**Description of the Study:**

Cognitive control is crucially involved in making (dis)honest decisions. However, the precise nature of this role has been hotly debated. Is honesty an intuitive response or is willpower needed to override an intuitive inclination to cheat? A reconciliation of these conflicting views proposes that cognitive control enables dishonest participants to be honest, whereas it allows cheating for those who are generally honest. Thus, cognitive control does not promote (dis)honesty per se; it depends on one’s moral default. In the present study, we tested this proposal using EEG in humans (males & females) in combination with an external localizer task to mitigate the problem of reverse inference. Our analysis revealed that the neural signature evoked by cognitive control demands in the Stroop task can be used to estimate (dis)honest choices in an independent cheating task, providing converging evidence that cognitive control can indeed help honest participants to cheat, whereas it facilitates honesty for cheaters.

**Neural Data:**

*Raw\_EEG.zip* – The raw EEG data from the 35 participants collected

*Stroop\_prepro\_repo.zip* – The preprocessed Stroop Data

*SPD\_prepro\_TFR\_repo.zip* – The preprocessed Data for the Spot-The-Difference task for TFR analysis

*SPD\_prepro\_TbT\_repo.zip* – The preprocessed Data for the Spot-The-Difference task for Similarity analysis

*Trialbytrial\_Power\_FactorAnalysis\_final.csv* – Trial by trial power for the channels of interest used for multilevel modelling

*Trialbytrial\_SimilarityBasedForRev\_repo.csv* – Trial by trial similarity between Stroop and SPD for the channels of interest used for multilevel modelling

**Scripts:**

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

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

**Behavioral Data:**

*Logfiles\_repo.zip* – The logfiles containing all the behavioral data

*CC\_all\_repo.csv* – CheatCount for each subject

*SPD-EEG-repo.csv –* Trial by trial task and behavior info

**Processing of included data:**

We recorded EEG from 64 active scalp electrodes using a Biosemi Active Two system (Biosemi, The Netherlands). Additional flat type electrodes were placed on the right and left mastoid, and in the eye region in order to record eye movements or electro-oculograms (EOG): Electrodes were placed below and above the left eye in line with the pupil to record vertical EOG, and at the outer canthi of both eyes to record horizontal EOG. The EEG and EOG signals were sampled at a rate of 512 Hz. All preprocessing was done using the MNE package in python (Gramfort et al., 2013). EEG data were filtered with a low cutoff filter of 1Hz to remove slow drifts and a notch filter of 50Hz to remove line noise.

Subsequently, bad and noisy channels were detected using several different approaches as implemented in the PREP pipeline (Bigdely-Shamlo et al., 2015): First, by means of correlation: Here we checked how well a given channel is correlated with all other channels (categorize as bad at *r*  < 0.4). Second, by using the robust z-score deviation aggregates per channel (categorize as bad at *z*  > 5). Third, by using the robust z-score estimates of high frequency noise per channel (categorize as bad at *z*  > 5). Lastly, by using the RANSAC channel correlations, which is the correlation for each channel with itself across the original data versus the RANSAC predicted data (categorize as bad at *r*  < 0.75), as implemented in the PREP pipeline (Bigdely-Shamlo et al., 2015). After detection, these channels were removed from the data and subsequently interpolated (i.e., estimated from surrounding channels). Interpolation was performed using the spherical spline method (Perrin et al., 1989) as implemented in MNE, which projects the sensor locations onto a unit sphere and interpolates the signal at he channels identified as bad on the signals for the good channels. The EEG data were then re-referenced to the average signal across channels. As a next step, ocular artefacts were removed by performing an independent component analysis (ICA) on the data and then correlating the resulting components with the EOG channels to see which of the components represents the ocular artefacts. The component that correlated the highest with the EOG channels was then removed from the EEG data.

*Epoching & Artefact Rejection for the Spot-The-Difference Task*

The EEG data from the Spot-The-Difference-Task was then segmented into 3 second epochs, time-locked to the onset of the decision phase. The epochs were baseline corrected using the last second of the fixation period preceding the presentation of the image pair, which occurred seven to six seconds before the decisions phase. The resulting epochs were then subjected to *Autoreject*, an automated artefact detection algorithm based on machine learning classifiers and cross-validation to estimate the optimal peak-to-peak threshold (Jas et al., 2017). On average 3% of trials (~4 trials out of 144 trials, SD=5%) were rejected. This algorithm was implemented to remove artefacts not identified by previous preprocessing steps and, depending on the number of bad sensors for a given trial, either repairs the trial based on interpolation or excludes it from further analysis. The preprocessed data were then submitted to a morlet wavelet analysis to transform the data into the time-frequency domain with 18 logscaled frequency bins ranging from 4Hz to 40Hz in order to have higher sensitivity in lower frequency ranges such as the theta band. To optimize both spectral and temporal resolution, the number of cycles to include in the sliding time window were defined by dividing each individual frequency by two. After transforming the data to the time-frequency domain the data were decimated by a factor of 4 (sampling every 4th timepoint) to increase computational efficiency.

*Epoching & Artefact Rejection for the Stroop*

Following the same preprocessing steps as described above, the EEG data from the Stroop task was segmented into two second epochs time-locked to the presentation of second word on the screen. The epochs were baseline corrected using the second preceding the onset of the second word. The same artefact rejection as described above was applied to the Stroop task and in this task on average 2% (SD = 4%) of trials for each subject was rejected. The same morlet wavelet analysis was used to transform the data to the time frequency domain.

**Participants**

EEG recordings were obtained from 35 participants. One participant completed the Spot-The-Difference task, but not the Stroop task, so for that task there were 34 participants. The data of 2 participants had to be discarded because for these participants 27 or more channels (out of 64) were identified as bad channels by the Autoreject algorithm (see preprocessing section; Mainak et al., 2017), which classified them as outliers (IQR rule, 2 standard deviation rule). The reported analyses are based on the remaining 33 (32 for the Stroop task) participants (18 females; age 18 to 29; M = 21, SD = 2.6), recruited from an online community for university students, where students can sign up for experiments. An initial 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. The study was approved by the University’s internal review board and was conducted according to the Declaration of Helsinki.

**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 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. However, images could actually contain one, two, or three differences. Participants were instructed 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). 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 the trials with three differences (the filler trials) were categorized as normal trials, whereas trials with fewer 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. 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 and to be fair to all participants, they were all paid out the maximum amount irrespective of their actual cheating behavior.

Participants were informed 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 Speer, Smidts & Boksem, 2020). Further, participants were instructed that the neurocognitive effect of motivation, elicited by monetary reward, on speed and accuracy of visual search would be investigated. To further reduce suspicion about the purpose of the study, we added twelve point-and-click trials. In these trials, participants had to click on the location in the images where they spotted the differences using a mouse. 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. In sum, there were 144 regular trials (of which 72 cheatable trials) and 12 point-and-click trials.

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 constantly aware of the potential gains of cheating. Next, an image pair was presented for 6s, a length determined by behavioral pilots (see Speer, Smidts & Boksem, 2020), 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 buttons corresponding to “Yes” and “No” were switched across trials to reduce a possible response bias associated with the side at which the response options were presented. 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)).



*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 or not 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.