**Description of the Study:**

Measurement of the determinants of socially undesirable behaviors, such as dishonesty, are complicated and obscured by social desirability biases. To circumvent these biases, we used connectome-based predictive modelling (CPM) on resting state functional connectivity patterns in combination with a novel task which inconspicuously measures voluntary cheating to gain access to the neurocognitive determinants of (dis)honesty. Specifically, we investigated whether task-independent neural patterns within the brain at rest could be used to predict a propensity for (dis)honest behaviour. Our analyses revealed that functional connectivity, especially between brain networks linked to self-referential thinking (vmPFC, temporal poles, and PCC) and reward processing (caudate nucleus), reliably correlates, in an independent sample, with participants’ propensity to cheat. Participants who cheated the most also scored highest on several self-report measures of impulsivity which underscores the generalizability of our results. Notably, when comparing neural and self-report measures, the neural measures were found to be more important in predicting cheating propensity.

**Neural Data:**

*DL\_NQ\_Overlap\_all.zip* – correlations between regions extracted from Dictionary Learning algorithm

*preproc\_files.zip* – preprocessed neural data, processed using the CONN pipeline (<https://web.conn-toolbox.org/>)

*Confounds\_extracted\_from\_CONN.csv* – Confounds extracted from the data via the CONN pipeline (<https://web.conn-toolbox.org/>)

*/different\_metrics*  - This folder contains the prediction performance for spearman as well as pearson correlation needed to create *Figure 7*.

**Scripts:**

*ScriptForRepo.ipynb* - Python scripts for the analysis reported in the manuscript (Organizing Data, Analysis and plotting)

*ScriptForRepo\_R.ipynb* - R scripts for the analysis reported in the manuscript (Organizing Data, Analysis and plotting)

*funcs.py*  - helper functions that make the code more concise, mainly for permutation testing

**Behavioral Data:**

*exp\_subs\_rename.csv* - csv files detailing which subjects belong to which sample

*Behavioral\_combined.csv* – csv containing all the behavioral data

*Q\_all\_rsfMRI\_recoded\_repo.csv –* Questionnaire data from all participants

*91subs\_names.csv –*  91 participants who filled out all questionnaires

**Preprocessing of neural data included data:**

Data were preprocessed using the standard pipeline of the CONN toolbox

(https://www.nitrc.org/projects/conn) in MATLAB. This pipeline included realignment of the

functional data using SPM12’s realign & unwarp procedure (Anderson et al., 2001), where all scans were coregistered and resampled to a reference image (first scan of the first session) using b-spline interpolation. Subsequently, outlier detection was performed from the observed global BOLD signal and the amount of subject-motion in the scanner. Acquisitions with framewise displacement above 0.9mm or global BOLD signal changes above 5 SD were marked as potential outliers. Framewise displacement was computed at each timepoint by considering a 140x180x115mm bounding box around the brain and estimating the largest displacement among six control points placed at the center of this bounding-box

faces. Afterwards, both the functional and the anatomical data were normalized into standard MNI space and segmented into grey matter, white matter, and CSF tissue classes using SPM12’s unified segmentation and normalization procedure (Ashburner and Friston, 2005). As a last step, functional data were smoothed using spatial convolution with a Gaussian kernel of 8mm full width half maximum (FWHM), in order to increase BOLD signal-to-noise ratio and reduce the influence of residual variability in functional and gyral anatomy across subjects. As a next step, denoising of the functional data was performed again using the standard pipeline from the CONN toolbox. For each participant, CONN implemented CompCor, a method for identifying principal components associated with segmented white matter (WM) and cerebrospinal fluid (CSF). In a first-level analysis, aCompCor components (Behzadi et al., 2007) and first-order derivatives of motion were entered as confounds and regressed from the BOLD signal. In addition, preprocessing steps included temporal band-pass filtering (0.008 Hz – 0.09 Hz), linear detrending, and regression of outlying

functional volumes (>97th percentile in normative sample; global-signal z-value threshold = 5, subjectmotion mm threshold = 0.09) identified using the artifact removal toolbox (ART) (https:// [www.nitrc.org/projects/artifact\_detect/](http://www.nitrc.org/projects/artifact_detect/)).

**Participants**

The reported analyses are based on 99 participants (65 females; 24 nationalities; age 18 - 43 years, *M* = 24.3, *SD* = 4.30) from four separate studies. The data were collected in three different scanners. We included data from four studies to increase sample size and diversity. Specifically, we collected data from two big and two small samples. This was done to obtain two independent big samples for the training and two independent samples for the test set. Having two independent samples in the training set increases the generalizability of the models as they are trained on a more diverse sample from different scanners and are less likely to learn idiosyncratic patterns related to a specific sample.The big samples were chosen as training set as they allow for the biggest number of observations to be used for feature selection and training of the models, which is necessary to ensure generalizability of the model and improve prediction accuracy (Sima et al., 2005; 2006).

The first sample of participants consisted of students (N = 40, 29 females; age 18 - 35 years, *M* = 23.8, *SD* = 3.25) from now on referred to as Sample 1. The second sample consisted of a general population sample from a different city and neural data was collected in a different scanner (N=41, 23 females; age 18 - 43 years, *M* = 24.8, *SD* = 5.4). The third sample were students at a different university in a different scanner (N=9, 7 females; age 19 - 24 years, *M* = 21.6, *SD* = 1.42). The fourth sample was scanned at the same scanner as Sample 1 but consisted of different students (N=9, 6 females; age 22 - 30 years, *M* = 26.9, *SD* = 2.6). Samples 1 and 2 will form the training set and samples 3 and 4 form the test set. Due to the fact that only 91 out of the 99 subjects completed all questionnaires (see below) all the analyses reported in the Results section focus on these 91 subjects, apart from the robustness checks that were done on all 99 subjects (see Results section).

We also assessed whether differences in demographics existed between the samples. Here we focused on the 91 subjects, since these were the ones that were used in all the analyses reported in the main text. Testing the differences in age between the four samples (on the 91 subjects) no significant differences were found (ANOVA, M1 = 23.7, M2 = 24.9, M3 = 21.9, M4 = 26.8, *F* = 2.23, *p* = 0.09). Testing for differences in Gender across the four samples (on 91 subjects) again no significant differences were found (*X2* = 3.91, *p* = 0.27). When comparing the training set (Samples 1 and 2) to the test set (Samples 3 and 4) again no significant differences were found for age (*t* = 0.04, *p* = 0.97) and gender (*X2* = 0.31, *p* = 0.58).

**Task**

*Spot-The-Difference Task*

In the Spot-The-Difference task, participants were presented with pairs of images and were instructed that there were always three differences present between the image pairs. Differences consisted of objects that were added to or removed from an image, or objects that differed in color between images. Participants were instructed that the purpose of the study was to investigate the underlying neural mechanisms of visual search for marketing purposes such as searching for a product in an assortment or information on a webpage. In order to increase credibility of this cover story, a simple visual search task was added at the beginning of the experiment (see Appendix 1). Further, participants were instructed that the neurocognitive effect of motivation, elicited by monetary reward, on speed and accuracy of visual search would be investigated. Although participants were told that there were three differences in all trials, in 25% of the trials there were only two differences and in 25% there was only one difference. Participants were requested to find three differences between the images. Since reward (see below) was contingent on participants *reporting* that they had found all three differences, without having to point them out, this design allowed and encouraged cheating behavior (i.e., reporting having found all three, even when objectively fewer than three differences were present in the images).

All stimuli were standardized in size and were presented on a white background on a computer screen. The ratio of 50% - 50% (three differences vs fewer than three differences) was chosen based on the results of pilot studies that indicated this ratio to be optimal in reducing suspicion that the pairs did not always contain three differences.

Trials were further categorized into normal (50%), hard (25%) and very hard trials (25%), for which participants could receive 5cts, 20cts, and 40cts, respectively. All the trials with three differences (the filler trials) were categorized as normal trials, whereas trials with less than three differences (the trials of interest) were randomly categorized as hard or very hard trials. Consequently, the reward was independent of the number of differences in the image pair for the trials of interest, which is important in order to be able to disentangle the effects of reward and cheating magnitude (the actual number of differences) on cheating behavior. The different levels of difficulty were added to reduce suspicion about the real purpose of the task. It was assumed that if trials are labeled as hard or very hard, it would be more credible to the participant that the image pair actually contained three differences, but they were just too hard to spot. In addition, levels of difficulty were introduced to eliminate possible demand effects: we wanted participants to cheat for monetary reward and not to prevent seeming incompetent, which may be associated with different underlying neural mechanisms and consequently confound the analysis.

To further reduce suspicion about the purpose of the study, approximately 10% of all trials were point-and-click trials. In these trials, participants had to click on the location in the images where they spotted the differences. Consequently, cheating was not possible on the point-and-click trials. Participants always knew prior to the start of a trial whether it was a point-and-click trial, indicated by a screen requesting participants to click on the image. This ensured that participants would not refrain from cheating on all other trials, while still reducing the suspicion about the real purpose of the study. Participants were told that only 10% of trials were point-and-click trials because it would take too much time to point out the differences for every pair. For the point-and-click trials 12 additional stimuli were used. All of the point-and-click trials had three differences. Six of the point-and-click trials were labeled as normal, three of them were labeled as hard and three were labeled as very hard. This was done to mirror the ratio of levels of difficulty in the regular trials. For the hard and very hard trials we used stimuli that during the pilot testing for the stimuli were too hard to be used as regular trials, that is trials with an average completion time (time to find all three differences) longer than 6s. In sum, there were 144 regular trials (of which 72 were cheatable trials) and 12 point-and-click trials. The maximum amount of money earned if a participant cheated on all cheatable trials was approximately 35 Euros, whereas if a participant did not cheat at all, he or she would earn approximately 7.50 Euros. After completion of the full study, participants were debriefed. To be fair to all participants, they were all paid the maximum amount irrespective of their actual cheating behavior.

Each trial started with a fixation cross which was presented for a variable amount of time between 1-3s (see Figure 1). Subsequently, the Level of Difficulty screen was presented for 2 seconds, informing the participants about the level of difficulty of the upcoming trial. This screen also displayed how much money could be earned on that trial. As a result, participants were aware of the potential gains of cheating. Subsequently, another jittered fixation cross appeared (jittered between 1-3s). Next, an image pair was presented for 6s, a length that was determined by behavioral pilot testing to be sufficient for most subjects to find all the differences (see Appendix 2), and participants engaged in the visual search (see Appendix 3 for examples of the image pairs). Afterwards, the participants were asked whether they spotted three differences (yes/no response). On this decision phase screen, again the potential reward for this trial was presented, to make the reward more salient and increase cheating behavior. After 3s, the response phase started in which participants’ responses were recorded. In the decision phase and the response phase the current balance was also shown, which was done to demonstrate to the participants that if they stated that they had found the three differences, their current balance increased immediately. It was assumed that this direct noticeable effect of behavior on the increase of the current balance would further motivate participants to cheat.

The buttons corresponding to “Yes” and “No” were switched across trials. Additional analysis confirmed that switching did not induce confounding effects (see Appendix 4). Once the participants responded, the choice was highlighted by a blue box for 500ms to indicate that the response was recorded, and the trial ended. If no response was made, the trial ended after 3s. In addition, there were five practice trials, in which participants could get acquainted with the task. Stimulus presentation and data acquisition was performed using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, [www.neurobs.com](http://www.neurobs.com)).

Diagram

Description automatically generated with medium confidence

*Figure 1.* One trial of the Spot-The-Differences paradigm. Participants viewed a screen indicating the difficulty and value of the trial, then the image pair appeared for six seconds and then participants had to indicate whether they spotted all three differences.

**Stimuli used in the task**

The stimuli used in this task can be found in a the repository of a related study:

[https://datarepository.eur.nl/articles/To\_cheat\_or\_not\_to\_cheat\_Cognitive\_control\_ processes\_override\_our\_moral\_default/12287807](https://datarepository.eur.nl/articles/To_cheat_or_not_to_cheat_Cognitive_control_%20processes_override_our_moral_default/12287807)

**Ethics**

The study was approved by the ERIM internal review board and was conducted according to the Declaration of Helsinki.