J Cogn Enhanc (2017) 1:26–30
DOI 10.1007/s41465-017-0009-4

OPINION

The Enhancement of Social Norm Compliance: Prospects and Caveats

Claudia Civai[1] · **Ili Ma**[2]

Received: 20 September 2016 / Accepted: 31 January 2017 / Published online: 27 February 2017

Abstract Societies are characterized by a shared system of social norms, which promotes cooperation among people. However, following social norms often means going against self-interest—imagine, for example, being required to choose whether or not to get richer from an unfair deal; ignoring social norms, on the other hand, may elicit disruptive antisocial behaviors that damage human relationships. Therefore, this type of value-based decisions is particularly tough and requires a complex trade-off between self- and other-regarding motivations. The advancement in cognitive neuroscience has shed light on the mechanisms underlying social norm compliance, describing the interplay between the emotional, reward, and self-control systems in shaping social norm preference (Fehr and Camerer, *Evolution and Human Behavior* 25(2), 63–87, 2007). The modulation of these systems, in particular self-control areas like dorsolateral prefrontal cortex (DLPFC), through TMS and tDCS has proven to be effective in modifying people's behavior in socio-economic contexts (Knoch et al., *Science* 314(5800), 829–832, 2006; Knoch et al., *Cerebral Cortex* 18(9), 1987–1990, 2008; Ruff et al., *Science* 342(6157), 482–484, 2013). The scope of the current paper is to discuss the potential benefits of the enhancement of social norm compliance in the context of therapeutic interventions, along with the issues of methodological, theoretical, and moral nature that may arise when considering the very definition of social norm: indeed, the benchmark for deciding what is right and what is wrong is not always easy to determine in the social context, and thus, the implications of proposing interventions aimed at modulating social norm compliance, although definitely promising, should also be considered carefully.

Keywords Social norms · Neuromodulation · Therapeutic intervention

Fairness, reciprocity and cooperation are some of the norms that promote social interactions. In her book "Grammar of Society," Cristina Bicchieri describes social norms as the grammar of social interaction, as they are implicit and define what is acceptable and what is not in a social situation. As a matter of fact, when we make decisions, big or small, that involve other people, we do assume the existence of a shared knowledge of how appropriate human interactions work. The concept of "appropriate interaction" is where the social norms weight in: people vary in the degree of their willingness to follow norms, but, at least within a similar cultural background, people have an intuition of what should be done in social contexts. Social norms specifically apply to those situations in which there is a conflict between self-interest and collective interest. In a behavioral economics mindset, social norms have been related to other regarding preferences, and they are useful to explain the apparent irrationality of individuals' behavior in a number of situations. Indeed, rational utility theories predict that an agent is driven by self-interested goals that aim to maximize the agent's own payoff. However, plenty of experimental evidence showed that people care about others' outcomes: for example, people sometimes sacrifice their own resources for another persons' benefit (Fehr and Fischbacher 2004). From a cognitive and psychological perspective, social norms and other regarding preferences have been related to constructs such as expectations and emotions.

✉ Claudia Civai
c.civai@kent.ac.uk

[1] School of Psychology, University of Kent, Canterbury, UK

[2] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands

It has been shown that people are more willing to accept unfair offers in the ultimatum game when they expect the offer to be unfair; however, when they expect fairness, they are much stricter in their decisions (Sanfey 2009; Chang and Sanfey 2013). Fairness judgments seem to be also associated with emotional activation: in fact, negative affect increases our proneness to reject unfair deals (Harlé and Sanfey 2007; Moretti and di Pellegrino 2010). Neuropsychological models describe social norm perception and compliance as the interplay between different neurocognitive systems, consisting of areas related to emotional activation and expectations (e.g. anterior insula: Sanfey et al. 2003; Civai et al. 2012), reward (e.g. ventral striatum: Fehr and Camerer 2007; Tabibnia et al. 2008; Tricomi et al. 2010), evaluation (ventromedial prefrontal cortex—VMPFC: Tricomi et al. 2010; Corradi-Dell'Acqua et al. 2013), and self-control (dorsolateral prefrontal cortex—DLPFC: Knoch et al. 2006; Ruff et al. 2013; Baumgartner et al. 2011). Recent studies show accumulating evidence to support these accounts, some of which will be described in the next paragraphs. Notably, the fact that social norm understanding is rooted in our cognitive system means that a disruption of this system, or specific parts of it, may lead to a disruption in social norm compliance. Therefore, novel ideas in the treatment of judgment and decision-making impairments that involve the use of neuromodulation techniques, such as transcranial direct current stimulation (tDCS), offer a long awaited and welcome societal advancement. The current body of research constitutes only the first stage of what hopefully will become a structured attempt to bridge laboratory findings and therapeutic interventions. However, because of the broad and multifaceted way in which social norms are defined, the cognitive enhancement approach to this aspect of cognition is particularly challenging. The aim of this short opinion paper is, therefore, to draw attention on the specific pitfalls and caveats of the neuromodulation of social norm perception, which need to be considered carefully alongside the benefits.

Neuromodulation of Social Behavior and its Prospects

Individuals with psychopathy and antisocial personality disorder fail to conform to social norms. They show chronic antisocial behavior, instrumental aggression, lack of remorse, impulsivity, and deceitfulness (APA 2013; Hare 1996). Multiple theories have been proposed regarding the neural mechanisms of psychopathy, including key roles of the amygdala (Blair 2003; Blair 2005), orbitofrontal cortex—OFC (Birbaumer et al. 2005; Veit et al. 2002), and DLPFC (Buckholtz 2015; Gordon et al. 2004; Müller et al. 2003). For example, individuals with psychopathy compared with controls show decreased amygdala responsivity during negative outcomes, suggesting impaired aversive conditioning (e.g., Rilling et al. 2007; Kiehl et al. 2001). Furthermore, they show increased frontal cortical (e.g., DLPFC) activity during tasks that require emotion recognition, which has been suggested to reflect cognitive, rather than emotional processing of emotional stimuli (e.g., Gordon et al. 2004; Contreras-Rodríguez et al. 2014). Additionally, Buckholtz (2015) proposed a crucial role of DLPFC dysfunction reflecting disrupted representation and integration of (social) rules, context, and cost, into computations that determine subsequent action selection.

tDCS has been suggested to potentially enhance the efficacy of therapeutic interventions for psychiatric disorders, as it may boost neuroplasticity (Kuo et al. 2014; Cornet et al. 2014). Clinical applications will require a deeper understanding of tDCS effects on the brain, including validation, replication, and large-scale trials that measure its effectiveness for the applied purpose, as for pharmacological research. The investigation on neuromodulation effects on social behaviour is currently at an early stage and has been predominantly conducted in healthy individuals, but results so far have been auspicious. For example, anodal tDCS applied to the right DLPFC diminished proactive aggression in men (Dambacher et al. 2015). Transcranial magnetic stimulation (TMS) and tDCS studies on social norm perception mostly aimed at understanding the neural circuits by disrupting the function of specific regions, thereby hampering social norm compliance: studies using rTMS (van't Wout et al. 2005; Knoch et al. 2006) found that interfering with the activation of the DLPFC during an ultimatum game task decreases the chances of rejecting unfair offers. Similar effects have been demonstrated using cathodal tDCS (Knoch et al. 2008). Together, these findings suggest that the DLPFC modulates self-other regarding preferences: in particular, Knoch et al. suggest that self control processes, underpinned by DLPFC, are necessary in order to overcome self-interested motivations of accumulating money (accepting unfair offers) and favor fairness-driven choices (rejecting unfair offers). More recently, however, other studies employed neuromodulation to enhance social norm compliance or, more in general, the effects of social learning. For example, a very promising study by Ruff et al. (2013) showed that changing the neural excitability of the right lateral prefrontal cortex through tDCS enhanced both voluntary and sanction-induced social norm compliance, suggesting that these results may be a starting point for developing interventions to deal with impairments in social behavior. Another recent study by Ligneul et al. (2016) showed that activation of the rostromedial PFC was positively associated with positive (victory) and negatively with negative (defeat) prediction errors in a perceptual decision-making game where participants competed against others. Anodal tDCS over the rostromedial PFC enhanced social dominance learning from victories but not defeat. Interestingly, tDCS effects were specific for learning about social dominance relationships but did

not significantly modulate non-social learning in a control task.

In summary, these studies are promising for future research using tDCS and encourage investigations on the clinical applications for antisocial behavior. However, we also would like to propose caution with the use of neurostimulation in relation to social norm compliance for the reasons that follow.

Caveats

The first ramification that we propose to carefully consider when thinking about interventions employing neuromodulation relates to generalization issues. Using neuromodulation to create a bias towards one type of behaviour may appear beneficial in a specific, well-controlled (laboratory) environment; however, such a bias may be costly to the brains' ability to flexibly adapt our goals and behaviour to a more dynamically changing environment. This notion is not specific to tDCS or TMS, but also extends to psychopharmacological interventions that aim to enhance cognition, and it was recently demonstrated in a methylphenidate study by Ter Huurne et al. (2015) using a visuospatial attention task where facial targets were either paired with low salient distractors (scrambled faces), or high salient distractors (non-scrambled faces). Salient distractors caused interference in both placebo and methylphenidate conditions but significantly more in the methylphenidate condition compared with placebo, suggesting that methylphenidate enhanced attention to the target, but its effect was detrimental when stimulus features were shared by distractors and targets. Biases towards one behavior may be costly for flexible adaptation to the environment, and, more in general, for other cognitive functions, as it has been discussed in Iuculano and Kadosh (2013): findings show that numeric learning was enhanced by the transcranial electrical stimulation (TES) of the posterior parietal cortex and impaired by the TES of the dorsolateral prefrontal cortex, whereas the reverse was true for automaticity for learned material. This argument furthermore outlines a potential harmful consequence of uncontrolled long-term tDCS use (also see diytdcs.com), in addition to previously reported risks of adverse effects (Brunoni et al. 2012). Despite not being specific for social norm perception, this issue is particularly valid for this topic given the complex and flexible nature of social norms, which are extremely susceptible to the smallest change in the environment: a behavior that is in line with a norm may turn out to be inefficient, or even inappropriate, as the context changes. For example, rejecting unfair deals in order to pursue fairness is a useful social norm-driven behavior; however, what is perceived as fair depends on a complex and individual-specific network of areas related to emotion, cognitive control, memory (Chang and Sanfey 2009). This suggests that any interference with one of these systems may lead to a wrong interpretation of the situation or of our own current motivations. Let us now discuss more in detail the complex and flexible nature of social norms.

The second ramification is specific to our topic, and it relates to the potential difficulty in determining a benchmark for social norms. Even if laboratory settings are deemed to be cold and not completely reliable for eliciting social interaction, they are useful for controlled manipulation of the variables at play and reveal the response that was expected by the participant (Falk and Heckman 2009). Clearly, the world is much more complex, and this complexity makes it more difficult to control for the effects of contextual factors in real life: cues can be left unattended, or they may be read differently, goals can differ, and this diversity can elicit different behaviors, which may not necessarily indicate impairment in social norm compliance, but rather the application of a different norm, or the application of the same norm in another direction. For example, if someone accepts a low unfair offer in an ultimatum game, we may conclude that they are not sensitive to social context, and more driven by self-interested motivations rather than by social norms; however, what if they need the money for charitable fundraising? What if they think that, for some reason, they deserve less than their opponents? What if they expect that the majority of people in their situation accept any offer? The acceptance behaviour, in this cases, acquires a different meaning. It is true that some norms may be more fixed than others, by being rooted more deeply in our moral value system: indeed, Bicchieri (2006, 2008) and Elster (2009) draw a distinction between social and moral norms. The idea is that, while social norms are followed conditionally upon satisfaction of expectations, moral norms are followed unconditionally, and thus are less sensitive to the context; examples of moral norms involve contemplating murder or incest. Nevertheless, first of all it is worth noting that the norms considered in the context of this discussion (e.g., reciprocity, fairness, honesty) are defined as social by Bicchieri and as quasi-moral, rather than moral, by Elster. Secondly, we could argue that even moral norms are, to a certain extent, influenced by the cultural and historic context: for example, moral norms concerning homosexuality have changed dramatically in the last 50 years. Lastly, the distinction between social and moral is based not on the content of the norm, but on our attitude towards it (Bicchieri 2006), making certain norms even harder to categorize. It is challenging to develop interventions when the dividing line between a right and a wrong answer can easily get blurred.

Conclusions

We conclude by saying that, in pointing out caveats, our intent is to encourage research on social norm compliance, both in

laboratory and, in the future, also in applied settings. In particular, as far as clinical settings are concerned, it will be of great importance to come up with an operational definition of social norm, or social norm impairment, that can tackle the complexity of the cognitive and brain networks involved in this type of processing, in order to define benchmarks for successful interventions. Clearly, the employment of neuromodulation, in particular tDCS, is becoming increasingly common not only in research and clinical practice but also outside these controlled environments, and for this reason, studies that can contribute to the advancements in the theory of social norms and data interpretation should be welcome and encouraged.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468–1474.

Bicchieri, C. (2006). *The grammar of society: the nature and dynamics of social norms*. New York: Cambridge University Press.

Bicchieri, C. (2008). The fragility of fairness: an experimental investigation on the conditional status of pro-social norms. *Philosophical Issues, 18*, 229–248.

Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy: a functional magnetic resonance imaging study. *Archives of General Psychiatry, 62*(7), 799–805.

Blair, R. J. R. (2003). Neurobiological basis of psychopathy. *The British Journal of Psychiatry, 182*(1), 5–7.

Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Development and Psychopathology, 17*(03), 865–891.

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., et al. (2012). Clinical research with transcranial direct current stimulation (tDCS): challenges and future directions. *Brain Stimulation, 5*(3), 175–195.

Buckholtz, J. W. (2015). Social norms, self-control, and the value of antisocial behavior. *Current Opinion in Behavioral Sciences, 3*, 122–129.

Chang, L. J., & Sanfey, A. G. (2009). Unforgettable ultimatums? Expectation violations promote enhanced social memory following economic bargaining. *Frontiers in Behavioral Neuroscience, 3*, 36. doi:10.3389/neuro.08.036.2009.

Chang, L. J., & Sanfey, A. G. (2013). Great expectations: neural computations underlying the use of social norms in decision-making. *Social Cognitive and Affective Neuroscience, 8*(3), 277–284.

Civai, C., Crescentini, C., Rustichini, A., & Rumiati, R. I. (2012). Equality versus self-interest in the brain: differential roles of anterior insula and medial prefrontal cortex. *NeuroImage, 62*(1), 102–112.

Contreras-Rodríguez, O., Pujol, J., Batalla, I., Harrison, B. J., Bosque, J., Ibern-Regàs, I., et al. (2014). Disrupted neural processing of emotional faces in psychopathy. *Social Cognitive and Affective Neuroscience, 9*(4), 505–512.

Cornet, L. J., de Kogel, C. H., Nijman, H. L., Raine, A., & van der Laan, P. H. (2014). Neurobiological factors as predictors of cognitive–behavioral therapy outcome in individuals with antisocial behavior a review of the literature. *International Journal of Offender Therapy and Comparative Criminology, 58*(11), 1279–1296.

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2013). Disentangling self-and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social Cognitive and Affective Neuroscience, 8*(4), 424–431.

Dambacher, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S., & Sack, A.T. (2015). Reducing proactive aggression through non-invasive brain stimulation. *Social Cognitive and Affective Neuroscience*, nsv018.

Elster, J. (2009). Social norms and the explanation of behavior. In P. Hedström & P. Bearman (Eds.), *The Oxford handbook of analytical sociology* (pp. 195–217). Oxford: Oxford University Press.

Falk, A., & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science, 326*(5952), 535–538.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*(10), 419–427.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87.

Gordon, H. L., Baird, A. A., & End, A. (2004). Functional differences among those high and low on a trait measure of psychopathy. *Biological Psychiatry, 56*(7), 516–521.

Hare, R. D. (1996). Psychopathy and antisocial personality disorder: a case of diagnostic confusion. *Psychiatric Times, 13*(2), 39–40.

Harlé, K. M., & Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the ultimatum game. *Emotion, 7*(4), 876.

Iuculano, T., & Kadosh, R. C. (2013). The mental cost of cognitive enhancement. *The Journal of Neuroscience, 33*(10), 4482–4486.

Kiehl, K. A., Smith, A. M., Hare, R. D., Mendrek, A., Forster, B. B., Brink, J., et al. (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry, 50*(9), 677–684.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829–832.

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cerebral Cortex, 18*(9), 1987–1990.

Kuo, M. F., Paulus, W., & Nitsche, M. A. (2014). Therapeutic effects of non-invasive brain stimulation with direct currents (tDCS) in neuropsychiatric diseases. *NeuroImage, 85*, 948-960.

Ligneul, R, Obeso, I., Ruff, C., & Dreher, J.C. (2016). Dynamical representation of dominance relationships in the human medial prefrontal cortex. Current Biology, 26(23), 3107-3115.

Moretti, L., & Di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion, 10*(2), 169.

Müller, J. L., Sommer, M., Wagner, V., Lange, K., Taschler, H., Röder, C. H., et al. (2003). Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: evidence from a functional magnetic resonance imaging study using pictures with emotional content. *Biological Psychiatry, 54*(2), 152–162.

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., et al. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry, 61*(11), 1260–1271.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482–484.

Sanfey, A. G. (2009). Expectations and social decision-making: biasing effects of prior knowledge on Ultimatum responses. *Mind & Society, 8*(1), 93–107.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755–1758.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19*(4), 339–347.

ter Huurne, N., Fallon, S. J., van Schouwenburg, M., van der Schaaf, M., Buitelaar, J., Jensen, O., et al. (2015). Methylphenidate alters selective attention by amplifying salience. *Psychopharmacology, 232*(23), 4317–4323.

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature, 463*(7284), 1089–1091.

van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport, 16*(16), 1849–1852.

Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neuroscience Letters, 328*, 233–236.