Fareri et al. (2012)	Good > Bad > Neutral > Lottery	14.09
	Fareri et al. (2012)	Good > Bad > Neutral	14.09

	Phan et al. (2010)	Trust Reciprocated > Trust Betrayed	7.70
2	Fairley et al. (2019)	Investment level parametric	5.77
	Fareri et al. (2015)	Friend Reciprocity > All Other Conditions	8.35
	Fareri et al. (2015)	Tracking Prediction Error Signal	8.34
	Fouragnan et al. (2013)	No-Prior > Prior	4.58
	Fouragnan et al. (2013)	Consistent > Inconsistent	5.78
	Smith-Collins et al. (2013)	Expected > Un-Expected	7.95
	Smith-Collins et al. (2013)	Reinforcement > No reinforcement	6.37
	Fareri et al. (2012)	Positive > Negative Feedback	7.70
	Fareri et al. (2012)	Good > Bad > Neutral > Lottery	5.21
	Fareri et al. (2012)	Good > Bad > Neutral	7.13
	Xiang et al. (2012)	Investor_Level (High > Low)	7.59
	Xiang et al. (2012)	Investor_Level_0 > Investor_Level_1	6.74
	Phan et al. (2010)	Trust Reciprocated > Trust Betrayed	6.80
	Delgado et al. (2005)	Positive > Negative Feedback	11.69
	Attachment (Maternal love + Romantic love)		
1	Bartel et al. (2004)	Own child > Acquainted Child	14.30
	Hoekzema et al. (2017)	Own child > Other Child	6.71
	Michalska et al. (2014)	Own child > Other child	29.77
	Wan et al. (2014)	Own child > Unknown	0.12
	Vila et al. (2019)	Loved faces > Neutral faces	16.08
	Yin et al. (2018)	High- > Low-romance conditions	32.95
2	Atzil et al. (2011)	Own Infant > Unfamiliar Infant	17.48

Bartel et al. (2004)	Own child > Acquainted Child	5.70
Hoekzema et al. (2017)	Own child > Other Child	5.92
Leibenluft et al. (2004)	Own child > Familiar Child	16.40
Noriuchi et al. (2008)	Own child > Other	17.58
Ranote et al. (2004)	Own infant > Unknown Infant	1.35
Bartels et al. (2000)	Partners > Friends	3.52
Ortigue et al. (2007)	Love > Friends	13.47
Hamilton et al. (2017)	Romantic > Neutral	17.29
Vila et al. (2019)	Loved faces > Neutral faces	1.27

Table 3. ALE analyses results

Cluster #	Voxels	Regions	BA	MNI coordinates			peak *Z*
				x	y	z	
Calculus-based trust (Trust decisions in the one-round TG)							
1	137	Dorsal anterior insula	13	40	20	2	5.13
Identification-based trust (Trust decisions in the multi-round TG)							
1	100	Nucleus accumbens	25	-5	7	-9	4.39
Feedback in the multi-round TG							
1	575	Caudate head+ body	25	12	18	-4	6.36
2	277	Putamen	34	-16	2	-12	6.39
Attachment (Maternal love + Romantic love)							
1	418	Caudate head+ body	25	14	18	6	4.86
2	102	Medial Frontal Gyrus (ventromedial prefrontal cortex)	10	-2	54	-4	4.06
Feedback ∩ Attachment							
1	93	Caudate head + body	25	6	12	0	4.14

Results of main effect analyses investigating regions consistently activated across studies during each stage of the TG and attachment responses. In parentheses, anatomical assignments from the SPM Anatomy toolbox. Peak Z are Z values at the peak of the activation cluster. BA, Brodmann area; MNI, Montreal Neurological Institute; mm, millimeters; ∩, conjunction.