Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Brain Activity Related to Cheating Identification: A Step Towards the Study of Corruption

Víctor E. Olalde-Mathieu
Mind Research Network
Daniel Atilano-Barbosa
National Autonomous University of Mexico
Giovanna L. Licea-Haquet
National Autonomous University of Mexico
Jonatan García-Campos
Universidad Autónoma Metropolitana
Fernando A. Barrios
National Autonomous University of Mexico
Paola Hernández-Chávez
hcpaola@gmail.com

Universidad Autónoma Metropolitana

Article

Keywords: Cheating, Mentalizing, Theory of Mind, fMRI, Executive control, Corruption

Posted Date: July 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8079058/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. F.A.B. receives compensation from Springer Nature for his work as editor. All the other authors have no conflict of interest to disclose.

Abstract

Cheating is a complex social behavior implying the violation of social public rules. It integrates moral, emotional, and cognitive processes. We adapted an Instrument for Detecting Cheating and its Emotional Reactions (INDETRAE scenarios)3 into an fMRI paradigm presenting five cost–benefit cheating conditions: detriment to an undefined person (1), detriment to a third party (2), detriment to yourself (3), neutral no cheating (4), and self-benefit (5) cases. Forty-one adults read each vignette while BOLD responses were recorded. Contrasting all the cheating conditions (1+2+3+5) versus neutral (4) revealed increased activation in the mentalizing network. Specific contrasts showed: (1 > 2+3+5) heightened activation in the precuneus and left lingual gyrus, associated with semantic integration; (2 > 1+3+5) activation in the right posterior middle frontal gyrus, reflecting executive control; (3 > 1+2+5) activation in the left inferior frontal gyrus, associated with semantic and conflict processing; and (5 > 1+2+3) widespread activity in salience, reward, central executive and mentalizing networks, suggesting integrated cognitive control, social inference, affective evaluation, and reward processing during self-serving scenarios. These findings show that different cheating circumstances elicit distinct neural signatures. This reinforces the value of the INDETRAE fMRI paradigm for elucidating the neural substrates involved in cheating identification and typical corruption scenarios.

Introduction

Understanding the neural processes underlying social interactions has been a primary focus of research in cognitive neuroscience. Cheating, as a form of social violation, poses unique challenges as it not only involves the evaluation of another's intentions but also integrates emotional, social, and moral assessments. Cheating has become an increasingly critical area of research, particularly in socio-economic contexts where corruption and dishonest behaviors impact societal and political policies. These concerns have encouraged the development of new questioning techniques [1], machine learning algorithms [2] and behavioral paradigms [3], related to cheating identification.

By studying different contexts of cheating, insights can be gained into how the brain deals with interpersonal deception, self-serving motivations and cost-benefit rationalizations. Identifying the neural mechanisms underlying cheating is essential for understanding how individuals evaluate moral transgressions and navigate social interactions. One key component for this process is the mentalizing network, which includes regions such as the medial prefrontal cortex and the temporoparietal junction. This network is fundamental for interpreting others' intentions and beliefs [4–6], supporting the cognitive processes involved in assessing complex social information and inferring motivations behind deceptive behavior.

Moreover, neuroimaging studies have identified that the prefrontal, temporal, and cingulate cortices are activated when individuals are presented with honest and dishonest decisions, suggesting these areas are involved in the neural processes underlying moral decision-making [7–9]. In addition, the activation of the caudate nucleus, precentral gyrus, and central executive network (CEN), has been associated with

the execution of both deliberately deceptive and honest claims [10,11]. Evidence also suggests that motor areas may be involved in the preparation of self-determined actions and specifically modulated by moral conflict[12,13] .

Conversely, when individuals anticipate detrimental or beneficial outcomes from cheating, there is increased activation in executive control regions, the salience network and areas involved in reward evaluation, such as caudate and putamen [7,10,14–16]. This variation in the recruitment of neural systems pinpoints the complexity of moral and cognitive evaluations that depend on the social context of transgressive events such as cheating.

Importantly, this context-sensitive neural recruitment is consistent with evolutionary theories suggesting that humans developed specialized cognitive adaptations designed to identify violations in social exchange contexts aimed to maintain elementary rules of cooperation [17–19]. Supporting this theory, Ermer et al. [20], showed that medial prefrontal and anterior cingulate cortices, key nodes in theory of mind and salience networks, activate selectively during social-exchange reasoning (e.g., cheating detection).

Various methodologies have been developed to understand how people identify and process cheating. The Wason Selection Task is a psychological experiment designed to assess deductive reasoning and conditional logic. It involves presenting participants with a scenario that requires them to decide which cards to turn over to assess a specific rule. Ermer et al. (2006)[20] employed a version with three content conditions, social-contract (cheating), precautionary (safety rule), and descriptive (neutral), to isolate neural circuits specific to cheating reasoning versus other rule types. Similarly, Liu et al. (2021) [21] introduced the Difference Spotting Task (DST), a nonverbal, unsolvable paradigm in which half of the picture pairs contained no actual differences. Participants are incentivized to report differences, so claiming differences on unsolvable items directly indexes cheating at the trial level. Nevertheless, these tasks lack the evaluation of the cognitive and emotional mechanisms, i.e., where the cheated party differs from the individual who identifies the cheating, specifically between different cost-benefit social situations where there is a benefit or a detrimental outcome due to a third-party cheating action.

To address the limitations of current literature and enhance a better understanding of what is a stake, we used the Instrument for Detecting Cheating and its Emotional Reactions (INDETRAE, in Spanish: Instrumento para la Detección de Trampa y sus Reacciones Emocionales) to assess content-valid stimuli consisting of five contrasting social cost-benefit scenarios based in meaningful corruption contexts. In these scenarios, the cheating situation affects either an undefined person, a third person, or a first person (yourself), a neutral category with no cheating, plus a case where the first person receives a benefit from a cheating situation [3]. The objective of this study is to identify the neural correlates associated to cheating identification in five different cost-benefit scenarios, applying an fMRI paradigm of the standardized instrument INDETRAE. We hypothesize that the fMRI paradigm will disentangle the contributions of the mentalizing, reward, salience, and executive control networks related to the process of cheating identification within the five different conditions. Altogether, this approach will further

advance our understanding of the neural substrates of cheating identification under diverse cost-benefit scenarios that are at play during corruption and rule breaking in social contexts.

Results

Whole brain results

Whole-brain analyses revealed distinct patterns of neural activation across the various cheating identification conditions. Comparing the cheating scenarios (conditions 1, 2, 3, and 5) to the neutral condition (no cheating and no detriment; 1+2+3+5 > 4) showed significant activation in key regions of the mentalizing network, including the medial prefrontal cortex (mPFC), bilateral temporoparietal junction (TPJ), and precuneus (Pcu). Additionally, significant activation was observed in the left caudate nucleus (lCN), left occipital pole (lOcPole), and right Crus I/II (rCrusI/II) (Table 1, Fig. 2).

When condition 1 (cheating to the detriment of an undefined entity) was contrasted with conditions 2, 3, and 5 (1 > 2+3+5), enhanced activation was observed in the precuneus, extending into the left lingual gyrus (Table 1, Fig. 2). The contrast isolating condition 2 (cheating to the detriment of a third person; 2 > 1+3+5) showed significant activation in the right posterior middle frontal gyrus (rpMFG). In contrast, condition 3 (cheating to the first person's detriment (Table 1, Fig. 2); 3 > 1+2+5) recruited the left inferior frontal gyrus (lIFG) (Table 1, Fig. 2).

Finally, condition 5 (cheating with a benefit to the first person), compared to conditions 1, 2, and 3 (5 > 1 + 2 + 3), was associated with significant activation across the Central Executive Network (CEN) and the Salience Network. Engagement of the CEN involved a large prefrontal cortex cluster containing the bilateral dlPFC, and separate activations in the bilateral superior parietal cortex. Similarly, engagement of the Salience Network involved a component within the aforementioned PFC cluster, and a separate activation encompassing the bilateral IFG/AI. Additional activations were observed in the STS/ITPJ, Pcu/PCC, mTS, and bilateral putamen. A more comprehensive display of these contrasts can be found in the Supplementary Material (SFig 1-5).

Discussion

This study aimed to examine the neural mechanisms underlying moral evaluation and deception identification in typical corruption scenarios using the INDETRAE instrument [3], a content-valid fMRI paradigm featuring socially relevant cost-benefit cheating scenarios. As hypothesized, distinct patterns of neural activation emerged across conditions, reflecting differences regarding the implicated social agent and the type of cost-benefit scenario.

Comparing all cheating conditions to a neutral condition with no transgression (contrast 1+2+3+5 > 4) revealed significant activation in key regions of the mentalizing network, including the mPFC, bilateral TPJ, and Pcu [26]. The mentalizing network, supports the inference of others' intentions and the integration of self-other perspectives [27], which are critical processes when evaluating deceptive acts. The TPJ, in

particular, helps decode others' transient mental states such as intentions and beliefs [28,29], and is especially active during deception detection and moral norms violations [4,26]. Bilateral TPJ recruitment in this contrast suggests that participants engaged in perspective-taking and social evaluation, even though the task did not require explicit moral judgments. The mPFC is involved in reflective reasoning and moral evaluation, integrating affective and cognitive signals to assess fairness and intentions [30,31], and is particularly responsive to perceived intentionality in moral violations [32]. The Pcu, implicated in self-referential thought, episodic memory, and visual imagery [33], may support internal simulation and reflection on the implications of each scenario. Additional activations were observed in the left occipital pole (lOcPole), left caudate nucleus (lCaudate), and right cerebellum (rCrus I/II, vermis VI). This combined pattern of activation suggests the engagement of processes spanning mentalizing, social prediction, value-based reasoning, and emotional salience.

Activation in the lOcPole suggests involvement in perceptual simulation of the cheating scenarios. While traditionally associated with low-level visual processing, the occipital pole is also involved in constructing vivid mental imagery and visualizing events in the absence of sensory input [34,35]. Given the social charged nature of the scenarios, top-down modulation from affective processing may have enhanced visual mental simulation [36], aiding participants in constructing meaningful representations of each event. In this contrast, the lOcPole activation occurs in tandem with the ToM network, probably supplying perceptual detail, grounding abstract moral reasoning in concrete visual simulations. This process aligns with models suggesting that social-cognitive tasks require the interplay of scene construction (via posterior cortices) and evaluation/inference (via mPFC) to enable nuanced moral judgment [37,38].

The lCaudate, a region involved in reward prediction and action-outcome monitoring [39], may have tracked the potential consequences of cheating behaviors, such as anticipated harm or retaliation. Its activation aligns with studies linking the striatum to social punishment decisions and adaptive learning in deceitful interactions [40]. In addition, the rCrus I/II activation, increasingly recognized for its role in social cognition, likely contributed to refining predictions about agents' future actions and moral reliability [41,42]. This region is thought to work in concert with cortical mentalizing regions, updating internal models of social behavior and contributing to adaptive response generation [43,44]. Similarly, the lCaudate and lOcPole may also operate synergistically with the mentalizing network, supporting a dynamic interplay between perceptual simulation, social inference, and outcome monitoring. Together, these regions could facilitate a multimodal representation of morally salient events, in which vivid visual imagery, mental state attribution, and reward-based evaluation converge to guide nuanced moral reasoning.

In the contrast comparing cheating directed at an undefined entity (condition 1) to those targeting specific individuals (conditions 2, 3, and 5; contrast 1 > 2+3+5), significant activation was found in the lLing/Pcu extending towards the posterior Pcu. This pattern may reflect enhanced demand for semantic integration and situation modeling in ambiguous scenarios. The lLing/Pcu supports the integration of

visual and semantic information into coherent representations [45], and are involved in internally oriented processing, such as episodic memory retrieval and elaboration of meaning under uncertainty[33,46]. The lLing/Pcu may help resolve ambiguity when contextual details are missing or vague, consistent with previous findings on semantic elaboration [48].

The contrast comparing cheating to the detriment of a third person (condition 2) versus other cheating conditions (2 > 1+3+5), showed increased activation in the rpMFG. This region has been associated with executive control and moral reasoning by regulating emotional responses and integrating contextual and social information [49,50]. Its activation suggests that participants engaged in complex evaluative processes to weigh the interpersonal consequences of cheating that negatively affected others [51].

When comparing scenarios in which the participant receives a detriment (condition 3) to the other cheating conditions (3 > 1+2+5), increased activation was observed in the lIFG. This region has been associated with control and conflict monitoring [51–53]. In this context, activation of the lIFG may reflect conflict monitoring and increased executive demands when processing morally salient scenarios with personal harm. Additionally, the lIFG has been related to semantic integration and contextual interpretation [54,55], which will suggest that participants recruited the lIFG to extract meaning and regulate responses to personally disadvantageous outcomes.

The contrast isolating the self-benefiting cheating condition (5 > 1+2+3) elicited the most widespread activation pattern, engaging a distributed network encompassing both the Central Executive Network and the Salience Network [56]. Collectively, this network supports a range of functions, including self-referential processing, cognitive control, social inference, affective evaluation, and motivational learning. Robust activation was observed in the prefrontal cortex (PFC), including the ventro and dorsomedial PFC (vmPFC/dmPFC) and bilateral dorsolateral PFC (dlPFC). These regions are consistently involved in value-based decision-making, goal-directed reasoning, and self-evaluation [57]. Beyond these roles, the PFC contributes to executive functions such as inhibitory control and the flexible regulation of behavior in response to social and emotional demands [58]. The medial PFC, specifically, supports mentalizing and self-referential thinking, playing a key role in evaluating intentions, both one's own and those of others [59,60]. In addition, the dlPFC facilitates cognitive control over decisions, helping to balance internal goals with external social norms and to suppress self-serving impulses when they conflict with moral standards [58,61]. Together, these PFC regions likely enabled participants to evaluate the self-serving nature of cheating within a broader moral and social framework, even though explicit moral judgment was not required. Notably, these scenarios model a core component of common corruption cases, in which a cost-benefit conflict results in direct personal gain from the transgression. This suggests that the brain regions activated in this condition are likely recruited in similar real-world corruption situations.

Complementing this executive-social control system, increased activation was also found in bilateral TPJ, SPC, lSTS, rmTS and Pcu/PCC, regions for mentalizing and perspective-taking. The TPJ plays a central role in representing others' mental states [4], and its bilateral engagement suggests that

participants may have spontaneously considered how others might perceive their gain through dishonest means [26]. The STS and mTS support the interpretation of dynamic social cues and communicative intent [62], aiding in constructing the social meaning of each scenario. The rSPC, associated with shifting between self- and other-related perspectives [63], may reflect participants' internal negotiation between self-interest and social expectations, contributing to potential moral dissonance. The Pcu/PCC have been particularly implicated in theory of mind, autobiographical memory retrieval, and moral reflection [33]. This activation likely reflects introspective and self-referential processes, such as evaluating one's emotional response to self-serving transgressions or recalling similar experiences.

Furthermore, bilateral IFG/AI was also recruited. The IFG/AI integrates emotional signals and regulates emotional conflict, particularly when social norms conflict with self-interest [51,52,64,65] . Even though the task did not ask for explicit moral judgments, AI engagement suggests that participants experienced internal affective responses when contemplating their own benefit through cheating. Moreover, Its activation likely reflects the need to interpret the meaning of morally ambivalent scenarios while regulating emotional tension. Finally, bilateral activation of the Pu, central to reward processing and goal-directed action [66–68], suggests that participants implicitly perceived self-benefiting scenarios as advantageous.

In sum, all cheating scenarios involve the mentalization network. Scenarios that imply harm to other entity or self-harm elicit activation in particular punctual regions. Similarly, self-serving cheating scenarios showed a particular activation; however, this was broader and encompassed a distributed network implicated in self-referential processing, cognitive control, social inference, affective evaluation, and motivational learning. Even without an explicit moral judgment task, participants' neural responses suggest that internally guided reasoning and affective appraisal mechanisms were deeply engaged, reflecting the morally ambivalent nature of profiting at others' expense. Our findings shed light on the neural mechanisms of cheating in corruption scenarios. The results were obtained using the content-validated INDETRAE instrument [3], which was designed to capture diverse forms of cheating in socially meaningful corruption contexts. Rather than treating moral decision-making, cheating detection, or corruption itself as unitary processes, this study's key strength lies in showing that various cost-benefit scenarios within corruption contexts are associated with different neural signatures. By comparing cheating scenarios that varied in the identity of the affected entity and the potential benefit for the self, we were able to identify condition-specific brain activation patterns, highlighting the functional specificity of domain-general and social-cognitive brain networks, including regions involved in self-referential processing, semantic integration, and social inference. These findings mirror the complexity of corruption scenarios where a cheating and a cost benefit to the first party is always present, revealing that seemingly simple judgments engage sophisticated neural processes that integrate affective evaluation, perspective-taking, reward sensitivity, and cognitive control.

To preserve its content validity while maintaining participant engagement in the scanner, the INDETRAE was adapted into an fMRI-compatible task. Nonetheless, several limitations should be acknowledged.

Although the instrument shows strong content validity [3], it lacks ecological realism, as participants read brief written scenarios rather than deeply and spontaneously engaging with interactive or immersive moral situations. This may constrain the generalizability of the findings to real-world social decision-making. Additionally, because the task did not explicitly ask participants to make moral judgments, the observed neural activity likely reflects implicit forms of moral evaluation, including emotional appraisal and social perspective-taking, rather than deliberate and slow ethical reasoning. Finally, while the sample size was adequate for group-level analyses, future studies with larger and more diverse populations are needed to explore individual variability in wider corruption-cheating scenarios and to replicate the neural specificity observed across cheating types.

Using the INDETRAE paradigm, this study reveals distinct neural patterns underlying cheating scenarios within different corruption contexts. Compared to the neutral condition, cheating scenarios engaged the mentalizing network along with regions for perceptual simulation and reward prediction. Ambiguous cheating scenarios activated semantic integration areas (left lingual gyrus, precuneus), while cheating to the detriment of others engaged executive control regions (right posterior middle frontal gyrus). Detriment to the self-scenarios showed increased activation in the left inferior frontal gyrus, reflecting conflict monitoring. Self-benefiting cheating, elicited the broadest network activation, involving prefrontal, mentalizing, and reward-related regions, indicating complex affective and cognitive processing. These findings reflect the neural specificity of cheating processing depending on social context and type of transgression, even without explicit moral judgment. Opening a path for future work to explore these processes in more realistic settings and diverse populations.

Methods

Participants

Forty-one healthy volunteers (21 woman) aged 19-42 (mean age of 25.8 years, SD=5.88), all right-handed, participated in this study. All participants were native Spanish speakers and provided their informed consent before starting the study. The study protocol was previously approved by the Research Ethics Committee of the Instituto de Neurobiología UNAM and complies with national and international guidelines.

Stimuli and design of the task

The stimuli employed in this study consisted in the cheating identification scenarios of the validated instrument INDETRAE [3]. This instrument includes 90 written scenarios describing situations that fall into five different categories: cheating to the detriment of an undefined entity (1-CDdU), cheating to the detriment of a third person (2-CDdT), cheating with detriment to the first person/yourself (3-CDdF), no cheating and no detriment (4-noCD), and cheating with a benefit to the first person/yourself (5-CDbF).

Each scenario is followed by four questions assessing whether the participant is able to accurately identify cheating, the cost-benefit involved in the situation, and the emotions evoked by the situation. To

minimize fatigue during the task, we selected the 75 scenarios with the highest rates of cheating detection, 15 of each category, and included only two questions designed to maintain participants´ attention and to evaluate their emotional responses.

The selected questions were:

"Who is the adversely affected party in this situation?" There were four possible responses: *the cheater/someone else / yourself / nobody*.

"How does the described situation make you feel?" There were four possible answers: *Glad / Indifferent / Annoyed / Angry*.

During the fMRI session, each event consisted of a written scenario presented for 6 seconds, followed by the two selected questions displayed for 3 seconds each. Participants responded while the questions were being presented. To optimize the timing of stimulus presentation and interstimulus intervals, the *optseq* program was used [22] . Scenarios were randomly presented across trials to prevent order effects. The task was divided into two runs, each lasting 11 minutes and 32 seconds or 11 minutes and 13 seconds respectively (Figure 1). The task was implemented using *PsychoPy2* (v2020.2.10; [23]). Stimuli were displayed on a gray background through goggles, while responses were recorded using button boxes, both compatible with the MRI scanner (NordicNeuroLab, Bergen, Norway).

Procedure

Participants were informed that during the fMRI task, they would be presented with written scenarios followed by two questions. They were instructed to read the scenarios and, based on their understanding of the described situation, respond as quickly as possible by pressing buttons with their thumb and index fingers.

Response options appeared on the same screen as the questions, positioned at the bottom. Each question had four response options: two in the corners and two in the middle (Figure 1). Participants were instructed to press with their thumbs if the selected response was in the corners and with their index fingers if it was in the middle, selecting either the right or left button depending on the word's position on the screen. Before the fMRI session, participants completed fifteen practice trials of the cheating identification task on a laptop outside the MRI scanner, receiving feedback to ensure their understanding. The practice trials included the remaining scenarios from the INDETRAE, with three scenarios for each category.

Image acquisition

All MRI data were acquired at the Unidad de Resonancia Magnética of the Instituto de Neurobiología using a 3T General Electric (GE) MR750 scanner (Waukesha, WI) equipped with a 32-channel head coil. Task-based functional images were acquired using a T2* weighted echo-planar imaging (EPI) with the following parameters: repetition time (TR) = 2000 ms, echo time (TE) = 40 ms, field of view (FOV) = 25.6

cm, and a matrix size of 64 × 64. This sequence included 38 axial slices with a 4-mm slice thickness, resulting in an isotropic voxel size of 4 × 4 × 4 mm^3. Additionally, high-resolution T1-weighted structural images were obtained using a 3D spoiled gradient recalled (SPGR) sequence with a voxel size of 1 × 1 × 1 mm^3. The parameters for this sequence included (TR = 8.1 ms, TE = 3.2 ms, flip angle = 12°), covering the whole brain for anatomical localization.

Whole brain analysis

To analyze fMRI data, we used FSL 5 (FMRIB's Software Library) and performed statistical analysis with the FMRI Expert Analysis Tool (FEAT) using FMRIB's Improved Linear Model (FILM) version 6.00. Each participant's data was motion and slice-timing corrected and then registered to the individual's high-resolution T1-weighted anatomical image before being normalized to Montreal Neurological Institute (MNI) standard space (EPI template, voxel size = 2 × 2 × 2 mm^3). Data were subsequently smoothed with a Gaussian filter (full width at half maximum = 6 mm) and high-pass filtered (cutoff frequency 0.01Hz) to remove low-frequency drifts.

BOLD responses were examined within 12-second event windows corresponding to the presentation of the cheating identification scenarios and associated questions. Statistical analysis of event-related hemodynamic changes was performed using the general linear model (GLM)[24]. The GLM included five regressors of interest corresponding to the five cheating categories—cheating to the detriment of an undefined entity (1-CDdU), cheating to the detriment of a third person (2-CDdT), cheating to the first person's detriment (3-CDdF), no cheating and no detriment (4-noCD), and cheating with a benefit to the first person (5-CDbF). Additionally, the fixation point presentation was modeled as a nuisance regressor[24]. First-level analyses were performed using a fixed-effects model to generate individual activation maps. The resulting maps were used to identify regions with increased BOLD signal for cheating detection events compared to control events. These contrast images were then utilized for group-level analysis using FLAME 1 (FMRIB's Local Analysis of Mixed Effects). Group-level activation maps were thresholded at a cluster significance level of $Z > 2.3$ ($p < 0.05$), corrected for multiple comparisons using Gaussian Random Field theory (GRF [25]). Brain regions reported in this work were identified using the Harvard-Oxford Cortical and Subcortical Atlases, with coordinates referenced to the MNI152 template.

Declarations

Acknowledgements

We thank the participants for their time and interest in this project. We thankfully acknowledge the imaging resources and support provided by the "Laboratorio Nacional de Imagenología por Resonancia Magnética (LANIREM)" and the "Universidad Nacional Autónoma de México (UNAM)". We also thank Leopoldo González-Santos, and Erick H. Pasaye for their technical support.

Author contributions statement

V.E.O-M.: Methodology; Formal analysis; Investigation; Data Acquisition; Data Curation; Writing-Original Draft; Visualization. D.A-B.: Data Acquisition; Data Curation; Writing-Review & Editing. G.L-H: Data Acquisition; Data Curation; Methodology; Writing-Review & Editing. J.G-C.: Conceptualization; Methodology; Writing-Review & Editing. F.A.B: Methodology; Resources; Supervision; Funding acquisition, Writing-Review & Editing. P.H-C.: Conceptualization; Methodology; Conceptualization; Methodology; Writing-Review & Editing; Original Draft; Supervision; Project administration; Funding acquisition.

Funding

This research was funded by the project “Trampa, engaño y corrupción. Observaciones teóricas y experimentales” , institusionally unded by the “Dirección de Apoyo a la Investigación de Rectoría General of the Universidad Autónoma Metropolitana” (P.H.C.).

This work was also funded by DGAPA-PASPA 2023-24 and grant IN207923 from DGAPA-PAPIIT UNAM (F.A.B.).

Data availability statement (mandatory)
Access to the data can be obtained upon request to the corresponding author.

Additional information

Conflict of interest statement

F.A.B. receives compensation from Springer Nature for his work as editor. All the other authors have no conflict of interest to disclose.

References

1. Meisters, J., Hoffmann, A. & Musch, J. A New Approach to Detecting Cheating in Sensitive Surveys: The Cheating Detection Triangular Model. *Sociol Methods Res* 53, 328–368 (2024).
2. Kamalov, F., Sulieman, H. & Calonge, D. S. Machine learning based approach to exam cheating detection. *PLoS One* 16, (2021).
3. Hernández-Chávez, P. *et al.* Validity of an Instrument to Detect Cheating Confirmed by the Elicited Emotional Reactions. *Front Psychol* 12, (2021).
4. Saxe, R. & Kanwisher, N. People thinking about thinking people: The role of the temporo-parietal junction in ‘theory of mind’. *Neuroimage* 19, 1835–1842 (2003).
5. Frith, C. D. & Frith, U. The Neural Basis of Mentalizing. *Neuron* vol. 50 531–534 Preprint at https://doi.org/10.1016/j.neuron.2006.05.001 (2006).
6. Van Overwalle, F. Social cognition and the brain: A meta-analysis. *Hum Brain Mapp* 30, 829–858 (2009).

7. Abe, N., Greene, J. D. & Kiehl, K. A. Reduced engagement of the anterior cingulate cortex in the dishonest decision-making of incarcerated psychopaths. *Soc Cogn Affect Neurosci* 13, 797–807 (2018).
8. Karim, A. A. *et al.* The truth about lying: Inhibition of the anterior prefrontal cortex improves deceptive behavior. *Cerebral Cortex* 20, 205–213 (2010).
9. Pornpattananangkul, N., Zhen, S. & Yu, R. Common and distinct neural correlates of self-serving and prosocial dishonesty. *Hum Brain Mapp* 39, 3086–3103 (2018).
10. Hu, X., Pornpattananangkul, N. & Nusslock, R. Executive control- and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. *Cogn Affect Behav Neurosci* 15, 475–491 (2015).
11. Zheltyakova, M., Korotkov, A., Cherednichenko, D., Didur, M. & Kireev, M. To lie or to tell the truth? The influence of processing the opponent's feedback on the forthcoming choice. *Front Psychol* 15, (2024).
12. Dupont, L., Santangelo, V., Azevedo, R. T., Panasiti, M. S. & Aglioti, S. M. Reputation risk during dishonest social decision-making modulates anterior insular and cingulate cortex activity and connectivity. *Commun Biol* 6, (2023).
13. Panasiti, M. S. *et al.* The motor cost of telling lies: Electrocortical signatures and personality foundations of spontaneous deception. *Soc Neurosci* 9, 573–589 (2014).
14. Speer, S. P., Smidts, A. & Boksem, M. Cognitive control increases honesty in cheaters but cheating in those who are honest. *PNAS* 117, 19080–19091 (2020).
15. Speer, S. P., Smidts, A. & Boksem, M. A. S. Cognitive control promotes either honesty or dishonesty, depending on one's moral default. *Journal of Neuroscience* 41, 8815 (2021).
16. Speer, S. P., Smidts, A. & Boksem, M. A. S. Different Neural Mechanisms Underlie Non-habitual Honesty and Non-habitual Cheating. *Front Neurosci* 15, (2021).
17. van Lier, J., Revlin, R. & de Neys, W. Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module. *PLoS One* 8, (2013).
18. Cosmides, L., Barrett, H. C. & Tooby, J. Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences of the United States of America* vol. 107 9007–9014 Preprint at https://doi.org/10.1073/pnas.0914623107 (2010).
19. Bell, R. & Buchner, A. How Adaptive Is Memory for Cheaters? *Curr Dir Psychol Sci* 21, 403–408 (2012).
20. Ermer, E., Guerin, S. A., Cosmides, L., Tooby, J. & Miller, M. B. Theory of mind broad and narrow: reasoning about social exchange engages ToM areas, precautionary reasoning does not. *Soc Neurosci* 1, 196–219 (2006).
21. Liu, J. *et al.* The Difference Spotting Task: A new nonverbal measure of cheating behavior. *Behaviour Research Methods* 53, 1935–1944 (2021).
22. Greve D. N. Optseq Home Page. https://surfer.nmr.mgh.harvard.edu/optseq/ (2002).

23. Peirce, J. PsychoPy - Psychology software for Python. *Integration The Vlsi Journal* https://doi.org/10.1007/s11517-010-0657-3 (2009) doi:10.1007/s11517-010-0657-3.
24. Friston, K. *et al.* Spatial registration and normal- ization of images. *Hum Brain Mapp* 2, 165–189 (1995).
25. Worsley, K. J. *et al.* A General Statistical Analysis for fMRI Data. *Neuroimage* 15, 1–15 (2002).
26. Koster-Hale, J. & Saxe, R. Theory of Mind: A Neural Prediction Problem. *Neuron* vol. 79 836–848 Preprint at https://doi.org/10.1016/j.neuron.2013.08.020 (2013).
27. Schurz, M., Radua, J., Aichhorn, M., Richlan, F. & Perner, J. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews* vol. 42 9–34 Preprint at https://doi.org/10.1016/j.neubiorev.2014.01.009 (2014).
28. Saxe, R. & Powell, L. J. *It's the Thought That Counts Specific Brain Regions for One Component of Theory of Mind*. http://www.fil.ion. (2006).
29. Van Overwalle, F. & Baetens, K. Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage* vol. 48 564–584 Preprint at https://doi.org/10.1016/j.neuroimage.2009.06.009 (2009).
30. Keysers, C. & Gazzola, V. Chapter 21 Towards a unifying neural theory of social cognition. *Progress in Brain Research* vol. 156 379–401 Preprint at https://doi.org/10.1016/S0079-6123(06)56021-2 (2006).
31. Hare, T. A., Camerer, C. F., Knoepfle, D. T. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience* 30, 583–590 (2010).
32. Young, L. & Saxe, R. Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia* 47, 2065–2072 (2009).
33. Cavanna, A. E. & Trimble, M. R. The precuneus: A review of its functional anatomy and behavioural correlates. *Brain* vol. 129 564–583 Preprint at https://doi.org/10.1093/brain/awl004 (2006).
34. Kosslyn, S. M., Ganis, G. & Thompson, W. L. Neural foundations of imagery. *Nat Rev Neurosci* 2, 635–642 (2001).
35. Pearson, J., Naselaris, T., Holmes, E. A. & Kosslyn, S. M. Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends Cogn Sci* 19, 590–602 (2015).
36. Sabatinelli, D., Lang, P. J., Keil, A. & Bradley, M. M. Emotional perception: Correlation of functional MRI and event-related potentials. *Cerebral Cortex* 17, 1085–1091 (2007).
37. Hassabis, D. & Maguire, E. A. Deconstructing episodic memory with construction. *Trends Cogn Sci* 11, 299–306 (2007).
38. Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D. & Warren, D. E. Investigating the neural correlates of schemas: Ventromedial prefrontal cortex is necessary for normal schematic influence on memory. *Journal of Neuroscience* 35, 15746–15751 (2015).

39. O'Doherty, J. *et al.* Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science (1979)* 304, 452–454 (2004).
40. King-Casas, B. *et al.* Getting to Know You: Reputation and Trust in a Two-Person Economic Exchange. *Science (1979)* 308, 78–83 (2005).
41. Van Overwalle, F. *et al.* The role of the cerebellum in reconstructing social action sequences: A pilot study. *Soc Cogn Affect Neurosci* 14, 549–558 (2019).
42. Sokolov, A. A., Miall, R. C. & Ivry, R. B. The Cerebellum: Adaptive Prediction for Movement and Cognition. *Trends in Cognitive Sciences* vol. 21 313–332 Preprint at https://doi.org/10.1016/j.tics.2017.02.005 (2017).
43. Heleven, E., van Dun, K. & Van Overwalle, F. The posterior Cerebellum is involved in constructing Social Action Sequences: An fMRI Study. *Sci Rep* 9, (2019).
44. Schmahmann, J. D. The cerebellum and cognition. *Neuroscience Letters* vol. 688 62–75 Preprint at https://doi.org/10.1016/j.neulet.2018.07.005 (2019).
45. Binder, J. R. & Desai, R. H. The neurobiology of semantic memory. *Trends Cogn Sci* 15, 527–536 (2011).
46. Seghier, M. L. & Price, C. J. Functional heterogeneity within the default network during semantic processing and speech production. *Front Psychol* 3, (2012).
47. Binder, J. R., Desai, R. H., Graves, W. W. & Conant, L. L. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* 19, 2767–2796 (2009).
48. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. & Cohen, J. D. The Neural Bases of Cognitive Conflict and Control in Moral Judgment. *Neuron* 44, 389–400 (2004).
49. Mansouri, F. A., Tanaka, K. & Buckley, M. J. Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience* vol. 10 141–152 Preprint at https://doi.org/10.1038/nrn2538 (2009).
50. Sevinc, G. & Nathan Spreng, R. Contextual and perceptual brain processes underlying moral cognition: A quantitative meta-analysis of moral reasoning and moral emotions. *PLoS One* 9, (2014).
51. Liakakis, G., Nickel, J. & Seitz, R. J. Diversity of the inferior frontal gyrus-A meta-analysis of neuroimaging studies. *Behavioural Brain Research* vol. 225 341–347 Preprint at https://doi.org/10.1016/j.bbr.2011.06.022 (2011).
52. Diveica, V. *et al.* Graded functional organization in the left inferior frontal gyrus: evidence from task-free and task-based functional connectivity. *Cerebral Cortex* 33, 11384–11399 (2023).
53. Tyler, L. K. *et al.* Left inferior frontal cortex and syntax: Function, structure and behaviour in patients with left hemisphere damage. *Brain* 134, 415–431 (2011).
54. Ishkhanyan, B. *et al.* Anterior and Posterior Left Inferior Frontal Gyrus Contribute to the Implementation of Grammatical Determiners During Language Production. *Front Psychol* 11, (2020).

55. Yoshioka, A. *et al.* The Role of the Left Inferior Frontal Gyrus in Introspection during Verbal Communication. *Brain Sci* 13, (2023).
56. Menon, V. & Uddin, L. Q. Saliency, switching, attention and control: a network model of insula function. *Brain structure & function* vol. 214 655–667 Preprint at https://doi.org/10.1007/s00429-010-0262-0 (2010).
57. Funahashi, S. Prefrontal contribution to decision-making under free-choice conditions. *Frontiers in Neuroscience* vol. 11 Preprint at https://doi.org/10.3389/fnins.2017.00431 (2017).
58. Friedman, N. P. & Robbins, T. W. The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology* vol. 47 72–89 Preprint at https://doi.org/10.1038/s41386-021-01132-0 (2022).
59. Grossmann, T. Novel Insights into the Social Functions of the Medial Prefrontal Cortex during Infancy. *eNeuro* 12, (2025).
60. Grossmann, T. The role of medial prefrontal cortex in early social cognition. *Frontiers in Human Neuroscience* Preprint at https://doi.org/10.3389/fnhum.2013.00340 (2013).
61. Bicks, L. K., Koike, H., Akbarian, S. & Morishita, H. Prefrontal cortex and social cognition in mouse and man. *Frontiers in Psychology* vol. 6 Preprint at https://doi.org/10.3389/fpsyg.2015.01805 (2015).
62. Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4, 267–278 (2000).
63. Carter, R. M. K. & Huettel, S. A. A nexus model of the temporal-parietal junction. *Trends Cogn Sci* 17, 328–336 (2013).
64. Craig, A. dD. How do you feel — now? The anterior insula and human awareness. *Nat Rev Neurosci* 10, 59–69 (2009).
65. Gu, X., Hof, P. R., Friston, K. J. & Fan, J. Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology* 521, 3371–3388 (2013).
66. Balleine, B. W., Delgado, M. R. & Hikosaka, O. The role of the dorsal striatum in reward and decision-making. *Journal of Neuroscience* vol. 27 8161–8165 Preprint at https://doi.org/10.1523/JNEUROSCI.1554-07.2007 (2007).
67. Delgado, M. R. Reward-related responses in the human striatum. in *Annals of the New York Academy of Sciences* vol. 1104 70–88 (Blackwell Publishing Inc., 2007).
68. Liljeholm, M. & O'Doherty, J. P. Contributions of the striatum to learning, motivation, and performance: An associative account. *Trends Cogn Sci* 16, 467–475 (2012).

Table 1

Table 1 is available in the Supplementary Files section.

Figures

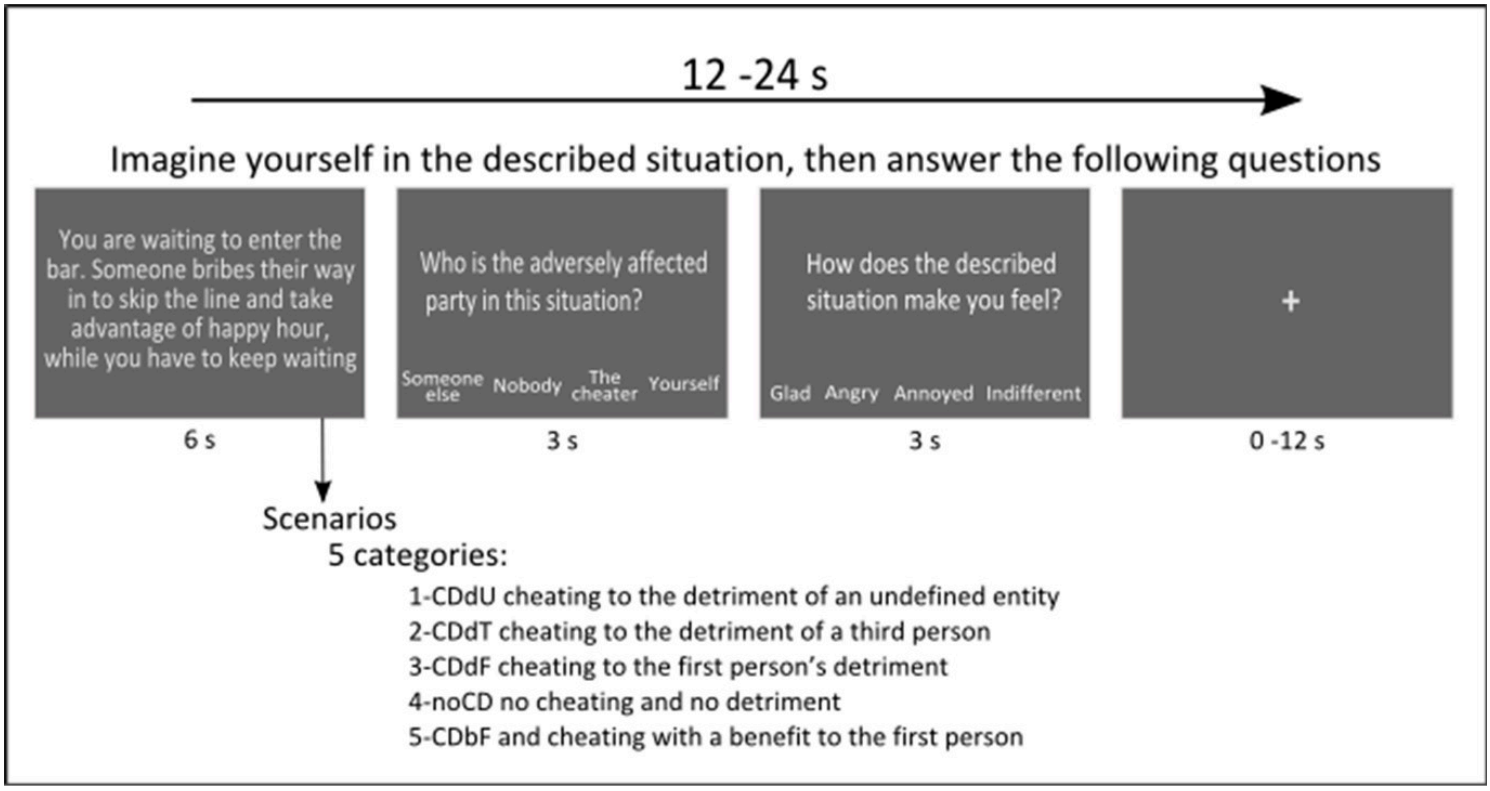


Figure 1

Example of an fMRI event stimuli. Participants read each scenario for 6 seconds after which they answered follow-up questions.

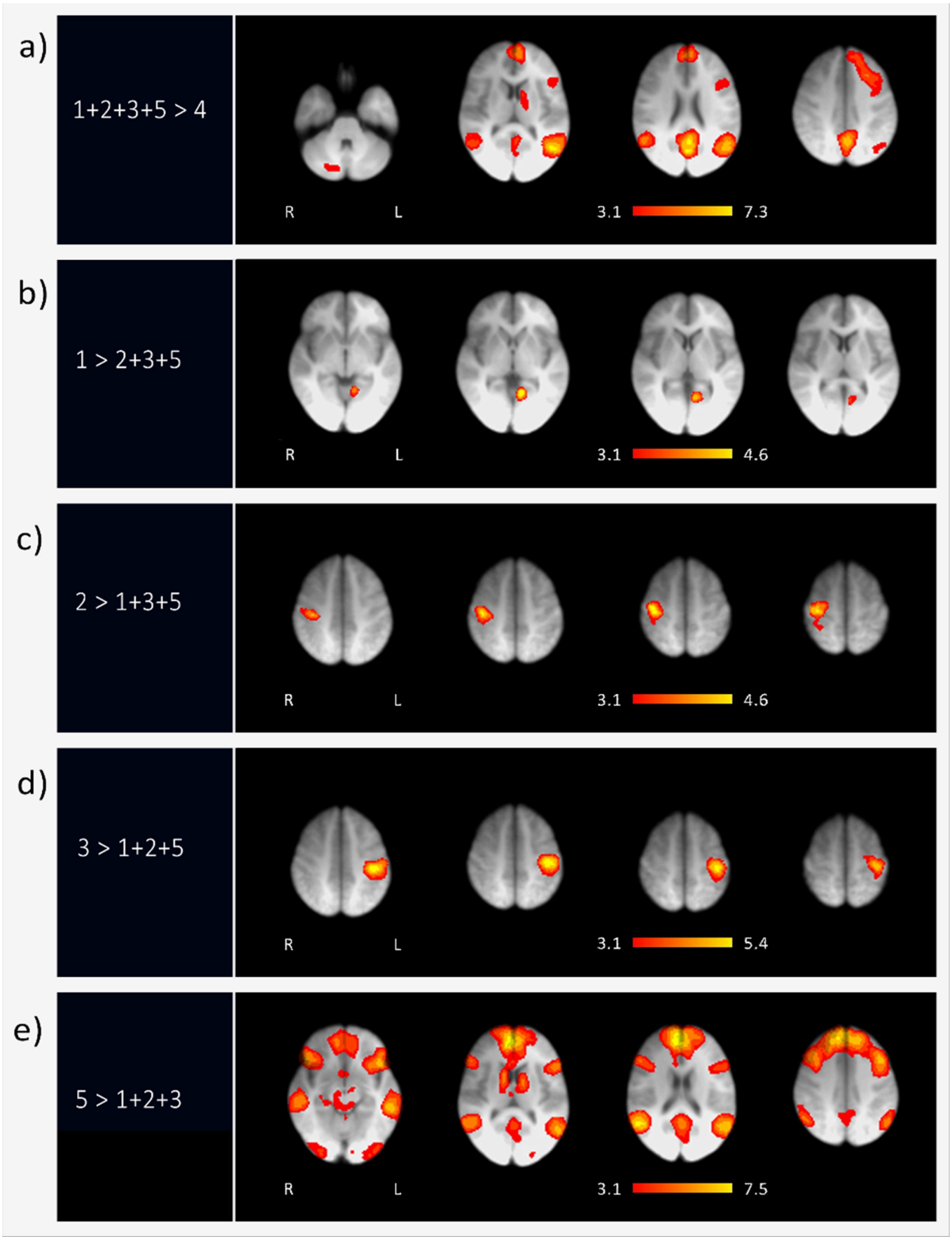


Figure 2

Whole-brain activation maps for the main contrasts across cheating scenarios (axial view)**. a)** 1+2+3+5 > 4: Increased activation in the left medial prefrontal cortex (mPFC), bilateral temporoparietal junction (TPJ), precuneus (Pcu), left caudate nucleus (lCN), left occipital pole (lOcPole), and right Crus I/II (rCrusI/II). **b)** 1 > 2+3+5: Activation in the precuneus and left lingual gyrus (lLing/Pcu). **c)**2 > 1+3+5: Activation in the right posterior middle frontal gyrus (rpMFG). **d)**3 > 1+2+5: Activation in the left inferior

frontal gyrus (lIFG). **e)** 5 > 1+2+3: Activity in the prefrontal cortex (PFC), superior parietal cortex/right TPJ (SPC/rTPJ), superior parietal and superior temporal sulcus/left TPJ (SPC/STS/lTPJ), bilateral inferior frontal gyrus/anterior insula (IFG/AI), bilateral dorsolateral prefrontal cortex (dlPFC/rdlPFC), middle temporal sulcus (mTS), precuneus/posterior cingulate cortex (Pcu/PCC), and bilateral putamen. Color bar indicates z-values. L: left hemisphere; R: right hemisphere.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuplementaryCheatingDetection.docx
- Table1.docx