Cognitive Attraction Theory and Moral Judgment

Zohair Chentouf

Department of Computer Science, King Saud University, Riyadh, KSA
Email: zchentouf@ksu.edu.sa

The present article deals with the processes that underpin moral judgment. In the specialized literature, some concepts are proven to be important mechanisms that build up the moral judgment. For instance, intuition, emotion, reasoning, moral rules, deontology and consequentialism. However, there is a lack of a comprehensive framework, which puts together those key concepts in a clear picture. The present article argues for a more comprehensive view under the light of the Cognitive Attraction Theory (CAT). The derived framework considers emotion as the main facilitator of moral judgment as it acts as the means of conceptual attraction between the different cognitive entities, including moral beliefs and rules. Based on this principle, we show how the moral judgment "evolves" from a moral intuition, sometimes endorsed by a reasoning fallacy, to an elaborated judgment that is a result of a conscious reasoning. With the help of a computer simulation performed with an artificial moral agent that incarnates a computational model of CAT, we show that moral judgment can be deontological or consequentialistic.

Keywords: Cognitive Attraction Theory; Moral Judgment; Emotion; Computational Model; Artificial Moral Agent

Introduction

Understanding moral judgment mechanisms is an active research endeavor (Baron & Spranca, 1997; Blair, 1995; Cushman et al., 2006; Greene et al., 2008; Gross, 1998; Nichols & Mallon, 2006; Oatley & Johnson-Laird, 1987; Schwarz & Clore, 1983). According to Bartels (2008), most proposed frameworks cross-pollinate between two concepts: deontology and consequentialism. The former states that moral judgment obeys moral rules (Darwall, 2003b; Davis, 1993). The latter asserts that it conforms to balancing costs and benefits (Darwall, 2003a; Pettit, 1993). An attempt of integrating the two theories of deontology and consequentialism into one framework has been reported in (Nichols & Mallon, 2006; Nichols, 2002). The proposed framework involves three processes: cost-benefit analysis, checking actions against moral rules, and emotional reactions. According to this account, moral cognition depends on an affect-backed normative theory. If rules that are backed by affect are used, the moral judgment is deontological; if the rules are not backed by affect, the judgment is consequentialistic. However, this account does not explain the moral judgment processes. Under the umbrella of deontology and consequentialism, some authors tried to explain how moral judgment is performed. For example, according to Haidt (2001, 2007) moral evaluations belong to immediate intuitions and emotions. The process is more akin to perception than reasoning in the sense that moral judgments suddenly appear in consciousness, including an affective valence, without any conscious awareness of reasoning. That is to say, moral intuitions (including moral emotions) come first and then cause moral judgments. According to Hauser (Hauser, 2006), moral judgments are the result of a conscious process that uses an innate moral grammar. After judgments are inferred, they trigger emotions. So, emotions come after moral judgments.

All these accounts highlight important concepts: intuition, emotion, reasoning, moral rules, and consequentialism. However, a more comprehensive one has to be elaborated. In particular, it should clearly describe moral judgment processes and determine the exact roles and interactions of those concepts. Many of these aspects of moral judgment have been addressed by the model of Haidt (2001, 2007). However, the here reported model based on CAT considers moral judgment as the product of two consecutive processes: a moral intuition sometimes directed by a reasoning fallacy, then a conscious moral reasoning. CAT considers these two processes as emergent from cognitive attraction mechanisms. This aspect will be detailed in the sequel.

In the rest of this paper we present our Cognitive Attraction Theory (CAT). Second, we show how this theory describes in detail the processes of moral judgment in a manner that accurately integrates the different concepts and properties of moral judgment that have been highlighted by the literature. Third, we demonstrate the potential of CAT through the analysis of the moral judgments of an artificial moral agent, which incarnates a partial computational model of CAT. Finally, we conclude the paper and highlight future direction.

Cognitive Attraction Theory

The here called Cognitive Attraction Theory (CAT) is part of the Recurrence Theory (RT) (Chentouf, 2000; Chentouf, 1997). RT is an onto-epistemological theory, which claims the existence of fundamental principles that rule biomolecular reactions, cells, neurons, human cognition, and human society. CAT is the part of RT that explains how cognition characteristics emerge from neuronal ones. Among the main findings of CAT is that moral judgment is leveraged by the same process that occurs during perceptual illusion (Chentouf, 2000: pp. 146-187). However, the moral judgment process is more complex than the perception's one since it involves emotions, intuitions, and

reasoning.

CAT adheres to the associanistic psychological theory. So, for CAT, mental entities (concepts, feelings, ideas, etc.) are stored in the brain as an associative network. Associations reflect the relations that exist between mental entities (Warren, 1921). For example, a friend's name and his face are memorized as having a link so that evoking the name immediately evokes the face. For CAT, mental associations are ruled by the Hebb's rule. Recall that according to this rule the efficacy of a synaptic link between two neurons increases when there is a repeated and persistent stimulation (Paulsen, 2000). CAT considers that synaptic links engender conceptual attractions between mental entities. Based on this central concept, CAT can be formulated as follows.

Mental Entities

Following the lead of folk psychology (Stitch, 1994; Wellman, 1990), RT considers the belief-desire-intention model (Chentouf, 2000; Chentouf, 1997). To model moral judgment, we need to differentiate those three concepts into the following mental entities: beliefs, moral beliefs, goals, actions, and intentions. The model also involves the concepts of events, moral rules, and emotions. For the latter two concepts we adopt the same definitions as in (Haidt, 2007). According to the folk psychology, beliefs are memorized facts like "the sky is blue". Actions may be general like "eating" or specific like "eating fish". Planned actions are intentions, e.g. "eating fish tonight". Longer run and more persistent planned actions are goals, e.g. "feeding oneself" or "getting a BSc in Computer Science". Events are propositions that relate something that has been perceived. For example, "the baby is crying". We conjecture that moral beliefs are the association of actions, intentions, and goals with either good (G) or bad (B). For example, "lying to teacher is bad" is a moral belief. When the action is a general one, the moral belief represents a moral rule, e.g. "lying is bad". Emotions are either positive (P), like happiness, or negative (N), like sadness. CAT does not deal with the source of emotions and the appraisal processes. Suffice it here to say that emotions are the result of the subjective cognitive evaluation of events, intentions, and goals (Roseman & Smith, 2001). All moral beliefs, goals, and intentions are associated with emotions and the latter have intensities. Some events are also associated with emotion, for example, "a car accident on the road". Other events like "hearing a colleague closing his office door" may evoke no emotion.

Cognitive Attraction Principle

The associanistic attitude of CAT comes from the concept of cognitive attraction. The latter means that a mental entity X may cognitively attract an entity Y. Such an attraction takes place if there is a conceptual relationship R between X and Y. R may be "implies", "causes", "similar to", "contradicts", "part of", etc. Besides, the strength of the attraction performed by X on Y is proportional to the X's emotional intensity and the degree of evidence of X, Y, and R. For example, the intention "cheating-in-exam" is attracted by the moral rule "cheating-is-bad". R represents analogy. The intensity of the negative emotion associated with "cheating-is-bad" and the degree of evidence of the analogy will determine the strength of the attractive link between the two entities.

Aberration Principle

At the neural level, cognitive attraction is materialized by the strength of the synaptic links. Thus, generalizing the Hebbian rule, we can say that cognitive attraction links and their strengths are dynamic. A new mental entity X (created after an event, for example) is immediately attracted by other mental entities. During this process, attraction links and intensities can be changed and new mental entities might be created. After attraction forces have stabilized, X may have been altered because of the impact of the attracting forces. For instance, the intention "I will cheat in the next exam" may be attracted by moral rules, beliefs, and goals. At the end, it might be altered to: "I will cheat in the next exam but never after that". Besides, the negative (N) and positive (P) emotions of the attracting entities are propagated to X. When all links are stabilized, X will have the emotion of the stronger attractor. If strong but emotionally antagonistic attractors have the same attraction strength, their simultaneous attraction assigns to X a combination of incompatible emotions, which engenders a contradictory feeling that is sometimes called hesitation.

Moral Judgment as Cognitive Attraction Process

The moral judgment problem can now be formulated under the light of CAT:
- The attraction made by a moral rule or moral belief Y = (y is v) on a new mental entity x will infer a new moral belief X' = (x is v), v ∈ {G, B}. Recall that, as stated earlier, attraction takes place only if R(Y, x).
- If there are more than one moral attractor and they are antagonistic or incompatible, there may be inference of more than one moral belief. In such a case, we call moral conflict the fact that both (x is G) and (x is B) are inferred.
- Moral judgment is the process of inferring new moral beliefs and solving moral conflicts.

For example, the attraction performed by the moral rule "cheating is B" on the intention "cheating-in-exam" will create a moral belief "cheating-in-exam is B". The attraction performed by "cheating-in-exam-implies-parents-happy is G" infers "cheating-in-exam is G". This is a moral conflict.

Let us now apply the CAT attraction and aberration processes on the moral judgment problem in order to derive a moral judgment process. In the sequel, moral beliefs, goals, and intentions will be denoted as 3-tuple (X, m, j) and beliefs as 2-tuple (X, m), where m ∈ {P, N} and j ∈ {G, B}. For example, (lying, P, G) is a moral belief, and (sky-is-blue, P) is a belief. The moral judgment process contains the following sequence of sub-processes.

Intuition

First, an event is received, e.g. "my classmates cheated in the exam and they got A; I did not and therefore got C". To simplify, let us write this event: "cheating-in-exam". First, such an event may be attracted by the goal (success, P, G) which has an intense positive emotion. At the early steps of moral intuition, only emotions are taken into consideration. Consequently, there is inference of (cheating-in-exam, P) on which the emotion of (success, P, G) has been propagated. At later steps of moral intuition, a moral judgment is intuitionally derived. We conjecture that, at the moral intuition stage, simple and inaccurate mechanisms are used. For example, identifying G as synonym

of P (what is agreeable is always good) and B as synonym of N (what is painful is always bad). For our example, there will be creation of (cheating-in-exam, P, G). Interestingly, this may be seen as the Appeal-to-Consequences fallacy (Gass, 2012): (cheating-in-exam implies success), success is G > cheating-in-exam is G.

Conflict

The subject may figure out the existence of a moral conflict. This happens because of the attraction made by the moral rule (cheating, N, B) on the event "cheating-in-exam", which will propagate a negative emotion N and a B judgment. This results in: (cheating-in-exam, N, B), which contradicts the result of the moral intuition: (cheating-in-exam, P, G). Such a conflict needs to be solved.

Aberration

To solve the moral conflict, attractions made by other mental entities (moral beliefs and goals) on both of the conflicting judgments will be evoked to support the conflicting entities. Their strengths (degree of evidence and emotion intensities) will play a key role. During this process, conscious reasoning will be used. As a consequence, aberration will take place: emotion of some entities will be increased or decreased, new entities will be created by analogy, deduction, etc. The result of these attractions and aberrations constitutes a "heuristic that mediates between events, beliefs, and goals" (Oatley, 1999). For example, the belief (dad-cheated-many-times, P, G) supports (cheating-in-exam, P, G) by attracting it and propagating its own positive emotion intensity. The conscious reasoning may use fallacies. Referring to the list of fallacies in (Gass, 2012), **Table 1** shows other beliefs and the corresponding fallacies that can be used to support (cheating-in-exam, P, G). Support to (cheating-in-exam, N, B) can come from (cheating-in-exam-implies-punishment, N) and (cheating-in-exam-might-not-imply-success, N). The former belief evokes bad consequences, which corresponds to the fallacy Bad-Consequence (Gass, 2012), and the latter corresponds to relativation, which decreases the evidence of the causal relation between cheating and success.

Final Judgment

Attraction ends by assigning either B or G to the new entity "cheating-in-exam", i.e. concluding either (cheating-in-exam, P, G) or (cheating-in-exam, N, B). The moral judgment might result in updating the intensity and kind of emotion assigned to some of the moral beliefs and moral rules that have been involved in evaluating the new entity. For instance, there may be a conclusion of (cheating, P, G), which means that cheating is absolutely good and desirable. There may also be conclusion of (cheating-in-exam, P, B), which means that cheating in the exam is still emotionally positive although it is judged as morally bad.

Figure 1 recapitulates the above example. To demonstrate the expressive power of CAT, **Figure 2** illustrates the trolley dilemma (Foot, 1997), which is the most intensively studied one. In the bystander version of this dilemma, one has to imagine an approaching train that will kill five railway workers, and the only way to prevent their death is to flip a switch that will

Table 1.
Fallacies and corresponding beliefs.

Fallacy	Beliefs
Authority appeal to popularity	Dad cheated many times Most students cheat
Hypostatization	Success is more important than everything else
Bifurcation	One has to choose either success or failure
Good Consequences	Cheating in exam implies parents happy Cheating in exam implies less effort Cheating in exam implies good job and salary

```
Event:
        (cheating-in-exam-implies-success)
Intuition:
        (success, P, G)
        (cheating-in-exam, P, G)
Conflict:
        (cheating-in-exam, P, G), (cheating, N, B)
Support to (cheating-in-exam, P, G):
        (dad-cheated-many-times, P, G)
        (most-students-cheat, P, G)
        (cheating-in-exam-implies-parents-happy, P, G)
        (cheating-in-exam-implies-less-effort, P)
        (cheating-in-exam-implies-good-job-salary, P)
Support to (cheating-in-exam, N, B):
        (cheating-in-exam-implies-punishment, N)
        (cheating-in-exam-might-not-imply-success, N)
Conclusion:
        (cheating-in-exam, P, G) or
        (cheating-in-exam, N, B)
```

Figure 1.
The cheating dilemma.

```
Event:
        (trolley-about-to-kill - 5 - and-I-can-save-them)
Intuition:
        (save - 5, P, G)
        (kill - 1 - save - 5, P, G)
Conflict:
        (kill - 1 - save - 5, P, G), (kill - 1, N, B)
Support to (kill - 1 - save - 5, P, G):
        (5 lives > 1 life, P)
        (become-hero, P)
Support to (kill - 1, N, B):
        (killing, N, B)
        (cannot-have-it-on-shoulders, N)
        (trouble-with-police, N)
Conclusion:
        (kill - 1 - save - 5, P, G) or (kill - 1 - save - 5, N, B)
```

Figure 2.
The trolley dilemma.

divert the train but will cause the death of one person on the side track. The derived decision is either (kill - 1 - save - 5, P, G) or (kill - 1 - save - 5, N, B), which mean: "I accept to divert the train to cause the death of one person and save five lives because this is morally good", or "I do not accept to divert the train to cause the death of one person and save five lives because this is morally bad", respectively.

Experimenting with a Partial Computational Model of CAT

Method

A partial computational model of CAT has been imple-

mented as artificial moral agent called Zed. The simulation involves Zed and another hypothetical agent, Zee. In their initial states, Zed and Zee are waiting for food. The food comes to them at random intervals of time. When the food comes, Zed has to decide whether to leave it for Zee or harm Zee in order to take it for himself. Sharing food or taking it without harming the other is not possible. Zee is programmed to randomly harm Zed and prevent him from eating. The simulation is done from the point of view of Zed. Zee has not been implemented as an agent, only his random harming actions have been presented as stimuli to Zed. Beliefs of Zed are: (hunger, N), (like-Zee, P, G), (Zee-harmed-me, N, B), (Zee-not-harmed-me, P, G). Every time there is food, an event is received by Zed. It may either be "Zee-harmed-me" or "Zee-not-harmed-me". The latter means that Zee will not be able to harm Zed until the next time foods will come and that Zed still can harm Zee. The former implies that Zed cannot harm Zee until the next time the food will come. The event is then processed by the attraction module based on the algorithm presented in **Figure 3**. For example, (like-Zee, −) means to decrease the emotion strength of "like-Zee" and (hunger, +) means to increase the emotion strength of "hunger". For this aim, an emotional sensibility parameter has been used. The sensibility of "hunger", "harmed-me", "not-harmed-me", and "like-Zee" are arbitrarily chosen to be .2, .2, .1, and .1, respectively. This implies that increasing or decreasing the emotion intensity of "hunger", for instance, means to add or subtract .2, respectively. When it is still possible for Zed to take food for himself, i.e. there is no event "Zee-harmed-me", Zed has to solve the intention moral conflict: (harm-Zee, P, G), (harm-Zee, N, B) and then perform the action "harm-Zee" or "not-harm-Zee" accordingly. The inference of "not-harm-Zee" means that Zed leaves the food for Zee despite he was able to take it. The inference of "harm-Zee" means that Zed harms Zee and takes the food even though Zee does not show any interest for the food. Zed's decision is also processed by the attraction module (see Algorithm 1).

Measures

The first four columns of **Table 2** contain the initial emotion strengths (attraction strengths). Negative values correspond to negative emotions, and positive values correspond to positive emotions. The aim of the experiments is to show the impact of the emotion strength on the final moral judgment of Zed.

Procedure

The simulations (lines in **Table 2**) have been fed with the same set of random input events. The first line represents the baseline with which all the other experiments have been compared. Every experiment has consisted to vary a single emotion initial strength (highlighted values). For example, in experiment 2 (line 2), the aim was to see the behavioral effects of deviating

```
If Zee harmed me then
        (hunger, +), (harmed-me, +), (like-Zee, −)
If Zee not-harmed me then
        (not-harmed-me, +), (like-Zee, +)
If I harmed Zee then
        (hunger, −), (like-Zee, −)
If I not-harmed Zee then
        (like-Zee, +)
```

Figure 3.
Attraction process.

Table 2.
Moral judgments made by Zed.

	Beliefs				Conflict	
	Hunger	Zee harmed me	Zee not harmed me	Like Zee	Harm Zee	Not harm Zee
1	−.5	−.7	.1	.2	15	15
2	**−.8**	−.7	.1	.2	22	8
3	−.5	**−.8**	.1	.2	19	11
4	−.5	−.7	**.5**	.2	5	25
5	−.5	−.7	.1	**.5**	7	23

the "hunger" intensity from the baseline.

Results

The last two columns of **Table 2** correspond to the number of times Zed harmed or did not harm Zee. The first line presents the model calibration which has been considered as baseline since the numbers of "harm" and "not-harm" are equal (15 times for each action). Each of the subsequent experiments consisted to vary one parameter and compare results (number of times Zed harmed Zee) with the baseline. In all the four experiments, raising the emotion related to own needs (hunger and safety, lines 2, 3) renders Zed more selfish and aggressive. Indeed, Zed harmed Zee 27% - 47% more times compared with the baseline behavior (19 compared with 15, and 22 compared with 15, respectively). This is a typical consequentialistic behavior. Raising the emotion assigned to Zee's kindness (line 4) and sentiments to Zee (line 5) makes Zed display a more altruistic and less aggressive behavior. Results show that aggressions decreased by 53% - 67% (7 compared with 15, and 5 compared with 15, respectively). Such a behavior is deontological.

Discussion

As a result, we can say that deontology and consequentialism are nothing but the consequence of attraction either performed by the moral rules or by the beliefs that contradict moral rules, respectively. To check the consistency of Zed's moral behavior, we repeated the experiments with different baselines; all results conformed to the pattern drawn by **Table 2**.

Interestingly, the behavior of Zed is predictable by the common sense. Moreover, it completely conforms to results of other studies. For instance, a study by Arsenio and Gold (2006) revealed that "difficulties in early parent-child interactions in combination with hostile larger social environments act to undermine emotional reciprocity, empathy, and concern for others in ways likely to promote proactive, uncaring forms of victimization and harm". Similar facts are related by Bartels and Pizarro: "Participants who indicated greater endorsement of utilitarian solutions had higher scores on measures of psychopathy, machiavellianism, and life meaninglessness" (Bartels & Pizarro, 2011).

The main limitation of the experiments is due to the fact that the computational model is a partial implementation of CAT; it does not comprehensively validate the proposed model. Besides,

attraction strengths and emotional sensibility values are not based on real experimental data.

It is worth noting an interesting point revealed by Zed: the fact that he successfully made decisions (harm Zee or not). This encourages us to embed in our future computational model the faculty of decision. Obviously, the decision problem involves more concepts, like acceptance, which means that a subject may judge immoral an action but accepts to do it. But we think that the main concepts of CAT will not have to be changed. Rather, CAT has to be enhanced with some moral-backed decision mechanisms presented in (Chentouf, 2000) such as altering the intention "cheating in exam" by: 1) replacing it with another which has less negative emotion, e.g. "cheating in exam one time only"; 2) adding another intention that is emotionally positive, like "cheating in exam and studying later"; 3) relativating the intention: "cheating is bad for others who have time to work; It is good for me because I am overwhelmed"; or 4) applying the Two-Wrongs fallacy (Gass, 2012): "*It is not wrong to cheat in exam because I have a bad teacher*".

Another possible enhancement shall be to base attraction strengths and emotional sensibility in Zed on real data. A study may have to be conducted for this aim.

Conclusion

This article presented the Cognitive Attraction Theory (CAT) and explained how its formulation of cognitive processes explains the processes of moral judgment and accurately define the role of intuition, emotions, reasoning, moral rules, and consequentialism. CAT does not deal with the sources, nature, and levels of appraisal processes. Further, it does not address when and how appraisal processes are activated. However, CAT does not contradict appraisal theories. Interestingly, it is compatible with Scherer's Stimulus Evaluation Checks (SEC) model (Sherer, 2001). The latter postulates that when an event is received it is checked at the sensory-motor level, then the schematic level, then the conceptual level. Appraisal processes at the sensory-motor level are mostly genetically determined. They are related to biological needs and engage biological mechanisms. The schematic level processing is automatic and outside of consciousness. It involves schemata learned from the society. The conceptual level is a slow process. It involves "highly cortical, propositional symbolic mechanisms, requiring consciousness and involving cultural meaning systems" (Sherer, 2001). We conjecture that CAT's moral judgment processes hold at two of SEC's levels. Moral intuition takes place at the schematic level, which is out of consciousness. Moral conflict, attraction, and conclusion engage the conceptual level as they imply conscious reasoning. According to Sherer, emotions at this level are more elaborated than at the schematic level. This is exactly what CAT postulates: emotion that is first assigned to a newly created belief may be updated at the end of the attraction sub-process. As future work, we intend to check this conjecture. The integration of CAT and the SEC model would contribute to advance research in both of the fields of moral judgment and appraisal theory. Such an integration also facilitates elaborating a computational model based on CAT and the SECs model. Yet, the experimental results obtained through Zed, the artificial moral agent, are encouraging despite the underlying computational model was only a part of CAT. Mainly, fallacies have not been implemented although they have a key role in moral judgment intuition and reasoning.

REFERENCES

Arsenio, W. F., & Gold, G. (2006). The effects of social injustice and inequality on children's moral judgments and behavior: Towards a theoretical model. *Cognitive Development, 21,* 388-400. doi:1.1016/j.cogdev.2006.06.005

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108,* 381-417. doi:1.1016/j.cognition.2008.03.001

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121,* 154-161. doi:1.1016/j.cognition.2011.05.010

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70,* 1-16.

Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57,* 1-29.

Chentouf, Z. (2000). *Homo informaticus*. Paris: L'Harmattan.

Chentouf, Z. (1997). *Invariable variable: Introduction to the recurrence theory*. Algiers: Office des Publications Universitaires.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science, 17,* 1082-1089. doi:1.1111/j.1467-928.2006.01834.x

Darwall, S. (2003a). *Consequentialism*. Oxford: Oxford University Press.

Darwall, S. (2003b). *Deontology*. Oxford: Oxford University Press.

Davis, N. (1993). Contemporary deontology. In P. Singer (Ed.), *A companion to ethics* (pp. 205-218). Malden, MA: Blackwell Publishing.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review, 5,* 5-15.

Gass, R. (2012). Common fallacies in reasoning. URL (last checked 17 October 2012). http://commfaculty.fullerton.edu/rgass/fallacy3211.htm

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107,* 1144-1154. doi:1.1016/j.cognition.2007.11.004

Gross, J. (1998). The emerging field of emotion regulation. *Review of General Psychology, 2,* 271-299.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316,* 998-1002. doi:1.1126/science.1137651

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834. doi:1.1037//0033-295X.108.4.814

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition, 100,* 530-542. doi:1.1016/j.cognition.2005.07.005

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84,* 221-236. doi:1.1016/S0010-0277(02)00048-3

Oatley, K. (1999). Emotions. In R. A. Wilson and F. Kiel (Eds.), *The MIT encyclopaedia of the cognitive sciences* (pp. 273-275). Boston: MIT Press.

Paulsen, O., & Sejnowski, T. J. (2000). Natural patterns of activity and long-term synaptic plasticity. *Current Opinion in Neurobiology, 10,* 172-179. doi:1.1016/S0959-4388(00)00076-3.

Pettit, P. (1993). Consequentialism. In P. Singer (Ed.), *A companion to ethics* (pp. 273-275). Malden, MA: Blackwell publishing.

Roseman, I. J., & Smith, C. A. (2001). Appraisal theories. In K. R. Sherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, method, research* (pp. 3-19). Oxford: Oxford University Press.

Sherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Sherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: theory, method, research* (pp. 92-120). Oxford: Oxford University Press.

Shwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgment of well-being. *Journal of Personality and Social Psychology, 45,* 513-523.

Stich, S., & Ravenscroft, I. (1994). What is folk psychology? *Cognition, 50,* 447-468. doi:1.1016/S0959-4388(00)00076-3.

Warren, H. C. (1921). *A history of the association psychology.* New York: Charles Scribner's Sons.

Wellman, H. M. (1990). *The Child's theory of mind.* Cambridge, MA: MIT Press.