Biological Psychology 94 (2013) 192–197

Contents lists available at SciVerse ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Who is honest and why: Baseline activation in anterior insula predicts inter-individual differences in deceptive behavior

Thomas Baumgartner*, Lorena R.R. Gianotti, Daria Knoch

Social and Affective Neuroscience, Department of Psychology, University of Basel, Birmannsgasse 8, CH-4055 Basel, Switzerland

ARTICLE INFO

Article history:
Received 22 January 2013
Accepted 27 May 2013
Available online 2 June 2013

Keywords:
Deception
Dishonesty
Honesty
Individual differences
Neural trait marker
Resting EEG
Source localization
Anterior insula

ABSTRACT

Humans engage in deceptive behavior that negatively affects others. The propensity to deceive is, however, characterized by vast inter-individual heterogeneity that is poorly understood. Attempts to investigate the origins of this heterogeneity have so far mainly relied on subjective measures and have shown little predictive power. Here, we used resting electroencephalography to measure objective and stable individual differences in neural baseline activation in combination with an ecologically valid deception paradigm. Results showed that task-independent baseline activation in the anterior insula, a brain area implicated in mapping internal bodily states and in representing emotional arousal and conscious feelings, predicts individuals' propensity for deceptive behavior. The higher the neural baseline activation in this area is, the lower individuals' propensity to deceive. Moreover, results provide evidence that high baseline activation in the anterior insula is associated with negative affect and dispositional tendencies to avoid aversive emotional situations. These results provide converging neural and psychological evidence that individuals might avoid a deceptive act due to a highly active negative emotional system which would make a deceptive act too stressful and bothersome.

1. Introduction

People lie frequently in everyday social interactions (DePaulo, Kirkendol, Kashy, Wyer, & Epstein, 1996). Although not all of these lies are harmful, people often engage in deceptive behavior that negatively affects others (DePaulo, Ansfield, Kirkendol, & Boden, 2004).

Despite the pervasiveness of deception in human society, considerable inter-individual differences in the propensity to deceive can be observed in laboratory and field studies (Kashy & DePaulo, 1996; Phillips, Meek, & Vendemia, 2011). Attempts to explain this inter-individual variation in deceptive behavior primarily focused on psychological variables such as stable personality traits (e.g. Kashy & DePaulo, 1996; McLeod & Genereux, 2008; Phillips et al., 2011) and personality disorders (e.g. Spidel, Herve, Greaves, & Yuille, 2011). Correlations between personality traits and deception, however, have yielded rather variable and inconsistent results. For example, machiavellian personality traits correlated in some studies with deception (e.g. McLeod & Genereux, 2008), but failed to show this correlation in others (e.g. Phillips et al., 2011). One reason for these mixed results might lie in the subjective nature of the measurements employed (i.e. self-reports of personality traits and deceptive behavior), which are known to be affected by various biases (e.g. demand characteristics, social desirability). Moreover, the predictive power of these personality traits was rather low; often below 5% of variance could be explained by a single personality trait (e.g. DePaulo, 2004). Thus, the use of more objective individual trait measurements in combination with more ecologically valid deception paradigms might be beneficial in understanding the inter-individual differences in the propensity to deceive.

An ideal trait measurement of this type is task-independent neural baseline activation measured by resting electroencephalography (EEG) because this measurement demonstrates high stability over time and high specificity (i.e. the extent to which an EEG pattern uniquely belongs to a given person). Studies investigating the stability of resting EEG revealed test–retest reliabilities of up to 0.8 over a period of up to 5 years (Cannon et al., 2012; Dunki, Schmid, & Stassen, 2000; Gold, Fachner, & Erkkila, 2013; Napflin, Wildi, & Sarnthein, 2007; Smit, Posthuma, Boomsma, & Geus, 2005; Williams et al., 2005) and studies exploring the specificity revealed recognition rates of up to 99% (Dunki et al., 2000; Napflin et al., 2007). Due to high intra-individual stability and specificity, this measurement provides an ideal neural trait marker to investigate the sources of inter-individual differences in deceptive behavior.

To measure variation in deceptive behavior in an ecological valid situation, we used a paradigm resembling an economic exchange situation between two interaction partners – an investor and a trustee. In this paradigm, individuals were free to deceive or to tell

---

* Corresponding author. Tel.: +41 61 267 0639.
E-mail address: t.baumgartner@unibas.ch (T. Baumgartner).

0301-0511/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.biopsycho.2013.05.018

the truth and all decisions had real consequences for all involved interaction partners. More precisely, the trustee first has to make a promise decision at the beginning of a series of three subsequent trust game trials, indicating whether he/she *always*, *mostly*, *sometimes*, or *never* plans to be trustworthy. In this context, being trustworthy means returning money so that both players earn the same amount. The investor is always informed about the trustee's promise and can then decide (based on the trustee's promise) whether to trust the trustee and invest money or to not trust him/her and thus to keep the initial endowment of 2 money units (MUs). If the investor trusts the trustee, the experimenter increases the amount the investor sends to the trustee by the factor of five. The trustee can then freely decide to be honest and keep the promise, or he/she may also decide to be deceptive and break the promise and thus violate the investor's trust by not returning money.

Previous studies on the neural underpinnings of deceptive behavior focused on brain activation during the decision-making process rather than examining individuals' task-independent neurophysiological characteristics (for recent reviews see Abe, 2011; Sip, Roepstorff, McGregor, & Frith, 2008). Thus, these studies do not identify neural traits responsible for predisposing subjects to deceive or tell the truth. Nevertheless, the findings of these studies can be used to generate hypotheses about the potential neural sources driving individuals' propensity for deceptive and truthful behavior.

One of the most consistent finding in these studies (e.g. Abe, Suzuki, Mori, Itoh, & Fujii, 2007; Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Sip et al., 2010; Spence, Kaylor-Hughes, Farrow, & Wilkinson, 2008) is the increased activation in areas of the prefrontal cortex (e.g. dorsolateral prefrontal cortex, DLPFC, ventrolateral prefrontal cortex, VLPFC) during the deceptive act. Due to the role of these areas in cognitive control (e.g. Miller & Cohen, 2001), response selection (e.g. Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000), and response inhibition (Aron, Robbins, & Poldrack, 2004), these prefrontal activation patterns have commonly been interpreted as to suggest that the suppression of the honest response and the generation of the deceptive response requires control-related processes (Abe, 2011). In line with this assumption is a recent study of patients with Parkinson's disease, which demonstrated that the reduced tendency to deceive in these patients is associated with decreased metabolic rates in the left DLPFC and right anterior prefrontal cortices (Abe et al., 2009). Moreover, studies in patients with a history of pathological lying showed increased white matter volume in the prefrontal cortex (e.g. Yang et al., 2007). These findings together with previous studies showing that the level of neural baseline activation in the lateral prefrontal cortex positively correlates with control abilities (Gianotti, Figner, Ebstein, & Knoch, 2012; Knoch, Gianotti, Baumgartner, & Fehr, 2010) led us to hypothesize that subjects with higher baseline activation in control-related areas of the prefrontal cortex might show an increased propensity to deceive.

Other regions that have been demonstrated to be activated during the deceptive act (e.g. Abe et al., 2007; Baumgartner et al., 2009; Kozel et al., 2005) comprised areas involved in processing emotions (e.g. anterior insula, Craig, 2009), and areas involved in processing conflict (e.g. anterior cingulate cortex, Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999). It has been argued that the decision conflict and the psychological stress (e.g. guilt) that often accompany the generation and enactment of a lie are encoded in these brain areas (Sip et al., 2008). Thus, it might be that differences in the neural functioning of these brain areas affect individuals' propensity to deceive. We hypothesize that subjects with higher baseline activation in emotion- and conflict-related brain areas might show a reduced propensity to lie because a deceptive act would cause too much stress or bothersome emotions (e.g. guilt).

Taken together, we hypothesize that inter-individual differences in the propensity to deceive might be positively associated with baseline activation levels in control-related areas and/or negatively associated with baseline activation levels in emotion- and conflict-related areas.

2. Materials and methods

2.1. Subjects

We measured neural baseline activation and the propensity for deceptive behavior in 50 healthy individuals (mean age ± SD = 23.3 ± 4.9, 19 men and 31 female). One male subject was excluded from analyses based on outlier brain data and regression influence statistics (*Cook's Distance* = 0.164), leaving 49 participants for analyses. All subjects were right-handed and had no history of neurological or psychiatric disorders or alcohol and drug abuse. The study was approved by the local ethics committee. All subjects gave written, informed consent and were informed of their right to discontinue participation at any time. Subjects received 40 Swiss francs (CHF 40; CHF 1 = $1 U.S.) for participating, in addition to the money earned in the deception paradigm.

2.2. Procedure

The deception paradigm and the EEG recordings took place during different sessions, which were separated by several weeks. The deception paradigm was conducted in our behavioral laboratory with interconnected computer terminals, while the EEG recordings were conducted in our EEG laboratory, where only one subject was measured at one time.

2.3. Deception paradigm

Subjects in the role of a trustee played 9 rounds of the deception paradigm with 9 different, anonymous interactions partners in the role of an investor. We deliberately chose anonymous one-shot interactions in order to exclude reputation effects and strategic spillovers across trials. At the beginning of such an interaction, an investor receives an endowment of 2 money units (MUs), whereas the trustee receives nothing. Then, the trustee has to make a promise decision at the beginning of a series of three subsequent trust game trials, indicating whether he/she *always*, *mostly*, *sometimes*, or *never* plans to be trustworthy and return the money. The investor is then informed about the trustee's promise and he/she can decide to send his/her endowment of two MUs to the trustee, or he/she can decide to keep his/her endowment. If the investor keeps his/her endowment, the round ends and the trustee get nothing. However, if the investor trusts the trustee and invests money (which occurred in most cases [89%] due to the high promise level that nearly all trustees chose, see results section for details), the experimenter increases the amount sent by a factor of five, so that the trustee receives 10 MUs. He/she then can freely decide whether to keep the promise or to break it. If the trustee decides to keep his/her promise and return the investment, both players earn 5 money units (5 money units = CHF 1, about $1). If the trustee decides to break the promise and keep the investment, he/she earns 10 money units (=CHF 2, about $2), whereas the investor receives no money in this round. The software package z-Tree, a program for conducting behavioral exchange experiments was used for presenting screens and for collecting behavioral data.

2.4. Questionnaires

We administered the following trait questionnaires: positive and negative affect schedule (PANAS, Watson, Clark, & Tellegen, 1988), behavioral inhibition and behavioral approach system scales (BIS/BAS scales, Carver & White, 1994), and Barratt impulsiveness scale (Patton, Stanford, & Barratt, 1995). The PANAS is a 20-item questionnaire which measures dispositional tendencies to experience negative affect (e.g. distress, unpleasurable engagement) and positive affect (e.g. pleasurable engagement) in life. The BIS/BAS scales consist of 24 items which assesses individual dispositional differences in the sensitivity of two general motivational systems underlying behavior. The behavioral approach system (BAS) is believed to regulate appetitive motives, in which the goal is to move toward something desired. A behavioral avoidance (or inhibition) system (BIS) is said to regulate aversive motives, in which the goal is to move away from something unpleasant. The Barratt impulsiveness scale consists of 30 items which assess individuals' self-control abilities, i.e. how good subjects are able to control impulsive behavior. High values on this scale indicate low control ability.

2.5. EEG recording and processing

Subjects were seated comfortably in a dimly lit, quiet room with intercom connection to the experimenter. Then 64 Ag-AgCl active electrodes were placed following the 10-10 montage covering the entire scalp, as recommended by the International Federation of Clinical Neurophysiology Standards for Digital Recording of Clinical EEG (Nuwer et al., 1998). During the recordings, the signals were