A Self for Others:
Joint Self-Other Representation of Value During Morally Relevant Action

Remya Nair (rnair@caltech.edu)
Division of Humanities & Social Sciences, California Institute of Technology
Pasadena, CA 91125 USA

Mark Graves (markgraves@fuller.edu)
Travis Research Institute, Fuller Theological Seminary
Pasadena, CA 91182 USA

Kevin S. Reimer (kreimer@uci.edu)
Department of Education, University of California Irvine
Irvine, CA 92697 USA

Warren S. Brown (wsbrown@fuller.edu)
Travis Research Institute, Fuller Theological Seminary
Pasadena, CA 91182 USA

Steven Quartz (steve@hss.caltech.edu)
Division of Humanities & Social Sciences, California Institute of Technology
Pasadena, CA 91125 USA

Gregory R. Peterson (greg.peterson@sdstate.edu)
Department of History, Political Science, Philosophy & Religion, South Dakota State University
Brookings, SD 57007 USA

Dirk Schümann
Institute for Systems Neuroscience, University of Hamburg Medical Center - Eppendorf
D-20246 Hamburg, Germany

Jan Gläscher
Institute for Systems Neuroscience, University of Hamburg Medical Center - Eppendorf
D-20246 Hamburg, Germany

Michael Spezio (mspezio@scrippscollege.edu)
Department of Psychology, Scripps College
Claremont, CA 91711 USA
Institute for Systems Neuroscience, University of Hamburg Medical Center - Eppendorf
D-20246 Hamburg, Germany
Division of Humanities & Social Sciences, California Institute of Technology
Pasadena, CA 91125 USA

Abstract

The cognitive science of moral action seeks accounts of moral cognition – and their conceptual and valuational structures – that explain stable or unstable, reasoned or unreasoned, moral commitments in the real world. To be successful, cognitive science requires experimental approaches that are relevant to the lives and choices of people who demonstrate stable moral commitment in real life. Further, cognitive science should be able to develop models analogous to the theories from other scholarly inquiries into moral cognition, such as moral philosophy and theology. We applied cognitive valuational modeling and Bayesian model comparison to analyze choices in groups of people who 1) demonstrate real-world stable and reasoned action for others in long-term commitments of compassionate care; 2) demonstrate stable and reasoned action in the laboratory over 2-3 years and across context; and 3) a large group of young adults. We compared 4 different models, intended to correspond with being insensitive to context (Model 1), with simple ethical utilitarianism (Model 2), with an ethics of nondual self (Model 3), and with an ethics of relationally nondual self (Model 4). In all 3 studies, greater action for others associated with having a joint representation of values for self and others while still differentiating between the two (Model 4). Our findings show that action for others is facilitated by having a "self for

others": a representation of value for self that is tied to value for others without losing the distinction between the two.

Keywords: moral action, moral cognition, virtue, character, decision science, valuational modeling, Bayesian model comparison

Introduction

Two primary questions motivate this research. The first is whether the application of cognitive modeling methodologies from decision science, using discrete choice theory (W. H. Greene, 2009; Mazzanti, 2003), Bayesian parameter estimation (Kruschke, 2010), and Bayesian model comparison (Gelman, Hwang, & Vehtari, in press; Vehtari & Gelman, 2014) could reveal how people value self and other during contexts that allow, but that do not require, compassion and costly care for others. Do people who act more often for others represent the value of self and of other differently than those who choose not to care when the opportunity arises? Do people who care have a cognitive representation of joint valuation that is absent or simply modulated in people who act less often on behalf of others? We compared four different cognitive valuational models intended to correspond to the following broad ethical theories: 1) context insensitivity, which might include a deontological ethics (Herman, 2007; Kant, 1996 (1798), 2005); 2) simple utilitarian ethics, with its focus on additivity in aggregates of value (J. D. Greene, Nystrom, Engell, Darley, & Cohen, 2004; Mill, 1871); 3) an ethics of nondual self, in which the values of self and other are merged such that one is indistinguishable from the other (Gethin, 2011; Heim, 2011; Lopez, 2008); and 4) an ethics of relationally nondual self, in which the values of self and other are held together while maintaining the distinction between them (Aquinas, 1964; Aristotle, 1992 (1925); Bonhoeffer, 1998, 2005(1949); Frick, 2008).

The second major question is whether laboratory tasks designed to study ethical action are useful for groups of people whose long-term decisions in the real world show clear evidence of costly action on behalf of others. Economists, educators, policy makers, religious leaders, and grant-making foundations have all raised serious doubts about whether tasks designed in the laboratory, sometimes dismissed pejoratively as "only games", are able to meet this challenge. Moreover, unless such an extension of laboratory tasks is possible, even scholars who are open to cognitive scientific approaches will object that the research findings from laboratory participants are unhelpful due to "moral averaging." (Peterson, Van Slyke, Spezio, Reimer, & Brown, 2010)

"Moral averaging" refers to the practice of making theoretical and mechanistic inferences about moral cognition from measures during ethically salient choices made by typical laboratory participants, whose actual histories of unstable, stable, reasoned, or unreasoned ethically salient choices are unknown. This practice is akin to developing a cognitive science of calculus or a cognitive science of language without first assessing whether the laboratory group understands calculus or the language under investigation, respectively. The results from such work would assuredly include replicable patterns across groups, but these patterns would likely be unhelpful to the understanding of the cognitive processes calculus and language.

We took two approaches to avoid moral averaging. In Study 1, we worked with a group of participants with long-term commitments to stable and compassionate care of adults with mild to profound neurodevelopmental disorders, primarily via close, one-to-one caring dyads. They are all members of the L'Arche organization (L'Arche USA; http://www.larcheusa.org/), which was recognized by Pope John Paul II as "a sign of hope in a divided world." We then applied cognitive valuational modeling to the choices that they made in a novel "rescue decision" task that allows, but does not require, costly care for others under ambiguous threat to self. In Study 2, we first used decisions about the common good to classify a people as Giving and as matched Controls. We tested the stability and the generality of both the Giving and Control groups by asking them to return after 2-3 years to complete the same "rescue decision" task as the L'Arche members had done. We also assessed whether those who stably acted on behalf of others displayed a valuational representation similar to the L'Arche sample. In Study 3, we applied the valuational modeling outcomes from Studies 1 and 2 to an analysis of data from a large group of young adult participants in a northern European city, to determine whether valuational representations from the two groups of stable givers in Studies 1 and 2 would also associate with giving in a more generic group of participants from a different cultural context. All studies involving the Rescuer Paradigm (RP) used real money, and participants always began with twice as much money as the Victim. In Studies 1 and 2, the Victim is a real person who is not present and who is unknown to the participant and who will never have an opportunity to reciprocate any help that the participant provides. In Study 1, participants began with a total of US$60, in Study 2, participants began with a total of US$90, and in Study 3, participants began with a total of 15 euro.

Study 1

In the first study, 48 members of L'Arche USA (Age: M ± SD = 40.9 ± 15.9, range = 21-84 years, 34 women) completed 30 rounds of the Rescuer Paradigm (RP; see Figure 1), in which on each trial a participant observes a perpetrator steal money from a victim ((Spezio, Brown, Peterson, Reimer, & Van Slyke, 2008); Figure 1). Briefly, on each round, the Participant (Observer) witnesses a Perpetrator stealing money from a Victim. The Participant has the option of helping or not helping the Victim. To help the Victim, the participant gives of her/his own money to make up for the amount stolen. Each time the participant helps the Victim, the probability that the Perpetrator will detect the participant increases. If the Perpetrator detects the

Participant, the Perpetrator steals all of the Participant's money. Thus, the RP tests participants' willingness to take action for others within a context of ambiguous threat/danger.

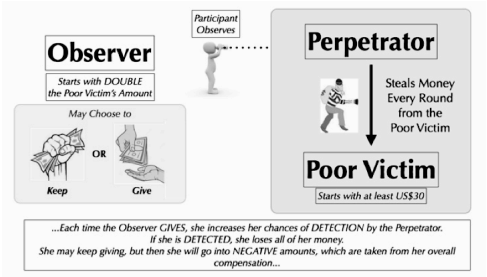

Figure 1: The Rescuer Paradigm. Participants observe money being stolen from an anonymous Victim and are given the choice of whether or not to help.

Study 1: Results

L'Arche USA members gave nearly 60% of the time, with half giving above this proportion of trials (Figure 2A). Several gave on every trial, and these are shown overlapping in the circle at the very top of the graph. Reports by the participants are consistent with the view that this level of giving was intentional and rational according to the participants' own value judgments. For example, one of the participants reported giving the *maximal* amount on *every other trial*, so as to balance a lower probability of detection with a high degree of caring action for the Victim. The loss ratios (loss to self divided by loss to other) prior to any detection events were greater than 1 for nearly all participants (Figure 2B), indicating more choices to give of one's own money than to allow the Victim to lose money.

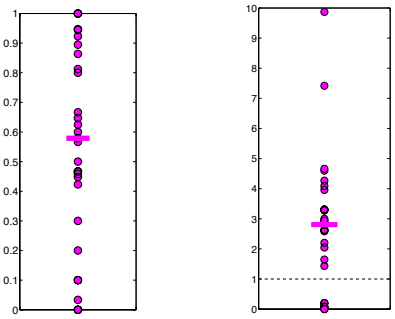

Figure 2. A (left). Proportion of "give" choices across the 30 trials, plotted for all participants. Horizontal bar is the median. B (right). Loss Ratio ([$Loss_{Self}$]/$Loss_{Victim}$]), plotted for all participants (3 points fall below a Loss Ratio of 1, shown by the dotted line) Horizontal bar is the median.

To determine how the L'Arche members represented their own losses with respect to the losses of the Victim, we compared four different logistic regression models in terms of fit to the trial-by-trial choices to keep or to give (0 v. 1) to the Victim: 1) Model 1, a model that included no contextual variables such as the amount stolen or how much the participant or the Victim had lost (i.e., Contextual Insensitivity); 2) Model 2, an additive model that separated loss to self from loss to the Victim as two predictor variables (i.e., simple Utilitarian); 3) Model 3, a model that tested for a multiplicative combination of loss to self and loss to Victim that lost all distinction between them (i.e., Nondual Self); and 4) Model 4, a model that offset loss to self and loss to Victim in a unified ratio (i.e., Relationally Nondual Self). RStan (RStan_Development_Team, 2014; Stan_Development_Team, 2014) generated Bayesian parameter estimates and model comparison used the WAIC statistic (Gelman et al., in press; Vehtari & Gelman, 2014; Watanabe, 2010). Almost all of our participants (78%) favored some form of joint self-other value representation, with 63% showing a ratio (Model 4) and 15% showing a multiplicative (Model 3) representation of joint value. Those who represented joint value as a ratio (Model 4) also gave more to the Victim (R^2 = 0.77).

Study 2

In the second study, 203 participants from the greater Los Angeles area completed one session of the Public Goods Paradigm (PGP) in groups of 10-12 people (15 rounds, $10 initial endowment per round which could be doubled to a $20 payout for the entire group if at least 25% of the group contributed). We defined a Giving group by selecting all participants who gave on at least 13 of 15 rounds, including the first and last round (N=17). Of the people who never gave or gave on at most 1 round, a Control group (N=17) matched the Giving group on self-reported age, gender, income, big 5 personality, empathy, and prosocial personality. After 2-3 years, participants from both groups returned to the laboratory to complete a 15-round Rescuer Paradigm.

Study 2: Results

After a delay of 2-3 years following group classification according to behavior on the PGP, participants in the Giving group gave a higher proportion of their money to the RP Victim (M ± SD = 0.45 ± 0.1; Figure 3A) compared to the Control group (0.08 ± 0.06; Figure 3B), demonstrating stable morally relevant decision making across behavioral contexts and extended periods of time. Cognitive valuational modeling followed by predictive model fitting using WAIC showed that most of the Giving group, but not the Control group, jointly represented losses to self and other as a unified ratio (Model 4).

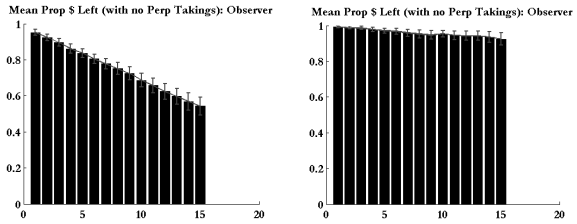

Figure 3: Trial-by-trial giving on the RP following a 2-3 year delay from the initial group classification on the PGP. A (left). Giving group. B (right). Control group.

Study 3

In the third study, 503 young adult participants recruited from Hamburg, Germany, and the surrounding area completed a 15-round Rescuer Paradigm as part of a large study of learning and decision making. The battery of testing materials included the DOSE assessment of risk and loss aversion (Wang, Filiba, & Camerer, 2010), the Portrait Values Scale (Schwartz, 2006; Schwartz & Boehnke, 2004), and the Temperament and Character Inventory (Cloninger, Svrakic, & Przybeck, 1993).

Study 3: Results

Only 413 of 503 participants chose to give to the Victim on at least 2 of 15 rounds. We found that self-report ratings associated only weakly with actual behavioral outcomes such as the proportion of giving or the loss ratio. We determined how participants represented their own losses and the losses to the Victim by again using the WAIC criterion compare each of the four models of their trial-by-trial choices. Most of the participants' choices were not fit best using a model that represented loss to self and Victim as a joint value. However, participants showing representation of joint value as a ratio of loss (Model 4) also showed a greater propensity to give to the Victim ($F(3,409) = 8.68$, $p < 0.0001$), compared to the context-free model (Model 1; $z = 3.02$, $p < 0.05$); the additive model (Model 2; $z = 2.00$, $p < 0.05$), and the multiplicative model (Model 3; $z = 4.34$, $p < 0.05$).

Discussion

We applied cognitive valuational modeling of choices on a "rescue decision" task, completed by two groups of participants demonstrating stable, caring action for others. The first group's stability of care came in the form of their long-term commitment to supportive relationships with adults who have neurodevelopmental disorders. Evidence for stable care in the second group came from measures over 2-3 years in different morally salient laboratory tasks. We found that when participants in both groups represented value jointly as a ratio of self and other (Model 4), action for others increased. Further, in a larger sample of young adults, we also saw that this ratio representation of joint value associated with a greater propensity to act on behalf of others, compared to models that represented value separately or multiplicatively.

Conclusions

Our findings show that action for others is facilitated by a "self for others," that is, a representation of value for self that is tied to the value for others, but that preserves a distinction between them. This finding opens up possibilities for interdisciplinary inquiry with deontological theories, with utilitarian theories, with theories emphasizing nondual self, and with theories holding to a relationally nondual self. We also show that laboratory methods and cognitive valuational modeling are relevant to understanding the moral cognition serving stable moral commitment in real life. Once this relevance is established and once the exemplary patterns of valuational representation are identified, those same methods can be applied to populations more generally.

Acknowledgments

We are grateful to Kathryn Aughtry, Andrea Beckam, and Catherine Holcomb for assistance in data collection and to James Van Slyke, Kristen Monroe, and Linda Zagzebski for helpful discussions. We gratefully acknowledge funding from the Science and Transcendence Advanced Research Series of the Center for Theology and the Natural Sciences in Berkeley, CA, from the Center of Theological Inquiry in Princeton, NJ, from the John Templeton Foundation (Grant 21338), and from the German Ministry of Research and Education to J.G. (01GQ1006).

References

Aquinas, T. (1964). *Commentary on Aristotle's Nicomachean Ethics* (C. I. Litzinger, Trans.). Notre Dame, IN: Dumb Ox Books.

Aristotle. (1992 (1925)). *Nicomachean Ethics* (W. D. Ross, Trans.). New York: Oxford University Press.

Bonhoeffer, D. (1998). *Sanctorum Communio: Theological Study of the Sociology of the Church* (R. Krauss & N. Lukens, Trans. Vol. 1). Minneapolis, MN: Fortress.

Bonhoeffer, D. (2005(1949)). *Ethics* (I. Todt, H. E. Todt, E. Feil & C. Green, Trans. Vol. 6). Minneapolis, MN: Fortress.

Cloninger, C. R., Svrakic, D. M., & Przybeck, T. (1993). A psychobiological model of temperament and character. *Archives of General Psychiatry, 50*(12), 975-990.

Frick, P. (2008). The Imitatio Christi of Thomas à Kempis and Dietrich Bonhoeffer. In P. Frick (Ed.), *Bonhoeffer's intellectual formation : theology and philosophy in his thought* (pp. 31-52). Tübingen: Mohr Siebeck.

Gelman, A., Hwang, J., & Vehtari, A. (in press). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*.

Gethin, R. (2011). On some definitions of mindfulness. *Contemporary Buddhism, 12*(1), 263-279. doi: 10.1080/14639947.2011.564843

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.

Greene, W. H. (2009). Discrete Choice Modeling. In T. C. Mills & K. Patterson (Eds.), *Palgrave Handbook of Econometrics* (Vol. 2, pp. 473-556). New York, NY: Palgrave MacMillan.

Heim, M. (2011). BUDDHIST ETHICS: A Review Essay. *Journal of Religious Ethics, 39*(3), 571-584.

Herman, B. (2007). *Moral Literacy*. Cambridge, MA: Harvard University.

Kant, I. (1996 (1798)). The Metaphysics of Morals (M. J. Gregor, Trans.). In M. J. Gregor (Ed.), *Practical Philosophy* (pp. 355-603). New York, NY: Cambridge University Press.

Kant, I. (2005). *Groundwork for the Metaphysics of Morals* (T. K. Abbot, Trans.). Orchard Park, NY: Broadview Editions.

Kruschke, J. K. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Burlington, MA: Academic Press.

Lopez, D. (2008). *Buddhism and Science: A Guide for the Perplexed*. Chicago, IL: University of Chicago.

Mazzanti, M. (2003). Discrete choice models and valuation experiments. *Journal of economic studies, 30*(6), 584-604.

Mill, J. S. (1871). *Utilitarianism*. London, UK: Longmans, Green, Reader & Dyer.

Peterson, G., Van Slyke, J., Spezio, M. L., Reimer, K., & Brown, W. S. (2010). The Rationality of Ultimate Concern: Moral Exemplars, Theological Ethics, and the Science of Moral Cognition. *Theology and Science, 8*, 139-161.

RStan_Development_Team. (2014). RStan: the R interface to Stan, Version 2.5. Retrieved from http://mc-stan.org/rstan.html.

Schwartz, S. (2006). Value orientations: measurement, antecedents, and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey*. London: Sage.

Schwartz, S., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality, 38*(3), 230-255.

Spezio, M. L., Brown, W. S., Peterson, G., Reimer, K., & Van Slyke, J. (2008). *Virtuous Decisions: Exemplarity in and out of the laboratory*. Paper presented at the Society for Neuroeconomics, Park City, Utah.

Stan_Development_Team. (2014). Stan: A C++ Library for Probability and Sampling, Version 2.5.0. Retrieved from http://mc-stan.org.

Vehtari, A., & Gelman, A. (2014). WAIC and Cross-Validation in Stan. http://www.stat.columbia.edu/~gelman/research/unpublished/waic_stan.pdf

Wang, S. W., Filiba, M., & Camerer, C. F. (2010). Dynamically optimized sequential experimentation (DOSE) for estimating economic preference parameters. http://people.hss.caltech.edu/~sweiwang/papers/DOSE.pdf

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research, 11*, 3571-3594.