Some Correlates of Agency Ascription and Emotional Value and their Effects on Decision-Making

Megan Strait, Gordon Briggs, and Matthias Scheutz
Department of Computer Science
Tufts University
Medford, MA 02155
Email: {mstrait, gbriggs, mscheutz}@cs.tufts.edu

***Abstract*—The prefrontal cortex (PFC) has been investigated extensively with functional magnetic resonance imaging (fMRI) and identified as a neural correlate of emotion regulation and decision-making, particularly in the context of moral utilitarian dilemmas. However, there are two limitations of previous work: (1) fMRI requires strict constraints on the physical experimental environment and (2) experimental manipulations have yet to consider the role of agency on the dilemma outcome and the corresponding neural activity. In this paper, we extend previous work by first evaluating an alternative neuroimaging technique, functional near infrared spectroscopy (NIRS), for observing decision-making processes in a less-constrained environment. We then examine the role of agency in deciding emotional (moral) and non-emotional dilemmas through a 2-part, 20-subject preliminary investigation. Our findings are two-fold: they suggest (1) NIRS is a potential alternative to fMRI in this decision-making context and (2) agency shows some influence on prefrontal neural activity, making NIRS a promising method for objective evaluation of agency and emotional value in human-agent interactions.**

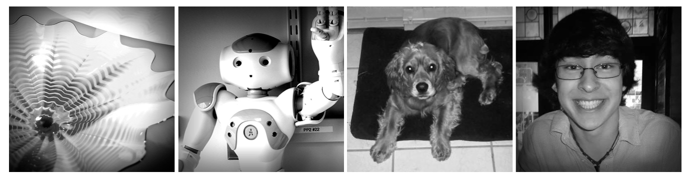

Fig. 1. Agents used in the following experimental protocols. The control agent, an inanimate glass object (far left), and human (far right) types were used in both experiments. The robot (Aldebaran Nao), middle left, and canine, middle right, agent types were added in the second experimental protocol to evaluate effects of perceived agency.

I. Introduction

Utilitarian dilemmas, which involve a conflict between competing imperatives, have long been used by philosophers and psychologists to study the cognitive processes involved in emotion regulation and decision-making. A standard example of a moral utilitarian dilemma is that of the "trolley problem", which is formulated as follows:

> *Suppose there is a runaway train which can only be steered from one track onto another. 5 people are working on one track and one person on the other; anyone on the track which is entered will be killed.*

The observer (the participant) must decide whether to exchange one person's life for the lives of five or to exchange five lives for one. The utilitarian view seeks to maximize welfare (or minimize harm), and as such, the morally preferred course of action is to steer the train to the track with only one person. However, an alternative view asserts moving to another track constitutes a participation in the moral wrong, making one partially responsible for the death when otherwise no one would be responsible.

Over the past decade this utilitarian dilemma has been employed in a multitude of neuroimaging studies to identify the psychological and neural substrates underlying emotion regulation and decision-making, highlighting, in particular, the role of the anterior prefrontal cortex (aPFC) [1]–[3]. However, this work is conducted within the confines of an fMRI, which imposes a number of constraints which limit experimental conditions to unrealistic settings. Furthermore, these neuroscientific studies largely ignore the role of the perceived agency of the moral patients[1](those affected by the dilemma outcomes) in the correlated neural activity.

Relative to fMRI, functional near infrared spectroscopy (NIRS), is more portable and less obtrusive, making it a promising alternative. Restrictions on participants' movement are substantially less using NIRS (see Figure 2. This avoids effects on participants' natural behavior (i.e. moving), yielding the ability to conduct experiments in more realistic settings. Furthermore, beyond the gains in portability and reduction in physical constraints, it affords the use of physical, embodied agents within close proximity to the participant. As agency attribution has been shown to be a dynamic process subject to experience and interaction [4], [5], it is important to extend neuroimaging techniques to environments in which interaction with embodied agents (e.g., robots) can occur.

In this paper, we first evaluate NIRS as a technique for measuring decision-making in the standard fMRI paradigm of utilitarian moral dilemma. We then extend related work by investigating the signal-based component of agency ascription and its correlation to behavioral outcomes in two (emotional and non-emotional) experimental protocols. We conclude with a discussion of the potential applications and limitations of NIRS for evaluating emotion regulation and decision-making cognitive processes in human-agent interaction settings.

Fig. 2. NIRS equipment and physical setup. Participant is fitted with two NIRS sensors, secured by a tight-fitting cap, and tethered to the NIRS oximeter (behind, right) and a laptop for data recording.

II. Related Work

Below we describe related neuroscientific studies concerning the investigation of moral decision-making and its neural correlates, followed by evidence of agency ascription as an important factor in decision-making.

A. Manipulations of Emotional Artifacts

Extensive work using lesion studies and transcranial magnetic stimulation (TMS) has identified the anterior prefrontal cortex (aPFC) as central to moral decision-making [1]–[3], [6]–[10]. Additional work has shown direct connections between the ventromedial prefrontal cortex (vmPFC) and the amygdala in emotion regulation (ER) tasks [11], [12]. As the standard utilitarian dilemma involves ER due to negative stimuli (killing), vmPFC activity is also observed in moral decision-making.

To further explore the aPFC and to what extent various factors implicit in utilitarian dilemma might have, additional studies have been conducted. These include, in particular, manipulations hypothesized to elicit increased emotional engagement [13]–[18], such as:

- *action immediacy*: participant performs the action versus the participant tells a surrogate to perform the action [19]
- *personal force*: participant performs direct harm (i.e. pushes a person in front of a train) versus indirect harm (i.e. switches the train tracks) [3], [20]
- *visual immediacy*: the participant imagines the scenario versus sees pictures of the scenario [21]

Other manipulations have investigated the effects of cognitive load [22], honesty [23], intent [24], and stereotype [25].

However, this work is limited in two key aspects: (1) the experimental conditions are highly constrained and thus not representative of realistic conditions surrounding moral dilemma and (2) they do not consider the role of agency of the patients (those being affected by the dilemma outcomes) in the dilemma. That is, dilemma employed in related work concern *only human patients*, as opposed to animal patients (e.g., cats and dogs). Thus it remains unaddressed whether it is the patient's ascribed agency or the emotional context or both that elicits the corresponding hemodynamic activity and how that affects participants' behavioral decisions.

B. Evidence for the Role of Agency in Decision-Making

Behavioral evidence from recent human-robot interaction studies suggests that agency factors into decision-making at least with respect to non-human agents. A two-part study mimicking the Stanley Milgram experiments showed that perceptions of agency in robotic artifacts play a role in moral decision-making [26]. It first demonstrated that people have less concern for robotic agents than human counterparts, and proceeded to show that humans had more willingness and enthusiasm to destroy robots of lower perceived agency. Another study found effects of perceived agency on how successfully a robot could dissuade a human participant from completing an emotionally-sensitve task [27]. Although these studies are not of the standard utilitarian dilemma employed in the above imaging studies and are limited in the range of agency levels evaluated, the observed differences in behavioral outcomes suggest an object's agency influences the processes involved in decision-making.

In addition, work regarding theory of mind and anthropomorphism suggest that agency ascription is dynamic and changes with experience and interactions [4], [5]. This is particularly important when considering the influences of perceptions of agency on decision-making, as static assessements of agency (e.g., as in pre or post-experiment surveys) may not correlate with behavior if there are small temporal variations in agency ascription during experimentation.

In the following sections, we attempt to address the physical limitations that fMRI faces by first employing NIRS for measuring hemodynamic activity in more realistic settings using a standard moral dilemma scenario. Second, we conduct a preliminary investigation of the role of agency in decision-making by varying the agent types (i.e. robot, dog, and human) involved in utilitarian decision-making tasks.

III. Materials and Methods

A 2-probe (2-channel), ISS OxiplexTS[2] near infrared tissue oximeter was used to record hemodynamic activity in the

[2]http://www.iss.com/biomedical/instruments/oxiplexTS.html

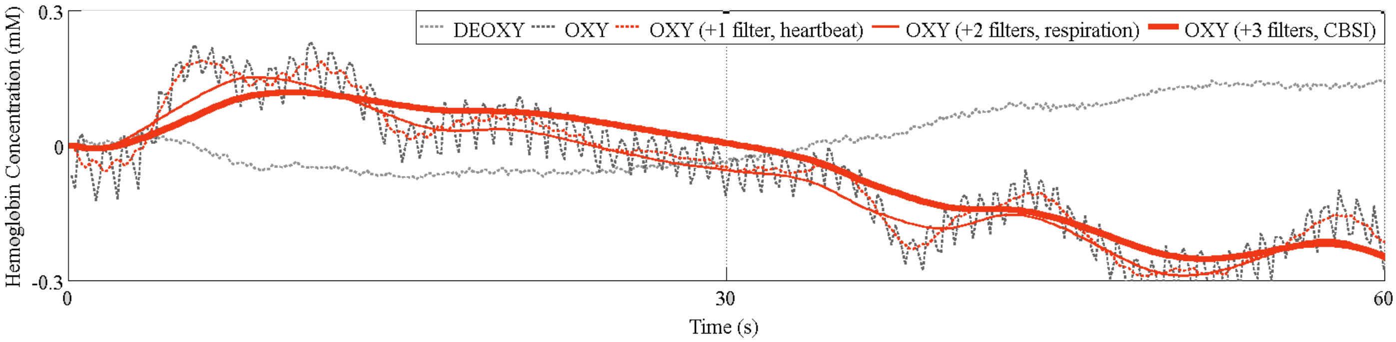

Fig. 3. Signal processing. Heartbeat artifacts are first filtered with a lowpass filter between 0.6Hz and 2Hz. From the resulting signal, respiration is removed. Finally, CBSI is applied to the denoised signal (shown in red, bold). The resulting signal shows the inverse correlation between oxy and deoxy measures.

aPFC at a temporal resolution of 6.25Hz. An elastic, black headband was used to securely fit the NIRS probes in place on the subject's forehead to sample the left and right aPFC respectively. The subject was seated in a generic office chair (Zoom Seating, model SP46105).

A. Signal Processing

The ISS OxiplexTS records relative absorption and scattering coefficients of the sampled tissue, which require additional processing prior to statistical analysis (see Figure 3).

1) Conversion to Hemoglobin: The raw measures are first converted to hemoglobin units using the modified Beer-Lambert Law (MBLL). The MBLL implentation from NIRS-SPM[3], a publicly available processing and analysis package for NIRS data, was used yielding a measure of deoxygenated (Hb) and oxygenated (HbO) hemoglobin for each probe (left and right).

2) Noise Reduction: Systemic (cardiac pulsations and respiration) and low frequency noise was reduced using a recursive low pass filter based on the NIRS Analysis Package [28], followed by a correlation-based signal correction (CBSI)[4] [29] and finally, averaging over trial repetitions.

As the CBSI correction is calculated from the correlation between Hb and HbO measures, the deoxygenated (Hb) measures become redundant and are thus discarded from statistical consideration at this point. The HbO signals are then truncated to a 20s window after the delivery of the dilemma instruction (5-25s, exclusive). This 20s truncated signal maximizes chances of capturing the peak of the hemodynamic response even if the peak of the response varies temporally as a function of condition. As a result, *two 20s HbO signals (left and right aPFC) per condition* remain for statistical analysis.

B. Statistical Inference

To determine whether experimental conditions elicited task-related hemodynamic changes, within-subjects paired t-tests (with Bonferroni corrections) were run on each experimental condition HbO compared with the baseline condition HbO. Experimental conditions showing significant activation were then compared pair-wise to determine whether significant differences existed as a main effect of condition.

To determine between-subjects trends, an area-under-the-curve (AUC) summary statistic is calculated by summing the signal change (condition hemoglobin concentration - baseline hemoglobin concentration) across the 20s truncated signal and calculating the mean AUC. This results in one mean AUC value for the left and for the right aPFC, for each subject.

IV. Moral Decision-Making

To investigate whether NIRS, as a neuroimaging technique, could measure the hemodynamic activity in the aPFC associated with moral decision-making, we conducted a within-subjects experiment replicating the aforementioned fMRI studies on moral decision-making.

A. Design

We constructed a set of 16 utilitarian moral dilemmas (two conditions, eight trials of each; see Figure 4, far left and far right) and a corresponding scenario to explain the task to the participants. Participants were instructed that they would be managing "emergency evacuations" to transport endangered patients to safety with the goal of *evacuating* as many as possible.

1) Conditions: We designed two conditions, a *control* condition to serve as a baseline comparison for our *test* condition, as we had a limited region of measurement (only the aPFC). In the *control* condition, the participant evacuated 8 nonliving, inanimate patients (glass blown objects), and in the *test* condition, the participant evacuated 8 standard human patients (see Figure 1). A set of 8 images of glass blown objects and 8 random images of people were collected and paired with the 16 dilemmas. Conditions were administered in blocks (of 8 trials, randomized) preceded by a 30-second resting sample (for the conversion to hemoglobin) and instructions (i.e. "there is a fire in office x, evacuate as many people as possible"), and counterbalanced.

2) Trial: A trial was composed of four parts:

- A pre-dilemma period (30s) consisting of a simple counting task to ensure the participant's attention to the computer screen.
- A textual description (15s) of a moral dilemma accompanied by a randomized photo of an agent. The type of the agent displayed (either glass or human)

[3]http://bisp.kaist.ac.kr/NIRS-SPM.html
[4]http://www.alivelearn.net/nirs/CBSI.m

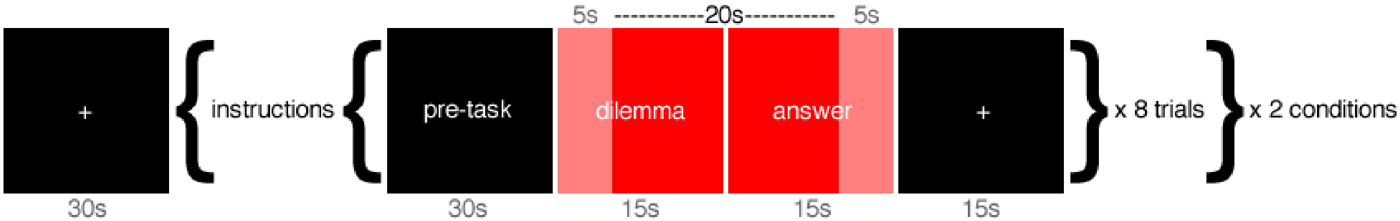

Fig. 4. Experimental design. A pre-dilemma task signaled the participant's attention and a fixation point signaled a resting period. The parts indicated in red contained a moral dilemma stimulus and decision, during which the participant's hemodynamic activity was sampled. The part indicated in bright red shows the truncated signal which was used in the statistical analyses. Both the conditions and the trials were randomized.

was in accordance with the condition (control or test, respectively). In the *control* condition, the dilemmas also employed physical impediments to evacuating the objects in a timely fashion (i.e. instead of 'a boy in a wheelchair' one might have a glass blown object too large or too heavy, etc. to evacuate quickly).

- An answer period (15s) for the participant to select an outcome (save or leave) in response to the dilemma.
- A rest period (15s) where the participant was instructed to focus on a fixation point and relax.

B. Population and Procedure

10 healthy, right-handed subjects (5 male), ages 19 to 33 (M = 22.0, SD = 3.8) were recruited via an affiliated university website and provided informed, written consent. To avoid any learning-based effects, the participant first completed four practice trials. Following, the participant completed 8 trials of each condition in succession.

C. Results and Discussion

As expected, the experimental condition with human agents showed significant activation ($p < .001$) from the baseline (glass control) condition for all 10 participants, in both the left and right aPFC. Such strongly significant differences – observed across all participants – suggest that NIRS is capable of measuring activity related to moral decision-making in the aPFC. Although it would be necessary to conduct a multimodal (NIRS, combined with fMRI) investigation to confirm this interpretation, we proceeded with a follow-up experiment to preliminarily investigate the role of agency in decision-making.

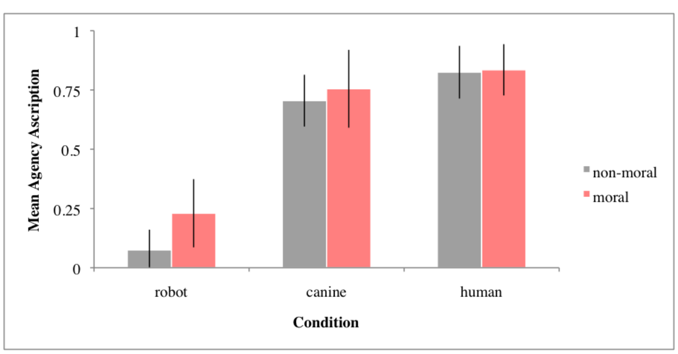

Fig. 5. Mean agency ascription across all subjects (N = 10) for each agent type (averaged across the 5 dimensions of mental state attribution). Ratings were normalized to a scale from 0 to 1 prior to averaging across subjects.

V. Agency Ascription and Decision-Making

To investigate the role of the agency in decision-making (emotional and non-emotional), we conducted a two-part follow-up to the previous experiment in which we introduced two additional agent types, dog and robot to represent four hypothetical levels of agency (see Figure 1). We selected robots as prior work suggested lesser agency is ascribed to robots in comparison to humans [26], [30], and specifically used the Aldebaran Nao for its humanoid appearance. Dogs were selected based on [5], which showed canines ascribed more agency than robots, but less than humans. We considered the moral stimuli used in the previous investigation as a subset of emotional stimuli and thus use moral and emotional interchangeably, as well as non-moral and non-emotional.

A. Design

We conducted both a non-moral and moral protocol, each with a distinct set of 10 participants, to avoid confounding emotional artifacts with agency ascription in the NIRS signal. Both protocols employed four conditions: the glass control and human agent of the previous experiment, and additionally robot and canine agents as well (see Figure 1).

1) Moral Protocol: We constructed a set 24 moral utilitarian dilemmas, for 6 dilemmas (trials) per each of the now four conditions. The number of trials was reduced from 8 to 6 in order to avoid significantly extending the total session time. No other modifications to the previous protocol (of the Moral Decision-Making replication experiment) were made.

2) Non-Moral Protocol: In this protocol, the evacuation scenario was modified to be a *relocation* scenario, where the goal was to *relocate* as many patients as possible. As in the moral protocol, we constructed a set of 24 non-moral utilitarian dilemmas (with one additional modification to the answer stimulus: instead of the option to "save", the participant now had the option to "take"). As an example of a non-moral dilemma, (in the instance of a glass blown object) the dilemma read as too large or too heavy, etc. to *transport* (instead of *evacuate*) quickly.

3) Additional Metrics: To assess participants' perceptions of the agency, we employed a post-questionnaire sampling five dimensions (e.g., capacity to feel pain) of mental capacities on a 5-point Likert scale (see Figure 5). Responses to the agency questionnaire reflected our hypothesis, with humans being attributed the greatest level of agency, followed by dogs, followed by robots; however, the difference in agency ratings between human and canines was not significant for either protocol population. The control condition (glass) was also

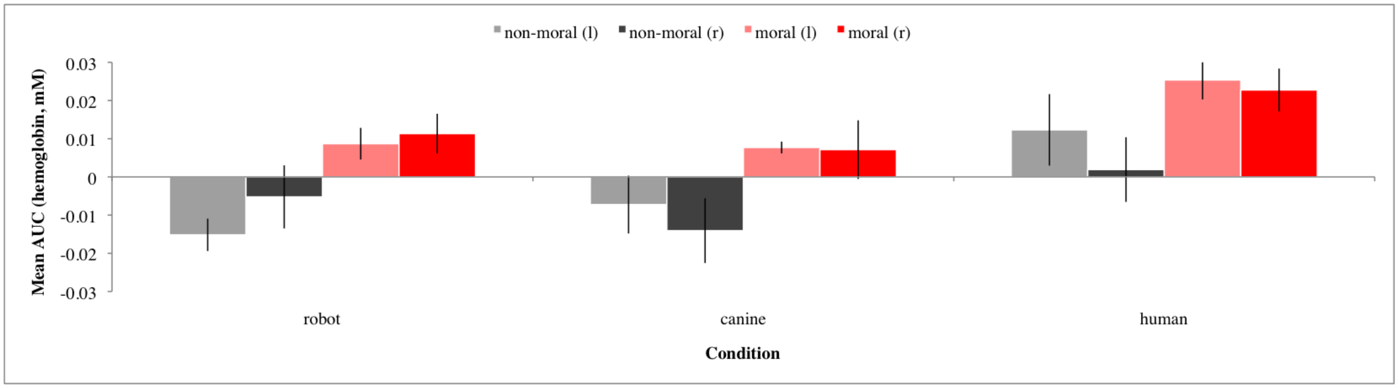

Fig. 6. Mean hemodynamic response across subjects, by agent type. Error bars represent the 95% confidence interval (computed within-subjects).

rated and received unanimous ratings of no agency. Participants' decisions on the dilemma tasks were also recorded. A measure of behavioral outcome likelihood from -1 (kill/leave) to 1 (save/take) was calculated by summing and then averaging decisions across the 6 trials for each condition.

B. Population and Procedure

20 healthy, right-handed subjects participated – 10 (3 male), ages 18 to 31 ($M = 20.8$, $SD = 3.6$) in the moral protocol and 10 subjects (3 male), ages 19 to 22 ($M = 20.2$, $SD = 1.3$) in the non-moral protocol. Participants again completed four practice trials prior to the experimental conditions and conditions counterbalanced.

C. Results and Discussion

1) Behavior and Subjective Ratings of Agency: Subjective ratings of agency (see Figure 5), significantly differed ($p < .05$) between living and non-living agent categories for both protocols. Additionally, the moral protocol showed a significant difference between robot and glass. Regarding behavioral outcomes, there was a significant difference between the human and robot in the moral protocol, with participants more likely to save human agents. All other comparisons in both protocols were non-significant (see Figure 7).

Further analysis using Pearson's product-moment correlation showed a significant correlation ($r = .553$, $p = .001$) between agency rating and behavioral outcome in the moral

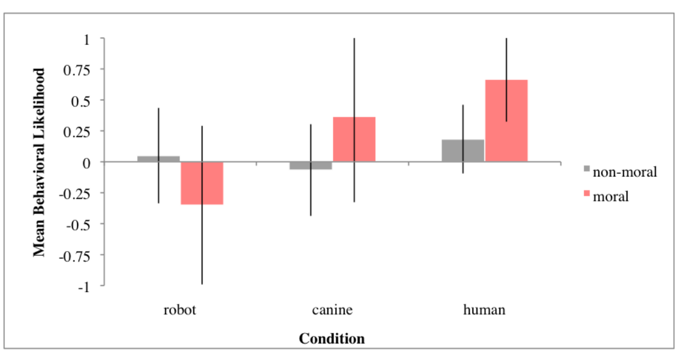

Fig. 7. Mean likelihood of behavioral outcomes across subjects (N = 10). Outcomes correspond to kill/leave (-1) and save/take (1).

(but not non-moral, $p = .496$) protocol. Combined with the significant increase in agency attributed to the robot agent type in the moral protocol compared to the non-moral protocol, the moral protocol might involve greater emotional engagement in decision-making, as opposed to the non-moral protocol which seems likely to operate without affect (as shown behavioral data without any skew towards one outcome or the other).

2) Neural Activity: Figure 6 shows the overall neural agent trends across subjects for both protocols. Qualitative analysis shows hemodynamic responses to human stimuli greatly exceed those of robot and canine stimuli, regardless of protocol. Regarding hemispheric effects, left and right probes show similar activation regardless of agent type in the moral protocol, whereas the non-moral protocol shows substantial change in the left channel to robot and human agents and little change in the right. Furthermore, this trend is reversed for the canine agent (greater change in the right channel).

Within-subjects analyses showed a main effect of condition on NIRS signal in both experimental protocols. Compared to the baseline condition signal, significant activation ($p < .001$) was elicited in the left aPFC for all 20/20 participants in all three experimental conditions (robot, dog, and human). Comparisons to determine whether agency-based NIRS activation was different between conditions showed significant differences (between all pairings of conditions) in the left aPFC for 16/20 participants (8/10 in the moral and 8/10 in the non-moral protocol). The same comparisons for the right aPFC showed significant activation (baseline versus experimental conditions) in 15/20 participants (moral: 9/10 and non-moral: 6/10), and significant differences (again between all pairings of conditions) in 13/15 (moral: 7/9 and non-moral: 6/6) participants (participants without significant activation from baseline were excluded from the between-condition analysis).

Correlation analysis (including subjects without significant activation) of mean AUC showed no significant trends between neural activity and behavior or agency ratings; however analysis of aPFC (r) x behavior showed a trend towards significance in the moral protocol ($r = -.292$, $p = .117$). The analysis also showed an overall trend towards significant correlation between aPFC (l) and agency ratings ($r = .227$, $p = .080$). The absence of correlation may potentially be due to the use of the AUC summary statistic, which ignores the temporal aspect of the NIRS signal, but further analysis that includes temporal NIRS features would be necessary to confirm.

VI. Conclusions

The first of these two investigations supports NIRS as a potential alternative to fMRI for measuring neural processes recruited in moral dilemma scenarios, thus allowing for a multitude of more realistic investigations on emotionally-sensitive decision-making tasks. This study was conducted still in a very controlled fashion, as participants were instructed to minimize their physical movement (e.g., avoid scratching, stretching, etc.). It would require further investigation to validate both (1) whether NIRS would be suitable for realistic, let alone 'in the wild' investigations, and (2) whether the activity measurable in this protocol with NIRS fully corresponds to that measured using fMRI; however, it demonstrates the evaluation of decision-making processes in more realistic settings than what is currently possible with fMRI or TMS.

The second of the two suggests agency ascription plays a role in decision-making. In particular, the significant correlation between agency ratings and behavioral outcomes as well as the trend towards significant correlation between neural activity and outcomes in the moral (and not in the non-moral) protocol show an interaction between agency and emotional severity. Qualitative analysis of between-subject trends also shows differential neural activity dependent on both emotional context (moral and non-moral) as well as agent type, suggesting agency is of varying relevance in decision-making based on the emotional context. Although replication using fMRI is necessary to confirm NIRS as a valid alternative, as well as TMS to pinpoint the prefrontal substrates of agency ascription, this paper provides a preliminary evaluation of NIRS for studying decision-making processes in the presence of emotion and agent-based artifacts. While the technology and results are both limited in scope and applicability, we hope they may serve as a basis for further investigation of agency in emotional and non-emotional decision-making.

Acknowledgment

This work was in part funded by the National Science Foundation under Grant No. ISS-1065154.

References

[1] W. D. Casebeer, "Moral cognition and its neural constituents," *Nature Reviews Neuroscience*, vol. 4, pp. 840–846, 2003.

[2] M. Koenigs, L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser, and A. Damasio, "Damage to the prefrontal cortex increases utilitarian moral judgements," *Nature*, vol. 446, pp. 908–911, 2007.

[3] C. E. Forbes and J. Grafman, "The role of the human prefrontal cortex in social cognition and moral judgment," *Annual Reviews of Neuroscience*, vol. 33, pp. 299–324, 2010.

[4] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," *Psychological Review*, vol. 114, pp. 864–886, 2007.

[5] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *Science*, vol. 315, p. 619, 2007.

[6] E. Ciaramelli, M. Muccioli, E. Ladavas, and G. di Pellegrino, "Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex," *Soc. Cog. and Affect. Neurosci.*, vol. 2, pp. 84–92, 2007.

[7] J. D. Greene, R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen, "An fmri investigation of emotional engagement in moral judgment," *Science*, vol. 293, pp. 2105–2108, 2001.

[8] J. D. Greene, L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen, "The neural bases of cognitive conflict and control in moral judgment," *Neuron*, vol. 44, pp. 389–400, 2004.

[9] H. R. Heekeren, I. Wartenburger, H. Schmidt, H. Schwintowski, and A. Villringer, "An fmri study of simple ethical decision-making," *NeuroReport*, vol. 14, pp. 1215–1219, 2003.

[10] L. Young, A. Bechara, D. Tranel, H. Damasio, M. Hauser, and A. Damasio, "Damage to the ventromedial prefrontal cortex impairs judgment of harmful intent," *Neuron*, vol. 65, pp. 845–851, 2010.

[11] H. L. Urry, C. M. van Reekum, T. Johnstone, N. H. Kalin, M. E. Thurow, H. S. Schaefer, C. A. Jackson, C. J. Frye, L. L. Greischar, A. L. Alexander, and R. J. Davidson, "Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults," *J. Neurosci.*, vol. 26, pp. 4415–4425, 2006.

[12] T. D. Wager, M. L. Davidson, B. L. Hughes, M. A. Lindquist, and K. N. Ochsner, "Prefrontal-subcortical pathways mediating successful emotion regulation," *Neuron*, vol. 59, pp. 1037–1050, 2008.

[13] F. Cushman and J. D. Greene, "Finding faults: How moral dilemmas illuminate cognitive structure," *Social Neuroscience*, vol. 7, pp. 269–279, 2012.

[14] F. Cushman, K. Gray, A. Gaffey, and W. B. Mendes, "Simulting murder: The aversion to harmful action," *Emotion*, vol. 12, pp. 2–7, 2012.

[15] J. Greene and J. Haidt, "How (and where) does moral judgment work?" *TiCS*, vol. 6, pp. 517–523, 2002.

[16] J. Moll, R. Zahn, R. Oliveira-Souza, F. Kreger, and J. Grafman, "The neural basis of human moral cognition," *Nature Reviews Neuroscience*, vol. 6, pp. 799–809, 2005.

[17] J. Moll and R. de Oliveira-Souza, "Moral judgements, emotions and the utilitarian brain," *TiCS*, vol. 11, pp. 319–321, 2007.

[18] P. Valdesolo and D. DeSteno, "Manipulations of emotional context shape moral judgment," *Psychological Science*, vol. 17, pp. 476–477, 2006.

[19] N. Paharia, K. S. Kassam, J. D. Greene, and M. H. Bazerman, "Dirty work, clean hands: The moral psychology of indirect agency," *Organizational Behavior and Human Decision Processes*, 2009.

[20] H. R. Heekeren, I. Wartenburger, H. Schmidt, K. Prehn, H.-P. Schwintowski, and A. Villringer, "Influence of bodily harm on neural correlates of semantic and moral decision-making," *NeuroImage*, vol. 24, pp. 887–897, 2005.

[21] E. Amit and J. D. Greene, "You see, the ends don't justify the means: Visual imagery and moral judgment," *Psychological Science*, 2012.

[22] J. D. Greene, S. A. Morelli, K. Lowenberg, L. E. Nystrom, and J. D. Cohen, "Cognitive load selectively interferes with utilitarian moral judgment," *Cognition*, vol. 107, pp. 1144–1154, 2008.

[23] J. D. Greene and J. M. Paxton, "Patterns of neural activity associated with honest and dishonest moral decisions," *PNAS*, vol. 106, 2009.

[24] J. D. Greene, F. A. Cushman, L. E. Stewart, K. Lowenberg, L. E. Nystrom, and J. D. Cohen, "Pushing moral buttons: the interaction between personal force and intention in moral judgment," *Cognition*, vol. 111, pp. 364–371, 2009.

[25] C. E. Forbes, C. L. Cox, T. Schmader, and L. Ryan, "Negative stereotype activation alters interaction between neural correlates of arousal, inhibition and cognitive control," *Social Cognitive and Affective Neuroscience*, 2011.

[26] C. Bartneck and J. Hu, "Exploring the abuse of robots," *Interaction Studies*, vol. 9, pp. 415–433, 2008.

[27] G. Briggs and M. Scheutz, "Investigating the effects of robotic displays of protest and distress," *Social Robotics*, pp. 238–247, 2012.

[28] T. Fekete, D. Rubin, J. M. Carlson, and L. R. Mujica-Parodi, "The nirs analysis package: Noise reduction and statistical inference," *PLoS ONE*, vol. 6, 2011.

[29] X. Cui, S. Bray, and A. L. Reiss, "Functional near infrared spectroscopy (nirs) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics," *NeuroImage*, vol. 49, pp. 3039–3046, 2010.

[30] C. Bartneck, T. Kanda, O. Mubin, and A. A. Mahmud, "Does the design of a robot influence its animacy and perceived intelligence?" *International J. of Social Robotics*, vol. 1, pp. 195–204, 2009.