Paper Title How to Make Moral Education More Effective: From a Brain Study to Policy Making

Author(s) Hyemin Han, The University of Alabama; Stephen J. Thoma, The University of Alabama; Firat Soylu, The University of Alabama; Kangwook Lee, KAIST

Session Title Moral Development and Education: Knowledge to Action

Session Type Paper

Presentation Date 4/27/2017

Presentation Location San Antonio, Texas

Descriptors Adolescence

Methodology Qualitative

Unit SIG-Moral Development and Education

DOI 10.302/1172082

How to Make Moral Education More Effective? From a Brain Study to Policy Making

Hyemin Han Stephen J. Thoma Firat Soylu

University of Alabama

Kangwook Lee

KAIST

Abstract

This paper suggests how to develop more effective moral educational programs by utilizing interdisciplinary research methods including neuroimaging, social psychological intervention, evolutionary modeling and deep learning methods. Our interdisciplinary research program consists of three steps: identification of core psychological processes involved in moral functioning at the neural level, development of small-scale intervention programs tweaking the identified psychological processes, prediction of long-term, large-scale outcomes of designed interventions using computational methods. We discuss how this research program will inform moral educators and educational policy makers who wish to develop effective moral educational programs in larger educational settings in the long term with actual neuroimaging and intervention experimental data.

Introduction

Recently, neuroscientists have conducted various experiments focusing on the neural correlates of human morality (Bzdok et al., 2012; Sevinc & Spreng, 2014). The findings have enhanced our understanding of the biological mechanism and substructure of moral functioning with evidence. However, it is still not clear how to apply the neuroscientific findings to moral education in practice and how to scale up the moral educational programs developed by psychological experiments in a lab in order to make them effectively work in a large scale, such

as a school or district level. Thus, we consider how to establish an interdisciplinary research program embracing neuroscience, experimental psychology, educational practice and policy making that can contribute to the improvement of moral educational in the reality (Han, 2016a). First, we discuss how neuroimaging methods including meta-analysis and functional magnetic resonance imaging (fMRI) can identify central neural mechanisms associated with moral functioning of interest, which will be targeted by moral educational interventions. Second, we demonstrate psychological experiments inspired by social psychology that are designed to target the psychological mechanisms identified by neuroimaging. Third, we discuss how to apply evolutionary modeling and computer simulation methods in order to develop appropriate interventions and inform educational policy makers. We demonstrate how the research program with these three procedures can produce applicable outcomes, the application of stories of moral exemplars in educational settings intending the promotion of moral motivation.

Utilizing Neuroimaging Methods

Neuroimaging methods can potentially contribute to the expansion of our knowledge regarding how human morality is functioning with biological evidence, which could not be approached through traditional research methods, such as paper-and-pencil survey (Ito & Cacioppo, 2007; Kristjánsson, 2013). However, how these scientific findings can eventually contribute to the development of moral education in practice is still unclear (Han, 2016b; Kristjánsson, 2015). Therefore, it would be necessary to consider how to connect the factual findings from neuroimaging studies to educational application in practice by establishing a valid interdisciplinary research program. As a first step in this process, we consider two specific neuroimaging methods, i.e., meta-analysis and fMRI, which can illuminate psychological processes involved in moral functioning, as components in the research program.

A meta-analysis of previously conducted neuroimaging studies can identify common activation foci of psychological functions of interest, moral functioning in case of our study. It can address several issues associated with traditional neuroimaging methods, such as the lack of statistical power originating from relatively small sample sizes, idiosyncrasies in experimental designs (Costafreda, 2009; Etkin & Wager, 2007; Wager, Jonides, & Reading, 2004) and possibility of reverse inference in interpretation (Poldrack, 2011). Several meta-analyses of previous neuroimaging studies focusing on moral functioning have demonstrated common activation foci associated with moral psychological processes (Bzdok et al., 2012; Han, 2017; Sevinc & Spreng, 2014). The common findings obtained by these meta-analyses provides a foundation for hypothesis setting in future fMRI studies (Han, 2016c). Taken together, the finding of the meta-analyses across diverse morality-related task conditions (see Fig. 1) found activation in brain regions associated with self-related processes, particularly autobiographical self and self-evaluation, that is, the default mode network (DMN) and cortical midline structures (CMS). Given these results, it is plausible that selfhood seems to commonly engage with moral functioning. Therefore, it would seem profitable to focus the next generation of fMRI experiments on examining psychological processes that will be targeted during intervention experiments.

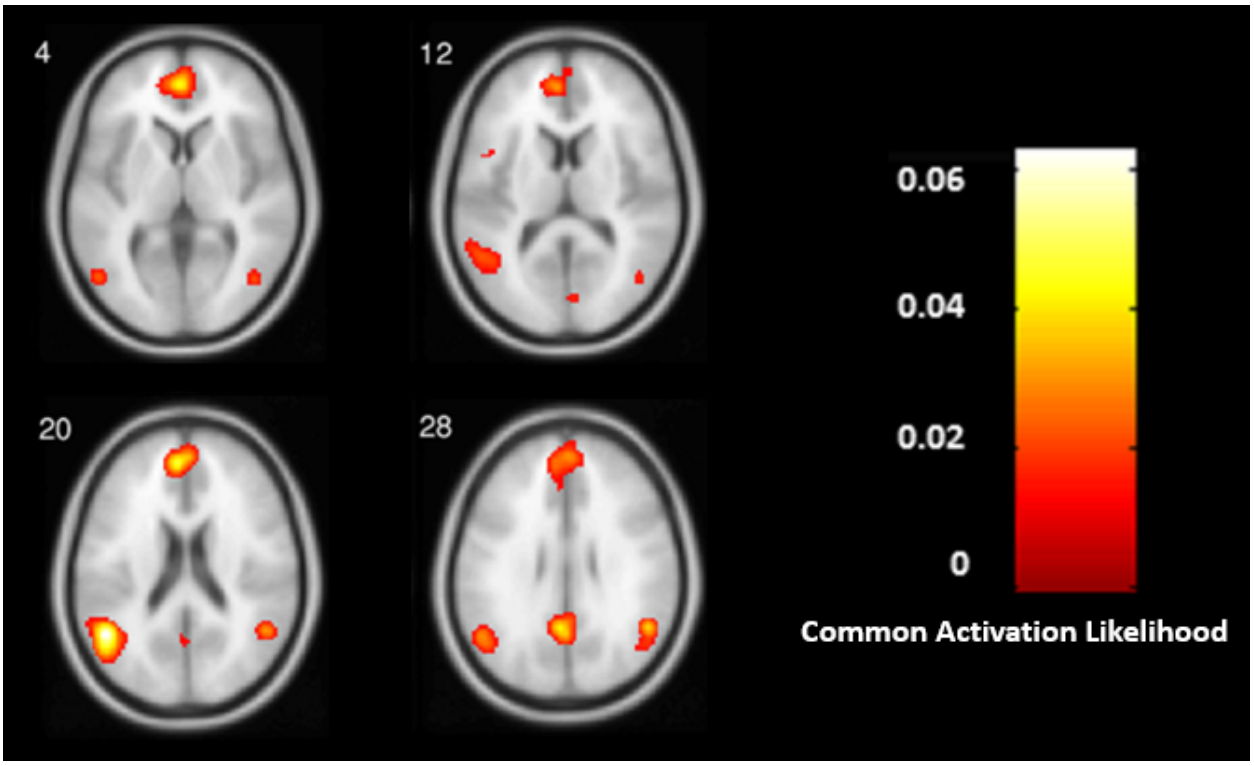

Figure 1. Common activation foci of moral functioning extracted from the meta-analysis.

In addition to the meta-analyses, fMRI experiments have demonstrated the neural correlates of moral function within diverse experimental paradigms (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Robertson et al., 2007). Particularly informative is, a recent fMRI experiment testing hypotheses based on the findings from the previous meta-analyses (Han, Chen, Jeong, & Glover, 2016). This fMRI experiment investigated how brain regions associated with selfhood, the DMN and CMN, moderated activity in other brain regions associated with moral emotion and motivation, such as the insula, while solving moral problems as assessed using the psychophysiological interaction analysis and Granger causality analysis methods (Friston et al., 1997; Seth, Barrett, & Barnett, 2015). Fig. 2 demonstrates the results of the analysis. As hypothesized, the analysis indicated that neural activity in regions associated with selfhood in the DMN and CMS, particularly the medial prefrontal cortex and posterior cingulate cortex, significantly moderated activity in moral emotion and moral motivation-related regions,

particularly the insula. Consequently, the fMRI experiment was able to support the hypotheses based upon previously published neuroimaging studies of morality and their meta-analyses and to identify psychological processes that will be targeted by intervention experiments, that is self-related psychological processes.

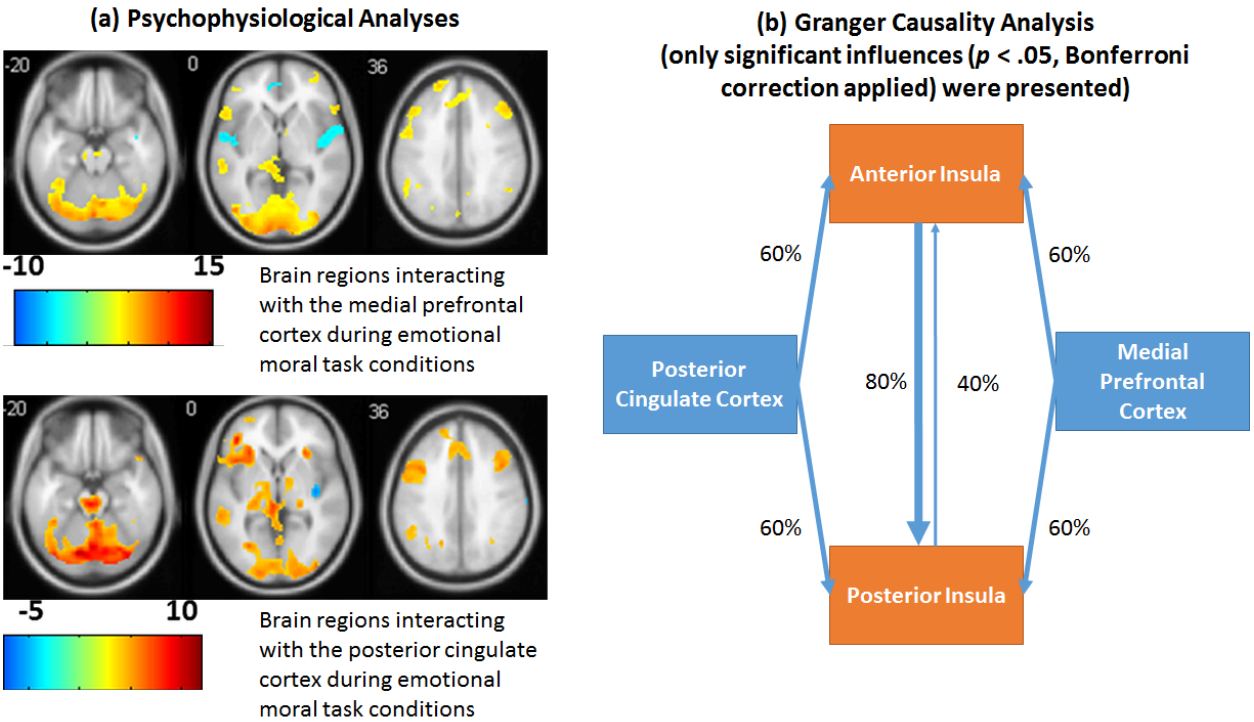

Figure 2. fMRI experiment results. (a) Psychophysiological analyses. (b) Granger causality analysis.

Psychological Intervention Methods

Interventions based on psychology, particularly social and educational psychology, have improved students' academic achievement and social adjustment in diverse educational settings (Cohen & Garcia, 2008; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009; Yeager, Trzesniewski, & Dweck, 2013; Yeager, Trzesniewski, Tirri, Nokelainen, & Dweck, 2011; Yeager & Walton, 2011). Thus, developed psychological intervention methods can provide

useful insights about how to design more effective moral education programs. Basically, psychological interventions are designed to tweak psychological processes that are fundamentally associated with the targeted developmental outcome (Miller & Prentice, 2012). Based on the purpose and nature of intervention research, two psychological intervention experiments using the stories of moral exemplars were conducted to test which type of exemplary stories better promoted motivation to engage in moral activity (Han, Kim, Jeong, & Cohen, 2017). Since the neuroimaging studies demonstrated that selfhood significantly moderated moral emotion and motivation, the experiments manipulated the perceived distance between presented moral exemplars and participants' self-concept.

The experiments presented two different types of exemplary stories: attainable and unattainable moral stories. Given the significant positive interaction between self-related and moral functioning-related brain regions, as the presented moral stories are perceived to be closely associated with participants' self-concept, the motivating effect of the stories would become greater (Walton, Cohen, Cwir, & Spencer, 2012); attainable stories (e.g., stories of peer exemplars) would more strongly promote motivation compared to unattainable stories (e.g., stories of historic figures), which seem distant from participants. In fact, previous social psychological intervention experiments focusing on non-moral motivation also reported that attainable stories better promoted motivation while unattainable stories might backfire (Cialdini, 1980; Lockwood & Kunda, 1997).

First, a lab experiment was conducted to examine the motivating effects of different types of moral stories among college students; it used participation in voluntary service activity as a proxy for moral motivation. This experiment was conducted among 54 Korean college students. They were randomly assigned to one of these three groups: attainable exemplar, unattainable

exemplar, and control group. Participants in the attainable exemplar group were presented with stories of college students who participated in voluntary service activities less than three hours per week. Those in the unattainable exemplar group were presented with stories of peers who engaged in service activities more than nine hours per week. These groups were requested to write a brief letter persuading their friends to participate in service activities while referring to the presented exemplary stories. Non-moral stories, such as sports news reports were presented to the control group.

Second, a classroom intervention experiment tested the same hypothesis among 8th graders after an eight-week intervention period. This classroom-level experiment was performed to apply the lab-level intervention to more realistic educational settings. A total of 187 Korean 8th graders participated in this experiment. They were randomly assigned to one of these three groups: peer exemplar, historic figure, and control groups. The interventions were conducted for eight weeks (1 hour / week). The peer exemplar group was requested to discuss moral virtues of peer moral exemplars, e.g., family members and friends. The historic figure group was asked to talk about moral behaviors of historic moral exemplars, e.g., Mother Teresa and Martin Luther King Jr. Students in the control group were not presented with any concrete moral exemplars.

In both experiments, pre-test and post-test (about two months later) voluntary service engagements were compared between conditions. The findings from these two experiments supported the hypotheses (see Fig. 3). Attainable and peer exemplars better promoted moral motivation compared to unattainable and historic moral exemplars; these findings are coherent with the neuroimaging experiments that contributed to the hypothesis setting and previous intervention studies.

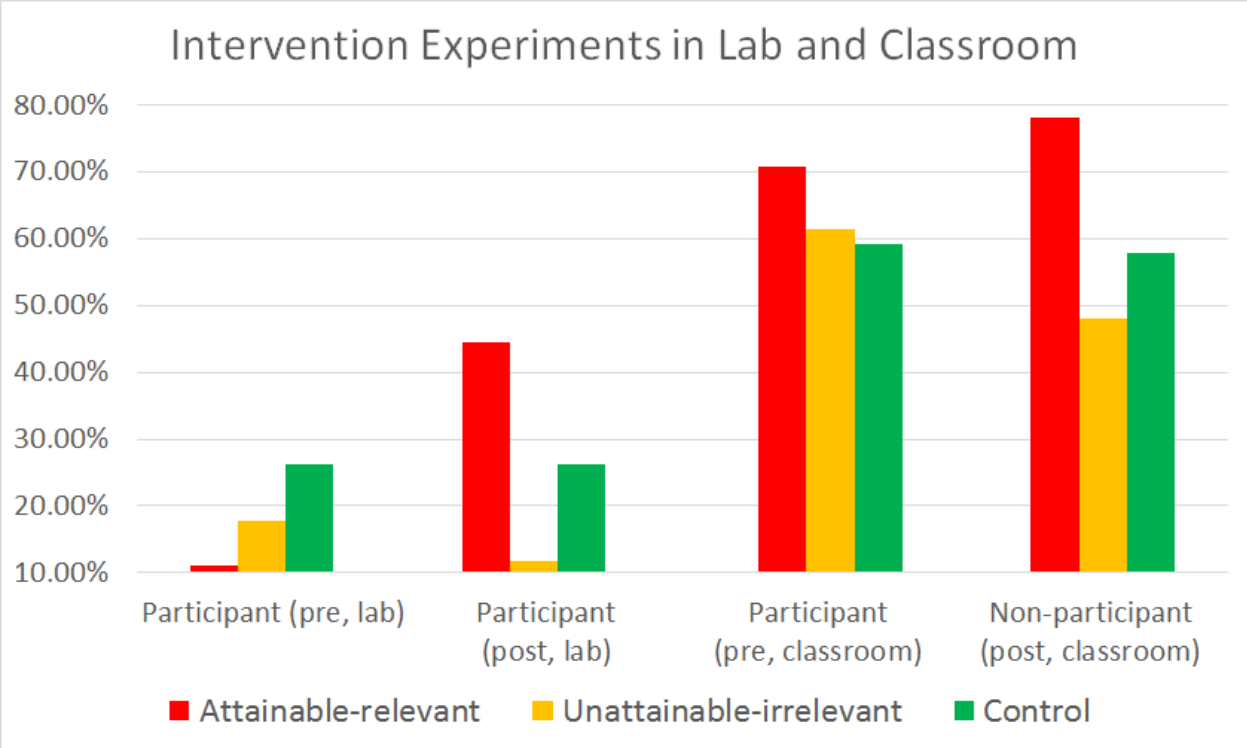

Figure 3. Results of two intervention experiments

Applying Evolutionary Modeling and Computer Simulation to Inform Educators and Policy Makers

In order to apply the intervention programs that were developed in a lab or classroom to wider contexts, such as districts, more long-term experiments with diverse experimental conditions should be performed. These additional experiments are necessary to determine the type and frequency of interventions. Since the application of intervention programs to larger-scale contexts, such as their application to curriculum, may result in significant long-term changes in students' development (Cohen et al., 2009; Yeager & Walton, 2011), these experiments would be important to educators and policy makers. However, due to the lack of time and resource, it would be difficult to conduct multiple long-term, large-scale experiments in real educational settings (Brown, 1992; Yeager, Fong, Lee, & Espelage, 2015). Given this

limitation, the evolutionary modeling and computer simulation predicting long-term outcomes of intervention programs can contribute to the large-scale implementation of developed interventions. For instance, predictions based on relatively small scale data can inform educators and policy makers of what kind of interventions would most effectively produce desirable developmental outcomes in larger educational settings in the long term.

The evolutionary modeling using the Evolutionary Causal Matrices (ECM) can predict the future status of a certain system consisting of different types of individuals (Claidière, 2009). The ECM predict the future status at t_0+n from the status change between t_0 and t_0+1 with iterative calculations; with n iterations, the predicted status at t_0+n can be calculated (Claidière, Scott-Phillips, & Sperber, 2014). In case of the moral educational intervention, we can set the t_0 status as the pre-test voluntary service engagement and t_0+1 as the post-test engagement. By performing iterative calculations, we can compare the effectiveness of interventions according to their types and application frequencies. As presented in Fig. 4, the attainable exemplar-applied intervention can better promote engagement. Its effect size declines as the frequency of application gets lower and the intervention should be performed at least once per every 18 months to produce a large effect (Han, Lee, & Soylu, 2016). We remark that the ECM-based prediction is useful at predicting future outcome sequences based on a simple stochastic model with a relatively small number of estimated parameters.

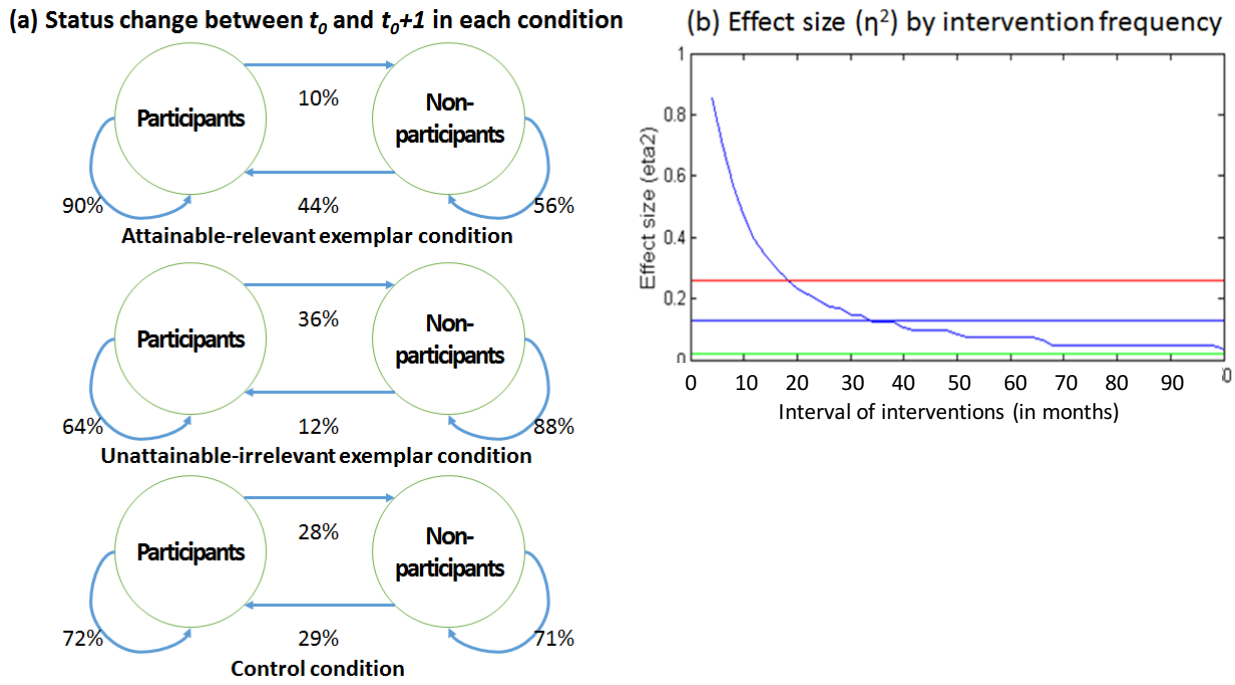

Figure 4. ECM-applied intervention outcome prediction. (a) Status change between t_0 and t_0+1 in each condition using ECM. (b) Effect size (η^2) by intervention frequency. Red line: threshold for large effect size (η^2 = .26). Blue line: threshold for medium effect size (η^2 = .13). Green line: threshold for small effect size (η^2 = .02).

When a large enough amount of training data is in hand, one may apply machine learning algorithms in order to develop a data-driven prediction model. Among various machine learning algorithms, artificial neural networks with many layers, or simply "deep learning", are the most popular choice of today due to its outstanding performance in many classical applications such as image classification, object recognition, speech recognition, etc. The deep architecture of deep learning corresponds to a hierarchy of features, factors, or concepts, where higher-level concepts are defined from lower-level ones, and the same lower-level concepts can help to define many higher-level concepts (Deng & Yu, 2014, p. 200). Using Google's TensorFlow (Abadi et al., 2015), we trained a two-layered convolutional network for a predicting intervention outcomes.

More precisely, our prediction network takes pre-test variables (i.e., service engagement, gender, intervention type, emotional responses to intervention activity, intention to engage in service) as inputs, and predicts post-test outcome indicator (i.e., whether or not engage in service at the post-test). We used an iterative training algorithm (called the stochastic gradient method), and we observed that the prediction performance is maximized after about 4000 iterations of the training algorithm.[1] The best prediction model with our convolutional network clearly outperforms simple logistic regression: while the accuracy of logistic regression was 75.47%, that of our best convolutional network reached 85.16%. These computer simulation models can provide researchers, educators and policy makers with a reliable way to predesign large-scale intervention experiments or applications.

[1] Note that the prediction performance decreases after a certain number of iterations. This is called overfitting, which happens when a prediction model starts capturing in the model noise of the data, losing predictability.

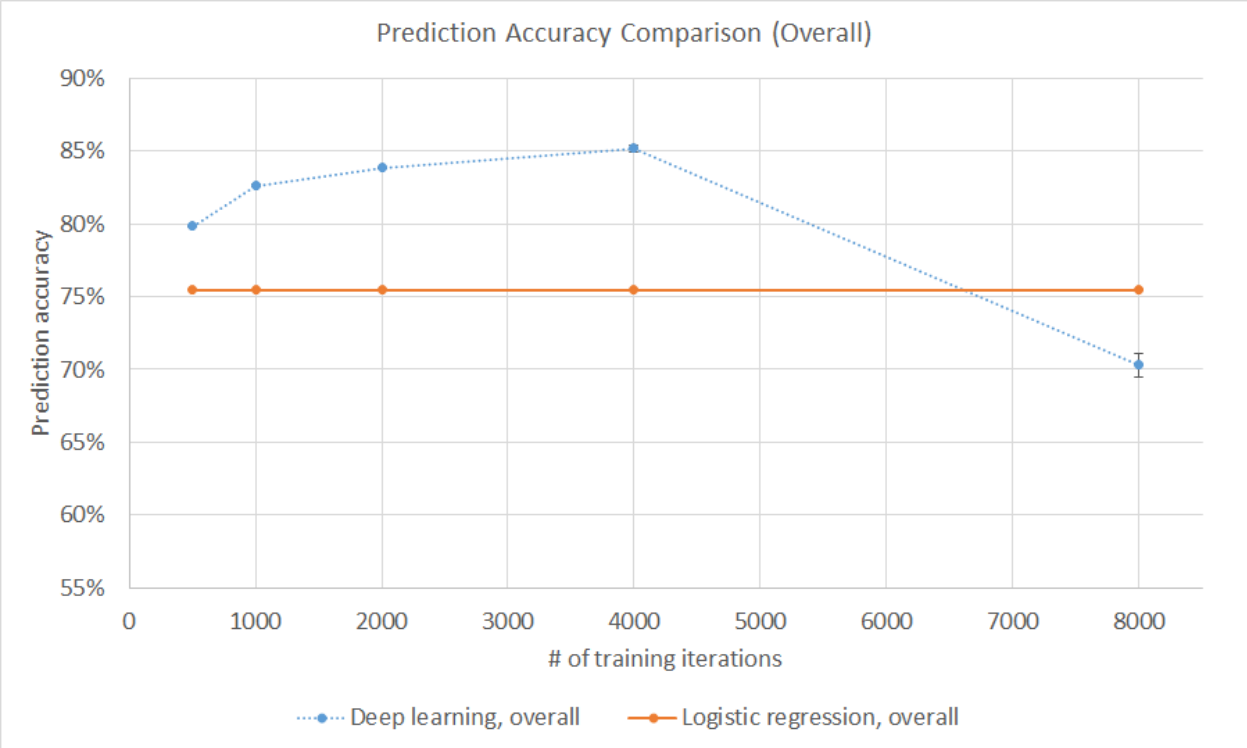

Figure 5. Prediction accuracy by different learning iteration.

Concluding Remarks

If and how findings from neuroscience studies can be beneficial for educational practice and policy making has been hotly debated in recent years (Ansari & Coch, 2006; Hruby, 2012; Immordino-Yang et al., 2007). Following the outline provided above we encourage the field to triangulate the information provided by lab studies, classroom interventions, computational modeling and meta-analyses of neural mechanism studies, to develop informed interventions based on empirically supported practices and procedures. Additionally, we see our approach using multiple empirical criteria for determining successful moral education interventions as providing a foundation for more general educational policy about what works in the moral domain and why. Taking these ideas a step further, we see our multi-source, multi-disciplinary approach as a blueprint for educational practice beyond the moral domain because it explicitly

heeds the call to build a two-way bridge between the lab and the classroom, and between neuroscience and education (Mason, 2009; Turner, 2011).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous distributed systems*. Mountain View, CA.

Ansari, D., & Coch, D. (2006). Bridges over troubled waters: education and cognitive neuroscience. *Trends In Cognitive Sciences*, *10*(4), 146–151.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178. doi:10.1207/s15327809jls0202_2

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, *217*, 783–796. doi:10.1007/s00429-012-0380-y

Cialdini, R. B. (1980). Full-cycle social psychology. *Applied Social Psychology Annual*, *1*, 21–47.

Claidière, N. (2009). *Darwinian theories of cultural evolution: models and mechanisms*. Ecole Normale Supeieure & Universite Pierre et Marie Curie.

Claidière, N., Scott-Phillips, T. C., & Sperber, D. (2014). How Darwinian is cultural evolution? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *369*(1642), 20130368. doi:10.1098/rstb.2013.0368

Cohen, G. L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions,

implications. *Current Directions in Psychological Science*, *17*(6), 365–369.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, *324*(5925), 400–403. doi:10.1126/science.1170769

Costafreda, S. G. (2009). Pooling FMRI data: meta-analysis, mega-analysis and multi-center studies. *Frontiers in neuroinformatics*, *3*, 33. doi:10.3389/neuro.11.033.2009

Deng, L., & Yu, D. (2014). Deep learning: Methods and applications. *Foundations and Trends® in Signal Processing*, *7*(3–4), 197–387. doi:10.1561/2000000039

Etkin, A., & Wager, T. D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *The American journal of psychiatry*, *164*, 1476–1488. doi:10.1176/appi.ajp.2007.07030504

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–29. doi:10.1006/nimg.1997.0291

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. doi:10.1016/j.neuron.2004.09.027

Han, H. (2016a). Attainable and relevant moral exemplars as powerful sources for moral education: From vantage points of virtue ethics and social psychology. In *Invited Oral Presentation at the 4th Annual Conference of the Jubilee Centre for Character and Virtues*. Oxford, UK.

Han, H. (2016b). How can neuroscience contribute to moral philosophy, psychology and education based on Aristotelian virtue ethics? *International Journal of Ethics Education*,

1(2), 201–217. doi:10.1007/s40889-016-0016-9

Han, H. (2016c). *Neuroscientific and social psychological investigation on psychological effects of stories of moral exemplars*. Stanford University.

Han, H. (2017). Neural correlates of moral sensitivity and moral judgment associated with brain circuitries of selfhood: A meta-analysis. *Journal of Moral Education*. doi:10.1080/03057240.2016.1262834

Han, H., Chen, J., Jeong, C., & Glover, G. H. (2016). Influence of the cortical midline structures on moral emotion and motivation in moral decision-making. *Behavioural Brain Research*, *302*, 237–251. doi:10.1016/j.bbr.2016.01.001

Han, H., Kim, J., Jeong, C., & Cohen, G. L. (2017). Attainable and Relevant Moral Exemplars Are More Effective than Extraordinary Exemplars in Promoting Voluntary Service Engagement. *Frontiers in Psychology*, *8*, 283. doi:10.3389/fpsyg.2017.00283

Han, H., Lee, K., & Soylu, F. (2016). Predicting Long-term Outcomes of Educational Interventions Using the Evolutionary Causal Matrices and Markov Chain Based on Educational Neuroscience. *Trends in Neuroscience and Education*, *5*(4), 157–165. doi:10.1016/j.tine.2016.11.003

Hruby, G. G. (2012). Three requirements for justifying an educational neuroscience. *The British journal of educational psychology*, *82*(Pt 1), 1–23. doi:10.1111/j.2044-8279.2012.02068.x

Immordino-Yang, M. H., Damasio, A., Immordino-Yang, M., Damasio, A., Immordino-Yang, M. H., & Damasio, A. (2007). We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. *Mind, Brain, and Education*, *1*(1), 3–10. doi:10.1111/j.1751-228X.2007.00004.x

Ito, T. A., & Cacioppo, J. T. (2007). Attitudes as mental and neural states of readiness: Using

physiological measures to study implicit attitudes. In B. W. N. Schwarz (Ed.), *Implicit measures of attitudes* (pp. 125–158). New York, NY, US: Guilford Press.

Kristjánsson, K. (2013). *Virtues and vices in positive psychology: A philosophical critique*. New York, NY: Cambridge University Press.

Kristjánsson, K. (2015). *Aristotelian character education*. London, UK: Routledge.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology*, *73*, 91–103. doi:10.1037/0022-3514.73.1.91

Mason, L. (2009). Bridging neuroscience and education: a two-way path is possible. *Cortex; a journal devoted to the study of the nervous system and behavior*, *45*(4), 548–9. doi:10.1016/j.cortex.2008.06.003

Miller, D. T., & Prentice, D. A. (2012). Psychological levers of behavior change. In E. Shafir (Ed.), *The Behavioral Foundations of Public Policy* (pp. 301–309). Princeton, NJ: Princeton University Press.

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. doi:10.1016/j.neuron.2011.11.001

Robertson, D., Snarey, J., Ousley, O., Harenski, K., Bowman, E. D., Gilkey, R., & Kilts, C. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*(4), 755–766. doi:10.1016/j.neuropsychologia.2006.08.014

Seth, A. K., Barrett, A. B., & Barnett, L. (2015). Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, *35*(8), 3293–3297. doi:10.1523/JNEUROSCI.4399-14.2015

Sevinc, G., & Spreng, R. N. (2014). Contextual and perceptual brain processes underlying moral

cognition: A quantitative meta-analysis of moral reasoning and moral emotions. *PLoS ONE*, *9*, e87427. doi:10.1371/journal.pone.0087427

Turner, D. a. (2011). Which part of "two way street" did you not understand? Redressing the balance of neuroscience and education. *Educational Research Review*, *6*(3), 223–231. doi:10.1016/j.edurev.2011.10.002

Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage*, *22*, 1679–1693. doi:10.1016/j.neuroimage.2004.03.052

Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of personality and social psychology*, *102*(3), 513–532. doi:10.1037/a0025731

Yeager, D. S., Fong, C. J., Lee, H. Y., & Espelage, D. L. (2015). Declines in efficacy of anti-bullying programs among older adolescents: Theory and a three-level meta-analysis. *Journal of Applied Developmental Psychology*, *37*, 36–51. doi:10.1016/j.appdev.2014.11.005

Yeager, D. S., Trzesniewski, K. H., & Dweck, C. S. (2013). An implicit theories of personality intervention reduces adolescent aggression in response to victimization and exclusion. *Child Development*, *84*(3), 970–988. doi:10.1111/cdev.12003

Yeager, D. S., Trzesniewski, K. H., Tirri, K., Nokelainen, P., & Dweck, C. S. (2011). Adolescents' implicit theories predict desire for vengeance after peer conflicts: Correlational and experimental evidence. *Developmental Psychology*, *47*(4), 1090–1107.

Yeager, D. S., & Walton, G. M. (2011). Social-psychological interventions in education: They're not magic. *Review of Educational Research*, *81*(2), 267–301. doi:10.3102/0034654311405999