**Description of the Study:**

There is a long-standing debate regarding the cognitive nature of (dis)honesty: Is honesty an automatic response or does it require willpower in form of cognitive control in order to override an automatic dishonest response. In a recent study (Speer, Smidts & Boksem, 2020) we proposed a reconciliation of these opposing views by showing that activity in areas associated with cognitive control, particularly the inferior frontal gyrus (IFG), helped dishonest participants to be honest, whereas it enabled cheating for honest participants. These findings suggest that cognitive control is not needed to be honest or dishonest per se but that it depends on an individual’s moral default. However, while our findings provided insights into the role of cognitive control in overriding a moral default, they did not reveal whether overriding honest default behavior (nonhabitual dishonesty) is the same as overriding dishonest default behavior (nonhabitual honesty) at the neural level. This speaks to the question as to whether cognitive control mechanisms are domain-general, or may be context specific. To address this, we applied multivariate pattern analysis to compare neural patterns of nonhabitual honesty to nonhabitual dishonesty. We found that these choices are differently encoded in the IFG, suggesting that engaging cognitive control to follow the norm (that cheating is wrong) fundamentally differs from applying control to violate this norm.

**Neural Data:**

*nonhabitualHonesty.zip* – The beta maps for nonhabitual honesty

*nonhabitualDisHonesty.zip* – The beta maps for nonhabitual dishonesty

**Scripts:**

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

**Behavioral Data:**

*CC\_final.csv* – Cheatcount for each participant

**Processing of included data:**

The fMRI data was preprocessed using fMRIPrep version 1.0.8, a Nipype based tool (Gorgolewski et al., 2011). The reason for choosing fMRIPrep was that it addresses the challenge of robust and reproducible preprocessing as it automatically adapts a best-in-breed workflow to virtually any dataset, enabling high quality preprocessing without the need of manual intervention (65). Each T1w volume was corrected for intensity non-uniformity (INU) and skull-stripped. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (66) was performed through nonlinear registration, using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w. Fieldmap distortion correction was performed by co-registering the functional image to the same-subject T1w image with intensity inverted (67) constrained with an average fieldmap template (68). This was followed by co-registration to the corresponding T1w using boundary based registration (69) with 9 degrees of freedom. Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1wto-template (MNI) warp were concatenated and applied in a single step using Lanczos interpolation. Physiological noise regressors were extracted applying CompCor (70).

Principal components were estimated for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). Six tCompCor components were then calculated including only the top 5% variable voxels within that subcortical mask. For aCompCor, six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (71) was calculated for each functional run using the implementation of Nipype. For more details of the pipeline see <https://fmriprep.readthedocs.io/en/latest/workflows.html>

The preprocessed data and more information about processing of the neural data can be found in 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)

**Participants**

The reported analyses are based on 40 participants (30 females; age 18-35 years; *M* = 23.7 *SD* = 3.2) recruited from an online community for university students, where students can sign up for experiments. An intial screening interview ensured that all participants were right-handed with normal or corrected to normal vision, spoke English fluently, were not on any psychoactive medication influencing cognitive function and had no record of neurological or psychiatric illness.

**Task**

*Spot-The-Difference Task*

In the Spot-The-Difference task, participants were presented with pairs of images and were told that there were always three differences between the image pairs. Differences consisted of objects that were added to or removed from an image, or objects that differed in colour between images. However, images could actually contain one, two, or three differences. Participants were asked to find three differences between the images. Because reward (see below) was contingent on participants *reporting* that they had found all three differences, without having to point them out, this design encouraged cheating behavior (i.e., reporting having found all three, even when objectively fewer than three differences were present in the images).

Participants were told 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 SI Appendix 1), which was also performed in the scanner while participants were undergoing localizer scans. Further, participants were instructed that the neurocognitive effect of motivation, elicited by monetary reward, on speed and accuracy of visual search was 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. 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 less 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 of 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 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 using a joy-stick. As a consequence, 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. Further, participants were instructed that excessive movement by manipulating the joystick would interfere with the brain signal. In sum, there were 144 regular trials (of which 72 cheatable trials) and 12 point-and-click trials. The maximum amount of money earned, in case a participant cheated on all cheatable trials was approximately 35 Euros, whereas in case a participant would not cheat at all he or she would earn approximately 7.50 Euros. After completion of the full study, participants were debriefed that the purpose of the study was to investigate the underlying neural mechanisms of (dis)honest decision-making. They were informed that the number of differences between pictures and level of reward were manipulated to encourage cheating. To be fair to all participants, they were all paid out the maximum amount, irrespective of their actual cheating behavior. In addition, participants received a flat fee of 10 Euros for participation in the scanning session.

Each trial started with a fixation cross which was presented for a variable amount of time between 1-3s (see Figure 8). 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 constantly aware of the potential gains of cheating. Next, an image pair was presented for 6s, a length determined by the behavioral pilots, and participants engaged in the visual search. Afterwards, the participants were asked whether they spotted all three differences (yes/no response). On this decision phase screen, again the potential reward for this trial was presented, in order 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 decision phase and response phase were separated to isolate the decision from motor responses. This was important for the fMRI analysis as we wanted to isolate the neural mechanisms underlying decision making from possible neural confounds related to button presses. Besides that, the buttons corresponding to “Yes” and “No” were switched across trials to further reduce confounding effects and to reduce the response bias for the dominant hand. 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 behavioral data acquisition was performed using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, [www.neurobs.com](http://www.neurobs.com)).

The main advantage of our experimental design is that it allowed tracking on which trials the participants cheated. As we knew how many differences there are in each image pair, we knew precisely whether the participants cheated or not. Further, by varying the number of differences, this design enabled us to assess the magnitude of cheating (i.e, cheating when only 1 vs 2 differences were found). It is therefore the first behavioural paradigm that allowed to assess cheating behaviour inconspicuously on a trial-by-trial basis in the scanner.



*Figure 1.* One trial of the Spot-The-Differences paradigm. Participants view a screen indicating the difficulty and value of the trial, then the image pair appears for six seconds and then participants have to indicate whether or not they spotted all three differences.

**Stimuli used in the task (***Images\_SpotTheDifferenceTask.zip***)**

Stimuli for the task consisted of 144 Spot-The-Difference image pairs that were downloaded from the Internet. Cartoon images of landscapes containing several objects were selected, to make them engaging and challenging enough for the participants. Landscapes were chosen as they generally satisfied the necessary criterion of containing several different objects. The stimuli consist of pairs of images that are identical apart from a certain number (1-3) of differences that were created using Adobe Photoshop. Differences consisted of objects added to or removed from the landscape picture or changed colors of objects. Differences were fully randomized across all pairs of images, which means that all image pairs could be presented with either one, two or three differences. To make sure that participants would be able to find the differences between the images in a reasonable amount of time and to minimize the chance of participants believing that they had seen a difference when they had not (false positives), we ran a pilot study on Amazon’s Mechanical Turk (N=205) to test the difficulty to spot the differences between the images and to determine the optimal duration of picture presentation (see SI Appendix 2).

**Ethics**

The study was approved by the ERIM internal review board and was conducted according to the Declaration of Helsinki. The Medical Ethics Committee of the Erasmus MC approved this study.

**Privacy protocols**

In order to keep the data anonymous only skull-stripped and normalized funtional scans have been uploaded.