The Version of Record of this manuscript has been published and is available in Social Neuroscience, 2019, http://www.tandfonline.com/10.1080/17470919.2019.1681507

I'll make you an offer you can refuse: How socially sensitive is the MFN?

Michiel M. Spapé[1, 2] & Kimberley Dundas[2, 3]

1. Department of Psychology & Logopedics, University of Helsinki, Finland; michiel.spape@helsinki.fi
2. Department of Psychology, Liverpool Hope University, United Kingdom
3. Department of Psychology, Keele University, United Kingdom

Medial frontal negativity (MFN) is an event-related potential thought to originate in the anterior cingulate cortex. It is evoked by outcomes being worse than expected, such as when presented with unfair economic proposals during the Ultimatum Game (UG). This could mean the MFN indexes a social-emotional response, as commonly suggested in accounts that relate it to a violation of a social norm of fairness. To examine the link between MFN and norm violation, we designed an EEG experiment with participants acting as representatives in an UG. Participants responded either as themselves, or as representatives of two charities. Of these, a norm-compatible charity conformed to the participant's values, while the norm-incompatible charity contrasted to them. The behavioural results showed that norm-incompatible representation reversed behaviour, with almost all fair offers being declined. The MFN, however, was unaffected by the norm representation, with unfair offers consistently evoking MFNs across conditions. We furthermore replicated the curious finding that unexpectedly generous offers evoke as much MFN as unfair offers. Thus, the MFN is not nearly as sensitive to higher-order social-emotional processes as commonly assumed. Instead, the perceived inequality that drives the MFN likely reflects a rational, probabilistic process.

Keywords: Medial frontal negativity; EEG; ERP; economic decision-making

Introduction

Successful cooperation requires swift behavioural adjustments to social feedback and perceived social slights. In the literature on human performance, this is commonly seen as enabled by an early response to outcomes that are detected as "worse than expected", such as when making an error during a task[1–3]. The electrophysiological correlate to this response is known collectively as the family of medial frontal negativities (MFNs[4]) and involves an anterior negativity, ca. 200 ms after errors (error related negativity; ERN[5]), negative feedback (feedback related negativity; FRN[6]), and problematic decision making scenarios (medial frontal negativity; MFN singular[7]). As we define errors through social learning[8], it makes intuitive sense that what counts as an error should be mediated by its social context, and MFNs could therefore represent socially sensitive ERPs.

The literature on social decision-making games generally supports this assertion. For example, in the Ultimatum Game (UG)[9], participants respond to fair and unfair shares of a set amount of money by either accepting an offer (and both the participant and the proposer get their share) or declining it (and neither gets anything). Behaviourally, this leads to economically costly rejections of unfair offers, implying irrational motives as the outcome-maximizing decision is to accept all offers[10]. In fMRI studies, the scenario was found to activate areas related to cognitive control[11], which was understood to regulate the conflict between financial and emotional motives. Likewise in EEG research, Boksem & De Cremer[7] showed unfair offers evoked MFN, which was found correlated with the sense of participants' justice sensitivity. Accordingly, they, and others who appreciate the social role of the MFN[12–14], inferred that the MFN responds to unfairness, an erroneous violation of social norms.

Yet, the variety of paradigms employing MFNs may hide the inconsistence of the specific MFN pattern across studies, and the uncertainty to which social-emotional processes contribute to changes in the MFN. For example, the first UG study reporting negativity in response to offer perception [15] showed negativity to *mid-value* but not unfair offers (their fig. 3). A more recent study replicated unfair MFN but showed this to be indistinguishable from unexpectedly *generous* offers[16]. Beyond the general sensitivity of the MFN, its social specificity was put to doubt by work suggesting social cues that affected decision making behaviour neither modulating MFN[16] nor FRN[17]. Likewise, while observing errors in others evokes socially-induced ERNs, this effect itself was found socially insensitive, being independent of competitive context or rewarding outcome[18]. Alternative explanations may account for lack of psychophysiological correlates of socially salient effects, but the most parsimonious lacks prominence: The MFN is not as sensitive to higher-order social emotional processes as previously suggested.

In the present study we used the UG to investigate the extent to which the MFN is socially sensitive. Contrary to previous studies that investigated the MFN in relation to abstract traits and subtle social-emotional cues[7,17,19], we manipulate a factor that should directly, maximally affect the MFN's social context. Participants played the UG as representatives of charities that either agreed (norm-compatible representation) or contrasted (norm-incompatible representation) with their own values. If the MFN reflects an emotional response to problematic outcomes, then incompatible norm representation should elicit mass rejection of offers. Furthermore, if the MFN causes the UG response pattern, we should predict norm incompatibility to eliminate the MFN. In other words, the MFN as a socially sensitive ERP predicts a strong interaction between norm representation and fairness.

Methods[1]

Twenty healthy, unpaid volunteers (8 females, age 32.2±13.9 years) participated in the experiment.

Stimuli

Participants were shown seven fictional charities and asked their financial support on a Likert-type scale. The single most and least appealing charities were used for norm compatible and norm incompatible conditions. Emotional involvement with these was validated after the experiment. Participants supplied their own photo for use in self conditions. Pictures of proposers were obtained from other students.

Procedure

After receiving instruction and signing informed consent, participants were informed on whose behalf they were responding (norm representation), during the block of 60 trials that followed. Trials started with the total amount to be divided shown between the proposer and responder's logo (see figure 1). Following a 500 ms cue, the offer was shown as numbers separated by a horizontal line. 1000 ms later, participants were to respond to offers by pressing Y (accept) or N (reject) on the keyboard. An animation followed to show the decision consequence; the two values moving to each party or disappearing.

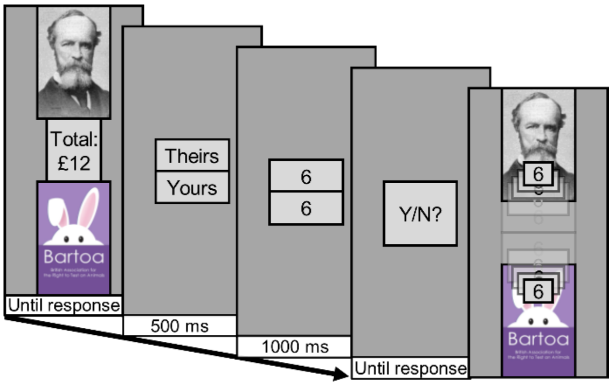

Figure 1: Trial procedure. A stranger divides an offer of £12 equally between themselves and the participant, who represents fictional charity *Bartoa* (British Association for the Right to Test On Animals). Responding results in an animation of the money moving towards each party.

Design

The experiment used a within-subject design with *norm representation* (compatible, incompatible, self) in counter-balanced order between 2 series of 3 blocks. Within each block of 60 trials, the 3 *fairness* conditions (unfair: <50% for the participant, fair, and generous: >50% for the participant) were presented orthogonally with 2 total sums of shares (£12 or £20)[2] and repeated 10 times in random order. The analysis used a repeated measures design with *norm representation* and *fairness* as factors. For behaviour, we analysed the acceptance rate whereas for EEG, we analysed the MFN, additionally including electrode (FCz, Cz, CPz) as factor. All ANOVAs were conducted with Greenhouse-Geisser adjustments where assumptions of sphericity were violated.

[1] Additional details on the methods are provided as supplemental materials 1

[2] We had no a-priori hypothesis regarding differences between sums of shares. However, an analysis exploring potential differences in the *stake* of the game is presented in the Supplementary material.

EEG

EEG activity was digitised at 1024 Hz from 64 scalp locations using active electrodes. Electro-oculographic activity was recorded using bipolar electrodes, one set placed lateral to the eyes, the other inferior and superior to the right eye. EEG analysis included average referencing, filtering (1-80 Hz BP), and independent component analysis. Following visual inspection for suspect components, source-level artefact-corrected EEG was reconstructed by applying weights of artefact-free components to the unfiltered data. Data were then further filtered (0.2-40 Hz) and segmented into 1200 ms epochs, time-locked to the offer and including 200 ms of baseline. An individually specified artifact rejection procedure further removed epochs based on thresholds. On average, the resulting procedure removed 19.01±5.12 % of epochs from analysis, which computed ERPs over the remaining 32.39 epochs per design cell.

Initial visual inspection of the MFN via grand average ERPs of fair and unfair conditions showed maximal negativity over the medial frontocentral to centroparietal cortex (Figure 3). To define the interval, we calculated the standardised difference between unfair and fair conditions, observing significant negativity from 290-470 ms. The MFN was measured as the mean activity within this window.

Results

Behaviour

A 4 x 3 repeated-measures ANOVA with *fairness* (very unfair, unfair, fair and generous) and *norm representation* (norm compatible, norm incompatible and self) on *acceptance* (percentage of offers accepted) showed significant main effects of *norm representation*, $F(1.22, 23.21) = 56.09$, $p < .0001$, $\eta_p^2 = .75$, and offer fairness, $F(1.53, 29.13) = 18.73$, $p < .0001$, $\eta_p^2 = .50$ respectively. *Norm representation* also significantly interacted with *fairness*, $F(2.30, 43.77) = 38.95$, $p < .0001$, $\eta_p^2 = .67$. Post-hoc comparisons showed all *norm representations* to significantly differ, $ps < .031$, with most offers accepted in the compatible, and least in the incompatible, condition. Contrasts between each level of offer fairness with its preceding level showed significant increases until the fair offer condition, *ps* < .002. The interaction was further analysed by applying the same contrasts to each *norm representation*. This showed a similar pattern for self and compatible conditions: increased acceptance with fairness until the fair ceiling level at, *ps* < .0007. Conversely, incompatible conditions revealed *decreasing* acceptance for unfair vs very unfair offers, $p < .004$, after which a floor level was reached. Figure 2 illustrates the reversed effect in norm incompatible conditions.

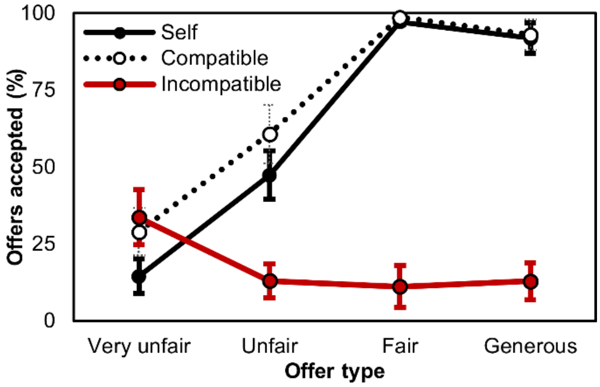

Figure 2: Acceptance rates as a function of offer types and norm representation. Error bars provide standard errors of means.

MFN

In repeated measures ANOVA on the average evoked potential with electrode (FCz, Cz, CPz), norm representation (*compatible, incompatible, self*), and offer fairness (unfair, fair, generous) as factors, neither *electrode* nor *norm representation* had a significant main effect, *ps* > .20. Fairness, however, did significantly affect the ERP, $F(2, 38) = 17.58$, $p < .0001$, $\eta_p^2 = .48$. Planned comparisons on the effect revealed this was due mainly to both unfair, $p < .0003$, and generous, $p < .002$, evoking negativity relative to fair offers, with no difference between unfair and generous offers being observed, $p = .48$. The predicted interaction[3] between *norm*

[3] Exploratory analyses on this interaction in addition to the MFN are provided as supplemental material 2.

representation and *fairness* was not observed, $F(3.26, 61.97) = 1.25$, $p = .30$, $\eta_p^2 = .06$, nor any other effect, *ps* > .18.

To determine whether the data favours the hypothesis that *norm representation* does not interact with *fairness*, we ran a full factorial Bayesian Repeated Measures analysis with the same factors. A null-model was defined as including only all main effects, which showed good fit, $BF_m = 96.69$, P(M|data) = .92. In contrast, low odds were observed for the alternative model that included the *norm representation* * *fairness* interaction, $BF_m = 0.39$, $BF_{10} = 0.050$, P(M|data) = .047 (P(M) = .111). In other words, the data show strong evidence *against* the interaction.

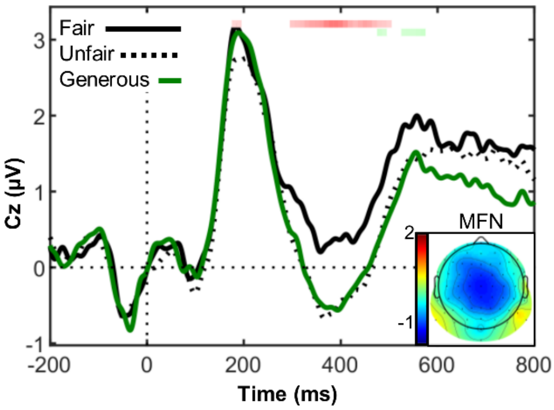

Figure 3: Event-related potential (ERP) to fair, unfair and generous offers. The evoked activity over Cz is shown, time-locked to the offer onset, averaged across social contexts. The displayed topography shows the subtraction of unfair from fair offers between 290 and 470 ms. Red dots to the top of the ERP indicate post-hoc comparisons between fairness conditions[2].

Discussion

Norm compatibility dramatically affects how we behaviourally respond to unfairness. While normally, unfair offers are accepted far less often than fair offers[10], but representing incompatible charities reversed this pattern. For example, people would almost never accept offers made to unpopular charities (e.g. Educate 'n' Arm), although they might accept the most unfair ones. Why would one accept just very unfair offers? Informal interviews with participants suggest participants favoured inconsequential donations to such charities if it meant a stranger would strongly benefit.

If this behaviour is a response to our provoked sense of justice, as indexed by the MFN, one would expect strong effects of norm-representation on the ERP. However, the MFN seemed insensitive to the behaviourally dramatic effect. Whether the participant themselves, a supported charity, or a loathed organisation such as Educate 'n' Arm was unfairly treated did not affect the MFN. It may therefore be that the MFN reflects an automatic effect in which all observed errors are treated alike; whether the social context concerns competitors, collaborators, or as here, antagonists, they are all socially equal. It may also be that the MFN is not socially sensitive.

Further dramatic evidence for this assertion is the observation that the MFN does not distinguish between unfair and generous offers. Are positive surprises and negative outcomes both violations of social norms[6]? They both signify unfairness, which is a social inference based on previous interaction experience. Yet, the present study calls into question the social-emotional quality of the response if it does not discern between hedonistically opposing offers. Rather than a violation of social norms, it may then be more neutrally described as an unequal response.

A more parsimonious alternative is that the MFN is not sensitive to top down control, and occurs as a bottom-up response to a low-level mismatch between features[20]. For example, in the present scenario, unfair offers were presented by unequal numbers, which necessarily involves complexity, while fairness has symmetrical features. Another possibility is that based on the UG itself, every offer size is imaginable, but fewer fair than unfair permutations exist. This probabilistic property makes it highly likely that responders would expect fair offers and consequently form an advance visual expectation.

That such lower-level considerations could account for effects on the MFN represents a serious problem for the state of the field. This is the case even if social cues and correlations with social traits have been observed, as a wealth of possibilities are open explanations if the MFN is not a-priori seen as a biomarker for social norm violations. In light of this, we urge the field to rule out all physical and statistical differences between conditions before considering a social account of their result. In the absence of such studies, we consider the MFN insensitive to social context.

The Version of Record of this manuscript has been published and is available in Social Neuroscience, 2019,
http://www.tandfonline.com/10.1080/17470919.2019.1681507

References

1. Dehaene, S., Posner, M. I. & Tucker, D. M. Localization of a neural system for error detection and compensation. *Psychol. Sci.* **5**, 303–305 (1994).
2. Gehring, W. J. & Willoughby, A. R. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* **295**, 2279–2282 (2002).
3. Van der Veen, F. M. & Sahibdin, P. P. Dissociation between medial frontal negativity and cardiac responses in the ultimatum game: effects of offer size and fairness. *Cogn. Affect. Behav. Neurosci.* **11**, 516–525 (2011).
4. Gehring, W. J. & Willoughby, A. R. Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. *Errors Confl. Brain Curr. Opin. Perform. Monit.* 14–20 (2004).
5. Falkenstein, M., Hohnsbein, J., Hoormann, J. & Blanke, L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr. Clin. Neurophysiol.* **78**, 447–455 (1991).
6. Hajcak, G., Moser, J. S., Holroyd, C. B. & Simons, R. F. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biol. Psychol.* **71**, 148–154 (2006).
7. Boksem, M. A. & De Cremer, D. Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* **5**, 118–128 (2010).
8. Ihssen, N., Mussweiler, T. & Linden, D. E. J. Observing others stay or switch – How social prediction errors are integrated into reward reversal learning. *Cognition* **153**, 19–32 (2016).
9. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388 (1982).
10. von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior*. (Princeton University Press, 1944).
11. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755–1758 (2003).
12. Hajcak, G., Moser, J. S., Yeung, N. & Simons, R. F. On the ERN and the significance of errors. *Psychophysiology* **42**, 151–160 (2005).
13. Luu, P., Collins, P. & Tucker, D. M. Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J. Exp. Psychol. Gen.* **129**, 43 (2000).
14. Cavanagh, J. F. & Allen, J. J. Multiple aspects of the stress response under social evaluative threat: An electrophysiological investigation. *Psychoneuroendocrinology* **33**, 41–53 (2008).
15. Polezzi, D. *et al.* Mentalizing in economic decision-making. *Behav. Brain Res.* **190**, 218–223 (2008).
16. Spape, M., Harjunen, V., Ahmed, I., Jacucci, G. & Ravaja, N. The semiotics of the message and the messenger: How nonverbal communication affects fairness perception. *Cogn. Affect. Behav. Neurosci.* 1–14 (2019).
17. Spapé, M. M., Hoggan, E. E., Jacucci, G. & Ravaja, N. The meaning of the virtual Midas touch: An ERP study in economic decision making. *Psychophysiology* **52**, 378–398 (2015).
18. De Bruijn, E. R. & von Rhein, D. T. Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Front. Neurosci.* **6**, 8 (2012).
19. Ma, Q., Hu, Y., Jiang, S. & Meng, L. The undermining effect of facial attractiveness on brain responses to fairness in the Ultimatum Game: an ERP study. *Front. Neurosci.* **9**, 77 (2015).
20. Firestone, C. & Scholl, B. J. Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behav. Brain Sci.* **39**, (2016).

Acknowledgements

The authors would like to thank Lisa Rosenberger for helpful comments on our Bayesian analysis.

Open data

Experiment code, EEG data, and analysis code available at OSF.IO/S2KMX