How to trust a perfect stranger: predicting initial trust behavior from resting-state brain-electrical connectivity

Tim Hahn,1 Karolien Notebaert,1,2 Christine Anderl,1 Vanessa Teckentrup,1 Anja Kaßecker,1 and Sabine Windmann1

1Department of Cognitive Psychology II, Johann Wolfgang Goethe University Frankfurt am Main, Germany and 2Research Center for Marketing and Consumer Science, Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

Reciprocal exchanges can be understood as the updating of an initial belief about a partner. This initial level of trust is essential when it comes to establishing cooperation with an unknown partner, as cooperation cannot arise without a minimum of trust not justified by previous successful exchanges with this partner. Here we demonstrate the existence of a representation of the initial trust level before an exchange with a partner has occurred. Specifically, we can predict the Investor’s initial investment—i.e. his initial level of trust toward the unknown trustee in Round 1 of a standard 10-round Trust Game—from resting-state functional connectivity data acquired several minutes before the start of the Trust Game. Resting-state functional connectivity is, however, not significantly associated with the level of trust in later rounds, potentially mirroring the updating of the initial belief about the partner. Our results shed light on how the initial level of trust is represented. In particular, we show that a person’s initial level of trust is, at least in part, determined by brain electrical activity acquired well before the beginning of an exchange.

Keywords: Trust Game; resting-state; electroencephalography (EEG); functional connectivity; reciprocity

INTRODUCTION

The key to cooperative behavior is trust. It is this willingness to take the risk of helping another person despite the possibility of nonreciprocation that enables us to overcome the fear of being exploited and work together. The absence of trust, on the other hand, renders us reluctant to engage in reciprocal exchanges in the first place, thereby leading us to forego the potential benefits of cooperation (Rilling and Sanfey, 2011; De Dreu, 2012). In line with this, trust is essential for almost all social interactions from romantic relationships (Rilling and Sanfey, 2011) to the large-scale reciprocal exchanges underlying economic growth (Knack and Keefer, 1997, Zak and Knack, 2001). Despite the classic game theoretical insight that the subgame-perfect Nash equilibrium is not to trust a partner, trust is ubiquitous in society and a vast number of empirical investigations show that humans consistently display trust and therefore engage in cooperative reciprocal exchanges with unknown partners (Camerer, 2003; McCabe and Smith, 2000; Rand et al., 2012).

Intuitively, reciprocal exchanges can be understood as the updating of an initial belief about a partner. As information becomes available through interaction with the partner, this initial belief is updated. In case the partner behaves according to our expectations, the current level of trust remains unchanged. If he, however, violates our expectations—in a positive or negative manner—our level of trust toward this partner is adjusted. Recent formal models of trust behavior mirror this notion in that they assume an initial level of trust which is updated when information about the partner is presented for the first time. In the second exchange, the initial level of trust updated using the partner’s behavior in the first exchange determines one’s own behavior. Consequently, behavior in the third exchange is determined by updating the initial level of trust using the partner’s behavior in the first and second exchange and so on (Ray et al., 2009). In short, this can be understood as the mathematical equivalent of getting to know the partner based on his behavior during past exchanges.

While the mechanism by which information about the partner is taken into consideration is fundamental to the progress and development of any reciprocal exchange, the initial level of trust plays the crucial role in starting to cooperate: for example, the initial trust level must be positive for a person to engage in a potentially costly interaction at all. In other words, cooperation cannot arise without a minimum of trust, which—by definition—is not justified by previous successful exchanges with this partner. A simple, but prominent, example of this is buying behavior in online shops where initial trust, which is not justified by previous successful exchanges, must be placed for cooperation to occur. E-commerce thus crucially depends on the customer’s initial trust in the unknown web-based vendor ([Torkzadeh and Dhillon, 2002]; note that shops commonly make previous customer evaluations publicly available to decrease dependence on initial trust levels). Later exchanges will then depend on previous experiences with this vendor as described above.

As the initial level of trust is obviously not determined by previous interaction with the partner, the question of how the initial level of trust is formed and represented arises. To this end, we asked whether there exists a representation encoding one’s initial trust level before any aspect of an exchange—such as the game’s context (e.g. the partner)—is known. If so, this would provide direct evidence for a subject-dependent component of the initial trust level, formed and maintained well ahead of actual behavioral expressions.

The attempt faced us with the issue of having to measure a representation without activating it. The dilemma is that measuring the representation using questionnaires, instructions or any experimental setups would induce context, thereby altering the to-be-measured variable that was supposed to be subject-driven and context-free. Circumventing such issues, we choose to measure brain electrical activity in participants during restful wakefulness several minutes before the start of a standard Trust Game (Berg et al., 1995). Using this particular source of information guarantees that our measurements are fully independent of information about the game-specific experimental setup used to measure trust behavior. In particular, we ensured that participants neither knew their partners’ identities nor did they...
know they would be playing a Trust Game after resting-state data acquisition. As previous tasks have also been shown to impact resting-state dynamics (Pyka et al., 2009, Pyka et al., 2013), we furthermore ensured that resting-state data were acquired not only before the Trust Game, of which we played 10 rounds, but also before any other tasks or questionnaires were administered.

We hypothesized that the Investor’s initial investment—i.e. his initial level of trust toward the unknown trustee in Round 1 of the Trust Game—can be predicted from resting-state functional connectivity acquired several minutes before the start of the Trust Game. As we assumed that belief updating ought to strongly affect initial trust levels, we additionally expected that resting-state functional connectivity is not significantly associated with the level of trust shown in later rounds.

**METHODS**

**Participants**

Sixty healthy subjects participated in the present study. Two subjects had to be excluded owing to missing questionnaire data or technical problems during electroencephalography (EEG) data acquisition. The remaining sample (n = 58) consisted of 29 females and 29 males with a mean age of 22.91 years (s.d. = 3.03). All were recruited from the student community at the University of Frankfurt am Main, Germany, through flyers and advertisements in social networks and electronic mailing lists. All participants gave written informed consent after a complete description of the study was provided. Our study was approved by the ethics committee of the German Society of Psychology (Deutsche Gesellschaft für Psychologie), and all of the procedures involved were in accordance with the latest version (fifth revision) of the Declaration of Helsinki.

**The Trust Game**

We conducted a standard 10-round Trust Game (Berg et al., 1995). One of the participants (Investor) started each round with 8 monetary units (MU), which he was asked to split between himself and the second player (trustee). The investor could choose to transfer 0 MU, 2 MU, 4 MU, 6 MU or the complete amount of 8 MU to the Trustee. The amounts transferred and kept by the investor were presented to both players simultaneously. In addition, the amounts were visualized using red bars whose heights corresponded to the amount given and kept relative to the overall available amount. The amount transferred by the investor was then tripled and transferred to the trustee who could choose any integer between 0 and the tripled amount invested by the investor to repay the investor. Again both players received the information about the trustee’s decision as described above. Then, a new round started with the investor again being endowed with 8 MU.

After the 10 rounds, each player’s MU were summed and converted into an amount in Euro that was paid in cash to the participants. Before the start of the game, both players were informed that their decisions during the Trust Game would determine the amount of their respective payment.

**Resting-state EEG acquisition**

Participants were asked to sit in a quiet room in front of a computer screen and relax. EEG data were continuously recorded during a 5 min period using BrainAmp DC-amplifiers (BrainProducts, Gilching, Germany; resolution 0.1 μV/bit) and Brain Vision Recorder (version 1.05; Brain Products, GmbH, Gilching, Germany), with alternating 40 s eyes-closed and 20 s eyes-open intervals separated by sound signals generated using Presentation (version 14.7, www.neurobs.com). During data acquisition, impedances were kept below 10 kΩ. An equidistance 64-channel EEG arrangement (Montage No. 10, EASYCAP GmbH, Herrsching, Germany) with the reference electrode positioned at Cz was used to record data at a sampling rate of 500 Hz. The ground electrode was located on the right mastoid. Two further electrodes were placed on the lower-outside of the right and left eye to record ocular artifacts.

**Resting-state EEG preprocessing**

Data preprocessing was conducted using the Fieldtrip-toolbox (version 20130617, Donders Institute for Brain, Cognition and Behavior, Nijmegen, The Netherlands) with Matlab (R2013a, The MathWorks Inc., Natick, Massachusetts). In accordance with previous work (Thatcher et al., 2007, Schlegel et al., 2012), only eyes-closed recordings were selected for further analyses. For each participant, the data were segmented into 2 s epochs employing a baseline correction using 0.1 s before each interval. Each channel was re-referenced to the average of all channels. Then, data were down-sampled 250 Hz and a band-stop filter (0.1–40 Hz range) using a two-pass reverse Butterworth filter was applied. If artifacts (e.g. ocular movements, muscle and cardiac contamination, spike or sharp waves) were identified via visual inspection, the entire 2 s epoch was discarded. Of the 5800, 2 s epochs (n = 58 * 100 epochs per person), 3733 epochs were artifact-free. On average, 64.36% of the epochs were retained (s.d. = 13.53%) for each participant.

Brain-electrical connectivity between all 61 head channels was assessed for each participant separately by computing Pearson correlations between the time courses of all channels for each epoch of each participant. This yields one matrix of correlations between all channels for each epoch and each participant. Then epoch-connectivity matrices were averaged over all epochs resulting in one connectivity matrix per subject. As the resulting connectivity matrices are symmetric, only the upper triangle [(61^2 − 61)/2 = 1830 unique connectivities per person] was used for further analyses.

**Predicting initial trust levels**

To predict a participant’s (initial) investment from brain-electrical connectivity during rest, multiple regression was performed using the Classification and Regression Tree algorithm (Breiman et al., 1984) as implemented in Matlab (The Mathworks, Natick, Massachusetts). To ensure the generalizability of the regression model, we used leave-one-out cross-validation (LOO-CV) to predict a participant’s (initial) investment. In each LOO-CV run, data from all but one sample (S-1 of the S subjects) is used to train the model. Subsequently, the investment of the remaining subject, which has so far been unseen by the algorithm, is predicted. This procedure is repeated S times, each time leaving out a different subject, yielding each subject’s predicted investment. The quality of the prediction is assessed by computing the mean squared error (MSE). To establish whether the observed MSE is significantly higher than chance level, we ran each model 1000 times with randomly permuted investments and counted the number of permutations, which achieved lower MSE (i.e. higher performance) than the one observed with the true investments. The P-value was then calculated by dividing this number by 1000.

To quantify the contribution of each EEG electrode, we computed electrode-specific importance scores by taking the mean of all feature importance scores over all cross-validation folds. Thereafter, we summed the mean importance scores for each electrode separately.

**RESULTS**

To test whether the resting-state functional connectome encodes the initial level of trust, we used a multivariate pattern recognition algorithm to predict the Investor’s initial investment (i.e. the investment in Round 1 of the Trust Game) from resting-state functional
connectivities. Avoiding circularity bias by using a LOO-CV procedure, we show that the Investor’s initial investment can be predicted from the human functional resting-state connectome acquired several minutes before the beginning of the Trust Game (MSE of the prediction = 5.35; \( P = 0.028 \)). Using the same approach to predict the investment in any of the other nine rounds of the Trust Game yielded nonsignificant results (\( P > 0.233 \); Figure 1).

Next, we investigated the contribution of each anatomical location (i.e. EEG electrodes) to overall prediction of the initial trust level by summing the feature importance scores from the multivariate pattern recognition algorithm for each electrode. As depicted in Figure 2, we found that the algorithm’s performance was driven to a large extent by connections of electrodes located over the parietal cortex (Electrodes 12, 14 and 28 in Figure 2). In addition, an electrode over the right frontal cortex substantially contributed to the prediction (Electrode 37). This, however, is solely due to its connection with one of the three parietal electrodes (Electrode 14). The contribution of these four electrode sites far exceeded the sum of importance scores of all other electrodes. In fact, importance scores from all other electrodes combined accounted for only \( \sim 77\% \) of the importance of the four sites alone.

**DISCUSSION**

In this study, we showed that the Investor’s initial investment—i.e. his initial level of trust toward the unknown trustee in Round 1 of the Trust Game—can be predicted from resting-state functional connectivity acquired several minutes before the start of the Trust Game. In accordance with our hypotheses, we also found that resting-state functional connectivity is not significantly associated with the level of trust in later rounds. This evidence suggests that there exists a representation encoding one’s initial trust level before any aspect of an exchange. Considering that trust behavior could not be predicted after interaction with the partner had occurred (i.e. in Rounds 2-10 of the Trust Game), the initial trust level does not appear to be hard-wired, though, but is dynamically adjusted depending on the partner’s behavior.

Generally, our finding that the initial trust level can be predicted from the pattern of brain electrical resting-state network connectivities is in agreement with a number of prior studies that found an association between resting-state dynamics and personality traits (Adelstein et al., 2011; Dawes et al., 2012; Hahn et al., 2012; Kunisato et al., 2013; Hahn et al., 2013). The resting-state dynamics are thought to impose constraints on the range of possible neural responses to stimulus input and task context, thereby defining personality (Kannurpatti et al., 2012; Başar, 1998; Basar, 1997). According to our results, the neural correlates of the initial level of trust might affect behavior in a similar fashion: implemented in resting-state dynamics, they might determine how the brain responds in social situations to the range observed in the phenotype [for a more detailed description and additional evidence supporting this notion, see (Hahn et al., 2012, Hahn et al., 2013)]. In particular, this framework might help to explain how the initial level of trust can be encoded in the brain in a stable manner—similar to personality traits—while allowing experience with the partner’s behavior to overwrite it.

While our findings show that the initial level of trust can be predicted before the beginning of an exchange, our data does not speak to the question of how long before an exchange the initial trust level is set. On the one hand, individual differences in initial trust level could be assumed to be highly stable over time, as trusting behavior is partly heritable (Cesarini et al., 2008, Kanagaratnam et al., 2009) and heavily depends on other-regarding preferences (Falk and Fischbacher, 2006; Fehr and Camerer, 2007; Fehr and Gintis, 2007), which, in turn, are highly stable over the life span (Eisenberg et al., 1999). On the other hand, it has been shown that the level of initial trust can be affected by a person’s emotional state with positively valenced emotions leading to increased trust and negatively valenced emotions (e.g. anger) decreasing it (Dunn and Schweitzer, 2005). In that sense, short-term fluctuations in resting-state connectivities may play a role as well. In our view, it appears most likely that situational changes affect resting-state dynamics depending on stable trait-like characteristics, so that resting-state dynamics that are continuously changing and those that are stable over time interact. Based on this view, we would predict individual behavioral variance of initial trust levels to be explained by, both, current emotional state and other-regarding preferences. Only further research can address the question of how fluid and stable characteristics of resting-state dynamics might interact to produce such behavior.

Investigating the initial level of trust and the mechanism by which it is updated, it has been shown that the two components can be affected independently: on the one hand, individuals who received oxytocin did not decrease their trust in response to a partner’s breach of trust, while those participants who had received a placebo adjusted their behavior as would be expected. In short, oxytocin affects the way in which information about the partner is incorporated, presumably by its suppression of amygdala activity, which is thought to dampen the fear of betrayal. However, it does not influence the initial level of trust (Baumgartner et al., 2008, Kosfeld et al., 2005). Against this background, the question arises of whether the mechanism by which the initial trust level is updated is also encoded by resting-state dynamics, i.e. not the behavior in later rounds itself could be encoded in resting-state dynamics, but the more general strategy with which a person responds to a partner’s behavior. To this end, future research should investigate the possible link between strategies in the Trust Game and resting state.

Considering the contribution of each of the connections to overall prediction of the initial trust level, we showed that the connections of electrodes over the parietal cortex provide unique information about a person’s initial level of trust compared with all other electrodes’ connectivities. As with all EEG-based approaches, these results do not allow for exact anatomical specification. Source localization is particularly hampered in our case by two issues: first, we used brain-electrical connectivities (not e.g. event-related potentials) in our analyses, requiring specific source localization algorithms, which are still animatedly debated within the community (Pascual-Marqui et al., 2011).
Secondly, when estimating the contribution of each functional connection to predictive performance, one should be aware that the maps describe a nonlinear multivariate pattern. Generally, importance scores can be meaningfully interpreted only in the context of the entire multivariate pattern—not in isolation. Nonetheless, previous studies investigating this area of the cortex, found the temporal parietal junction (TPJ) activated during investment decisions (Fett et al., 2013) and when the investor’s decision was revealed to the trustee (van den Bos et al., 2011). Generally, the TPJ has been described as part of the mentalizing system (Fletcher et al., 1995; Ruby and Decety, 2004; Gobbini et al., 2007; Van Overwalle, 2009) and it has been argued to play a role in mentalizing during decisions about how much to trust while predicting the game partner’s behavior. In this context, it appears plausible that parietal regions such as the TPJ play an essential role in the kind of trusting behavior observed in our study. Generally, it has been argued that the properties of resting-state dynamics in a certain region might define the extent to which this area can respond to internal or external input (Hahn et al., 2012, Mennes et al., 2010, Hahn et al., 2013), and recent evidence appears to support this notion (Kannurpatti et al., 2012). Considering the limitations of EEG-based functional localization outlined above, interpreting our results against the background of the current literature hints toward a central role of parietal regions in accordance with previous evidence from functional neuroimaging implicating the TPJ.

As we used a multivariate regression approach to predict the initial level of trust, we cannot infer causality of our results, i.e. a third variable might determine both initial trust behavior and resting-state dynamics. The fact that we cannot infer causality, however, does not compromise our findings, showing that the initial trust level is at least partly determined before—versus during—an exchange. Future studies using experimental designs using, for example, transcranial magnetic stimulation or neurofeedback will have to clarify whether resting-state dynamics in the parietal cortex are directly causally responsible for the initial trust behavior.

In summary, we asked whether there exists a representation encoding one’s initial trust level before any aspect of an exchange—such as the game’s context (e.g. the partner)—is known. Shedding light on how the initial level of trust is represented, we showed that a person’s initial level of trust is, at least in part, determined by brain electrical activity acquired well before the beginning of an exchange.

CONFLICT OF INTEREST
None declared.

REFERENCES
Adelstein, J.S., Shehzad, Z., Mennes, M., et al. (2011). Personality is reflected in the Brain’s intrinsic functional architecture. PloS One, 6(11), e27633.
Başar, E. (1998). Brain Function and Oscillations. New York, NY: Springer, Berlin.
Başar, E. (1997). Brain alpha activity—New aspects and functional correlates—Preface. International Journal of Psychophysiology, 26(1–3), 1–3.
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), 639–50.
Berg, J., Dickhaut, J., Mccabe, K. (1995). Trust, reciprocity, and social-history. Games and Economic Behavior, 10(1), 122–42.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). Classification and Regression Trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. New York, NY: Princeton, NJ: Russell Sage Foundation; Princeton University Press, xv, 550.
Cesarini, D., Dawes, C.T., Fowler, J.H., Johannesson, M., Lichtenstein, P., Wallace, B. (2008). Heritability of cooperative behavior in the Trust Game. Proceedings of the National Academy of Sciences of the United States of America, 105(10), 3721–6.
Dawes, C.T., Loewen, P.J., Schreiber, D., et al. (2012). Neural basis of egalitarian behavior. Proceedings of the National Academy of Sciences of the United States of America, 109(17), 6479–83.
De Dreu, C.K.W. (2012). Oxytocin modulates cooperation within and competition between groups: an integrative review and research agenda. Hormones and Behavior, 61(3), 419–28.
Dunn, J.R., Schweitzer, M.E. (2005). Feeling and believing: the influence of emotion on trust. Journal of Personality and Social Psychology, 88(3), 736–48.
