



PRINZ'S CONSTITUTION MODEL OF MORAL
JUDGMENT: A CRITICAL READING

José Oliverio Tovar Bohórquez

Universidad del Valle

Abstract

The purpose of this paper is to critically review Prinz's constitution model. According to commonly suggested models, moral judgment is the result of specific cognitive processes that are intuitive, emotional or rational. According to Prinz, sentimentalist views argue that such judgments are caused by emotions. In contrast, he argues that moral judgment is constituted, not caused, by them. I will expose Prinz's argument to support his proposal and outline some inconsistencies of it.

Keywords: *Sentiment; Emotion; Moral Judgment; Constitution Model; Cognitive Processes.*

Recibido: febrero 18 de 2016 - Aprobado: mayo 02 de 2016

Praxis Filosófica Nueva serie, No. 44, enero-junio 2017: 107 - 118

ISSN (I): 0120-4688 / ISSN (D): 2389-9387

Modelo constitutivo del juicio moral de Jesse Prinz:

Una lectura crítica

Resumen

El propósito del presente ensayo es presentar una lectura crítica del modelo de constitución concebido por Jesse Prinz. De acuerdo con los modelos más significativos que se han producido hasta el momento, el juicio moral es el resultado de procesos cognitivos específicos, ya sean estos de tipo intuitivo, emocional o racional. Según afirma Prinz, la perspectiva sentimentalista sostiene que tales juicios son causados por emociones. En contraste, él defiende una tesis según la cual el juicio moral está constituido, no es causado, por aquéllas. En este artículo se presentará el argumento con el que Prinz sustenta su propuesta y se expondrán algunas inconsistencias.

Palabras clave: *sentimiento; emoción; juicio moral; modelo de constitución; procesos cognitivos.*

José Oliverio Tovar Bohórquez. Profesor asistente de la Universidad del Valle. Doctor en Filosofía y Filósofo por la Universidad Nacional. Sus principales áreas de trabajo y de investigación son: la filosofía de la educación, la filosofía moral, la neuropsicología moral y la psicología moral.

Dirección Postal: Universidad del Valle. Ciudad Universitaria Meléndez. Calle 13 No 100-00. Edif. 386

Dirección electrónica: jose.o.tovar@correounivalle.edu.co

PRINZ'S CONSTITUTION MODEL OF MORAL JUDGMENT: A CRITICAL READING¹

José Oliverio Tovar Bohórquez
Universidad del Valle

Introduction

Which mental or brain process is involved in the production of a moral judgment? Researchers working on different areas have recently offered answers to this question. For instance, Haidt (2001) proposed the *social-intuitionist model*, according to which moral judgment is caused by emotions, and reasoning follows its expression; Greene et al. (2004) put forth a *dual-process theory* where they argue that moral judgment is caused by both emotional and cognitive processes; finally, Nichols (2004) has suggested that moral judgment is expressed by means of sentimental rules, which are constituted by an affective mechanism and a normative theory. Prinz, for his part, and in agreement with sentimentalist theories, proposes the *constitution model*, and claims that emotions do not cause but constitute themselves the moral judgment.

In contrast with these views, Hauser (2006) presents a *rawlsian model*, according to which a moral faculty causes the moral judgment, and both emotions and reasoning appear afterwards. Each of these proposals, as well as the research conducted by Damasio (2007), Moll et al. (2002), Prehn et al. (2008), Heekeren et al. (2005, 2003), Ciaramelli et al. (2007), Koenigs et al. (2007), Cushman and Young (2011), Ditto and Liu (2012), to name a few,

¹Preparation of this research paper was supported by an Outstanding Graduate Students Scholarship, National University of Colombia.

offer a significant advance in the research of the psychological architecture of moral judgment.

A critical analysis of Blair (1995), Nichols (2004), Haidt (2001), Greene (2004), and Hauser (2006) can be found in the first two chapters of Tovar and Ostrosky (2013). In this paper, in turn, I carry out an analysis of Prinz's proposal. Prinz's work (Prinz, 2007, 2008) takes place within the debate regarding the mental or brain processes involved in the expression of a moral judgment. He gathers a considerable part of the discussion and defines important concepts, such as 'emotion' and 'moral judgment', which are neither defined by Blair nor by Hauser in their works. Prinz aims to build a model that responds to various problems encountered in the previous ones. The purpose of this paper is to establish a dialogue with this author in order to disclose some of the deficiencies of his proposal.

The constitution model

In what follows, I present some important definitions before discussing the model with which Prinz explains the psychological processes involved in the expression of a moral judgment.

According to Prinz's embodied appraisal theory (2004a, see also 2004b), *emotions* are perceptions of programmed bodily changes that carry information about our relationship with an environment. Such perceptions represent danger (loss, offense, etc.) as they have evolutionarily developed to trigger in the face of it (loss, offense, etc.). On the other hand, *Sentiments* are dispositions to experience emotions: this implies that emotions should not be considered sentiments, since *sentiments* are –to say it again– merely *dispositions* to experience emotions.

Prinz calls sentiments *moral rules*, and emotions *moral judgments* (Prinz, 2007, 96). To put it in another way, a moral rule is a sentiment and a moral judgment is an emotion. On the basis of his theory, Prinz establishes the *constitution model*, according to which emotions constitute moral judgments (Prinz 2007, 23-99; 2008, 162). This model contrasts with the one he calls *causal model*, which affirms that moral judgments cause emotions, and are therefore independent from them (Prinz 2007, 23).

According to Prinz, moral concepts allow us to be *disposed to* other-blame emotions when we are victims of their transgressions, and self-blame emotions when we are the transgressors (i.e. emotions of self-blame refers to the set of emotions we experience while committing a transgression, Prinz 2007, 90 and 95). If you believe that doing X is wrong, you have a long-term representation in your memory that disposes you to feel guilt or shame for doing X; alternatively, you feel anger, contempt or disgust if another does X.

Prinz presents the model through which he aims to explain the stages of information-processing that lead to a moral judgment in the following way: first, a perceived event is categorized; we classify it according to knowledge we have obtained during previous experiences. Second, long-term memory recovers a moral rule, which in turn activates a sentiment; third, this sentiment produces a certain emotion depending on the subject's context (e.g. if you perceive someone stealing a wallet, you feel anger toward her); fourth, such emotion – together with the representation of the perceived action – produces the moral judgment (see Prinz 2007, 96).

Prinz argues that one of the main differences between his model and those of Haidt (2001) and Nichols (2004), which also belong to the Humean tradition, is that emotions *constitute* the moral judgment, whereas the model of Haidt and Nichols considers that emotions cause it. Establishing a causal relation, as Haidt and Nichols do, implies that the moral judgment could be produced by something other than emotion, which contrasts with Prinz's rigid proposal, as in his view emotions co-occur with moral judgments (Prinz 2007, 99).

Prinz's model may be captured by the following sequence: on the first level, we perceive an event and categorize it, that is, we classify it according to knowledge we have gathered through previous experiences. On the second level, this event activates a relevant emotion if in our long-term memory we have a moral sentiment regarding this type of event (e.g. guilt, if it is me who performs the action; anger, if it is someone else). On a third level, a mental state results, which consists of a representation obtained by perceiving the action with an emotion toward the action; this complex (the representation of the action and the emotion) constitutes the judgment of the action as right or wrong.

Accordingly, the moral judgment does not appear at a later stage to the appearance of the emotion, but is constituted by this and by the representation of the action. Afterwards, the subject might deliberate, express the judgment in words or reevaluate the case and adjust her/his sentiments, etc. That is how Prinz's model explains the process that takes place during the expression of a moral judgment. In what follows, I describe some of the deficiencies of the constitution model.

Remarks on Prinz's Model

i. There appears to be an inconsistency in the second stage of Prinz's model. He states that he calls the sentiment a 'moral rule'; however, while describing the second stage of his model he argues that the moral rule causes

the sentiment.² In other words, Prinz first establishes an identity relation between moral rules and sentiments and then defends that the moral rules cause the sentiments. An argument might perhaps be construed to show that causality relations are for Prinz actually identity relations. However, while comparing his model with those of Haidt or Nichols, Prinz clearly distinguishes these two types of relations from one another: Haidt and Nichols defend a causality relation between emotion and moral judgment; in contrast, Prinz defends an identity relation and criticizes their posture. But, if causality relations were shown to be identity relations there would be no difference among Haidt's, Nichols's and Prinz's models. In this case, Prinz's posture would no longer be interesting since it would not state anything different to what Haidt and Nichols did. Why does Prinz, in presenting his model, claim that moral rules cause the sentiments? Why does he not maintain an identity relation between moral rules and sentiments, but does maintain it between emotions and moral judgments? Does breaking up the first relation (between moral rules and sentiments) imply breaking up the second one (between moral judgments and emotions) as well? Prinz does not answer these questions, which could cause his model to be inconsistent.

i.i. According to Prinz, some of the critics of emotionist theories³ argue that no relationship can be established between emotions and moral values since the former fluctuate, while the latter are stable. However, he claims that this argument does not affect his theory since his model is dispositional. Given that it is dispositional, it does not demand from each of our moral judgments that it contains an emotional manifestation (Prinz 2007, 97).

As I show in what follows, Prinz's answer to these critics compromises his constitution model. This model's main characteristic is that moral judgments are constituted by emotions, which is what distinguishes it from other models. In contrast to what I argued in the last paragraph, this means that each of our moral judgments contains an emotional manifestation. Following Prinz's theory, emotions are completely different from sentiments. An emotion is a type of perception (an embodied appraisal), whereas a sentiment is a concept (to the extent that it represents secondary qualities),

² “[...] to introduce a useful piece of terminology, we can call the sentiment a *moral rule*, and we can call a particular emotional manifestation of that sentiment a *moral judgment*.” “[...] The rule causes the sentiment to become active.” (Prinz 2007, 96, emphasis in the original). And further on: “A rule is retrieved from memory, which activates a sentiment.” (97)

³ ‘Emotionism’ is the term used by Prinz to refer to any theory that considers emotions as essential in some way. This term should not be confused with ‘emotivism’, which is a specific version of emotionism (Prinz 2007, 13).

as well as a disposition to have an emotion.⁴ Likewise, the brain activations related to each are completely different: working memory activates emotions, whereas long-term memory activates sentiments. Moral judgment is constituted by emotions, not by sentiments. Prinz's response to the critique mentioned (to argue that his model is dispositional) implies that moral judgments might be constituted by sentiments.

In sum, if emotions are different from sentiments, and moral judgments are constituted by emotions, moral judgments cannot be constituted by sentiments; they cannot lack of emotions. If moral judgments are constituted by emotions, and sentiments are different from emotions, moral judgments are, then, different from sentiments. Therefore, Prinz's model cannot be dispositional as long as it requires that each of our moral judgments contain an emotional manifestation.

As noted, the only response that Prinz provides to the critique is that his model is dispositional. This argument implies that moral judgments might be produced by sentiments without the intervention of emotions. Given that sentiments are different from emotions, Prinz happens to be in a certain dilemma: either he defends a constitution or a dispositional model. If he decides to defend the former, he is not able to respond to the critique against emotionists; however, if he decides to defend the latter, his constitution model is refuted.

i.i.i. A third remark on Prinz's model is related to the cognitive basis of moral judgment. According to Prinz, Haidt (2001) holds that moral attitudes typically arise in the absence of deliberative reasoning; this means that no deliberation is needed to express a moral judgment. Moral reasoning comes for Haidt always after the expression of a moral judgment; for this reason he argues that moral judgment is caused solely by emotions. Tovar and Ostrosky (2013, 24-28) show an inconsistency in Haidt's theory regarding this matter. In contrast, Prinz accepts that there are cases in which we carry out a deliberative reasoning before the action perceived that triggers inside of us an emotional response, and consequently a moral judgment as well (first level of the model) (Prinz 2007, 98).

According to Prinz, deliberation is what allows us to recognize an action as moral. Establishing the difference between Haidt's model and his own,

⁴ I think of psychological dispositions as physically realized states of the mind. If realism about the mind is true, then there is a fact of the matter whether someone has a sentiment; there is a thing in the brain that is the realization of that disposition. [...] So rather than saying that moral properties exist in virtue of causing certain emotions under certain conditions, we can say that they exist due to the fact that some observers have sentiments that dispose them to have those emotions.

Prinz claims that Haidt's has a certain deficiency: it does not allow for the deliberation required for moral attitudes to arise:

Unlike Haidt, I do not want to insist that moral attitudes typically arise in the absence of deliberative reasoning; it may take a lot of inference before we see an action in a way that triggers an emotional response (Prinz 2007, 98).

Prinz's theory is emotionist, which means that he believes morality is constituted by emotions and not by deliberation; nonetheless, his model may suggest that deliberation (or inference, which seems to be equivalent) is the foundation of morality (at least of what Prinz calls *moral attitude*). We recognize an action as moral, Prinz would argue, on account of a deliberative, and not of an emotional, process. It follows that emotions are posterior to this process, which appears in the first and second levels of the model (as presented above).

114 What has just been said implies that there are cases where deliberation is at the basis of moral judgment, that is, one could argue that the basis of moral judgment is not emotional but deliberative, although it is constituted by emotions. Prinz's purpose in *The Emotional Construction of Morals* (2007) is to show that emotions are at the basis of morality. Prinz's proposal of an explanation of the psychological architecture of moral judgment seems to be inconsistent with the general proposal of his moral theory. This inconsistency is evident in his remarks to Haidt's model. Soon after presenting his proposal on the psychological architecture of moral judgment, Prinz argues the following:

The model depicted [i.e., the constitution model] here has several nice features, which bear mention. First, it helps to diagnose cases in which moral judgments can be said to be erroneous. Consider Wheatley and Haidt's (2005) study described in chapter 1. They found that some people who were hypnotized to feel disgust ended up morally condemning a perfectly innocent individual. I think such condemnations qualify as errors because they were not caused by sentiments in long-term memory, but rather by extraneous facts; they do not qualify as legitimate expressions of the subjects' moral attitudes. In short, a wrong action is an action against which an observer has a moral rule. If an action is condemned because of hypnotically induced disgust, it does not qualify as wrong (Prinz 2007, 96).

Afterwards, in an article published in 2008 entitled *Is Morality Innate?* Prinz presents a very different interpretation from the one above:

I cannot adequately support the claim that moral norms are sentimental norms here, but I offer three brief lines of evidence.

First, psychologists have shown that moral judgments can be altered by eliciting emotions. For example, Wheatley and Haidt (2005) hypnotized subjects to feel a pang of disgust whenever they heard an arbitrary neutral word, such as ‘often’. They gave these subjects stories describing various individuals and asked them to make moral assessments. Compared to a control group, the hypnotized subjects gave significantly more negative moral appraisals when the key word was in the story, and they even morally condemned individuals whom control subjects described in positive terms (Prinz 2008, 369).

In this quotation from the book in which he defends the emotionist model, Prinz (2007) states that in Haidt’s and Wheatley’s experiment (2005) subjects are *not* making moral judgments, by which he aims to refute Haidt’s model. Again, “if an action is condemned because of hypnotically induced disgust, it does not qualify as wrong” (Prinz 2007, 96). Thus, this kind of judgment is not a moral one. Later, in the quote just presented, where Prinz criticizes the innateness hypothesis, he uses the same study as evidence in favor of his emotionist proposal, arguing that the individuals participating in the experiment *do* make moral judgments.

In sum, Wheatley and Haidt’s (2005) study supports Prinz’s emotionist theory, regardless of being valid or not. On the one hand, Prinz uses Wheatley and Haidt’s experiment to support his own theory (Prinz 2008, 369), but, on the other, he criticizes this experiment to show a quality of his emotionist theory (Prinz 2007, 96). Prinz’s inconsistent interpretation of both Wheatley’s and Haidt’s experiment (2005) questions the rigidity of his emotionist model.

i.v. A last remark on Prinz’s proposal is related with empirical evidence. Prinz (2007 22s) argues that the experiments conducted in neuropsychology serve as evidence in favor of emotionist postures, since all of them show that the zones that activate when the subject is in an emotional state coincide with those that activate while the subject makes a moral judgment.

Regarding that I will not present an argument to criticize the Prinz’s perspective. I only want to point out that there are studies that seem to prove something contrary to his view. In effect, recent studies in neuropsychology have found that neural zones that activate while a subject is in an emotional state are also active when the subject carries out high-level cognitive processes (e.g. deliberation). See Pessoa and Adolph (2010); Salzman and Fusi (2010); Prehn and Heekeren (2009); Pessoa (2008); Davidson (2000,

2003); and Dolan (2002).⁵ As these researchers argue, the distinction between emotional and rational processes does not correspond with the brain's architecture. It means that brain zones we thought activate only in emotional processes are also active in deliberative processes — such as the amygdala in the domain of emotion and the lateral prefrontal cortex in the case of cognition (Pessoa, 2008)—. What I believe follows from this finding, as such both emotionist and rationalist models ought to play a role in the production of moral judgment. Accordingly, it is innocuous to defend a strong interpretation of emotionism as Prinz does (2007, 9 and ch. 2).

In sum, the four remarks presented above allow me to claim that the constitution model, according to which moral judgments are constituted by emotions, is inconsistent. This also seems to affect Prinz's emotionist theory, at least in a tangential way. Certainly, as Prinz allows, there are cases in which we make moral judgments without being in an emotional state (second remark), and cases in which such a state is the product of a deliberation (third remark). There also seems to be an inconsistency regarding the way in which he presents the relationship between moral norms and sentiments (first remark).

⁵ Volume 52 (1-133) of *Brain and Cognition* is a special edition dedicated to affective neuroscience. Various articles contained in it show that emotions imply an activation of the cortical and subcortical zones. See specially Schulkin et al., Ericsson and Schulkin, and Adolphs et al.

References:

- ADOLPHS, RALPH, TRANEL, DANIEL, and DAMASIO, ANTONIO. R. 2003. "Dissociable neural systems for recognizing emotions." *Brain and Cognition*, 52:1, 61–69.
- BLAIR, JAMES. 1995. "A cognitive developmental approach to morality: investigating the psychopath." *Cognition*, 57:1, 1–29. Reprinted in Thomas, Nadelhoffer, Eddie, Nahmias, and Shaun Nichols, editors. 2010. *Moral Psychology: Historical and Contemporary Readings* (pp. 48–63). Maiden, MA: Wiley-Blackwell.
- CIARAMELLI, ELISA, MUCCIOLI, MICHELA, LADAVAS, ELISABETTA and DI PELLEGRINO, GIUSEPPE. 2007. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2:2, 84–92.
- CUSHMAN, FIERY and YOUNG, LIANE. 2011. Patterns of moral judgment derive from nonmoral psychological representations. *Cognitive Science*, 35:6, 1052-1075.
- DAMASIO, ANTONIO. 1996. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society London Series Biological Science*, 351:1346, 1413-1420.
- _____. 1994. *Descartes' error: emotion, reason, and the human brain*. New York: Putnam.
- DAVIDSON, RICHARD. 2003. Seven sins in the study of emotion: Correctives from affective neuroscience. *Brain and Cognition*, 52:129–132.
- _____. 2000. Cognitive neuroscience needs affective neuroscience (and vice versa). *Brain and Cognition*, 42:89–92.
- DITTO, PETER and LIU, BRITTANY. 2012. Deontological Dissonance and the Consequentialist Crutch. In Mikulincer, M., & Shaver, P. (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (pp. 51-71). Washington, D.C.: American Psychological Association.
- DOIDGE, NORMAN. 2007. *The Brain That Changes Itself: Stories of Personal Triumph from the frontiers of brain science*. New York: Viking.
- DOLAN, RAY. 2002. "Emotion, cognition, and behavior". *Science*, 298:1191–1194.
- GREENE, JOSHUA, NYSTROM, LEIGH, ENGELL, ANDREW, DARLEY, JOHN, COHEN, JONATHAN. 2004. "The neural bases of cognitive conflict and control in moral judgment". *Neuron*, 44: 389-400.
- HAIDT, JONATHAN. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment". *Psychological Review*, 108:814–834.
- HAUSER, MARC. 2006. *Moral Minds: how nature designed our universal sense of right and wrong*. New York: Harper Collins Publishers.
- HEEKEREN, HAUKE, WARTENGURGER, ISABEL, PREHN, KRISTIN, SCHWINTOWSKI, HANS-PETER, ARNO VILLRINGER. 2005. "Influence of bodily harm on neural correlates of semantic and moral decision-making". *NeuroImage*, 24: 887–897.
- KOENIGS, MICHAEL, YOUNG, LIANE, ADOLPHS, RALPH, TRANEL, DANIEL, CUSHMAN, FIERY, HAUSER, MARC AND DAMASIO, ANTONIO. (2007). "Damage to the prefrontal cortex increases utilitarian moral judgements". *Nature*, 446: (7138), 908–911.

- MOLL, JORGE, DE OLIVEIRA-SOUZA, RICARDO, BRAMATI, IVANEI AND GRAFMAN, JORDAN. 2002. "Functional networks in emotional moral and nonmoral social judgments". *Neuroimage*, 16: 696–703.
- NICHOLS, SHAUN. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- PESSOA, LUIZ. 2008. "On the relationship between emotion and cognition". *Nature Reviews Neuroscience*, 9: 148-158.
- PESSOA, LUIZ AND ADOLPHS, RALPH. 2010. "Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance". *Nature Reviews Neuroscience*, 11: 773-783.
- PREHN, KRISTIN and HEEKEREN, HAUKE. 2009. Moral Judgment and the Brain: A Functional Approach to the Question of Emotion and Cognition in Moral Judgment Integrating Psychology, Neuroscience and Evolutionary Biology. In Jan Verplaetse, Jelle De Schrijver, Sven Vanneste, Johan Braeckman, editors. *The Moral Brain: Essays on the Evolutionary and Neuroscientific Aspects of Morality* (pp. 129 – 154). Dordrecht, The Netherlands: Springer.
- PREHN, KRISTIN, WARTENBURGER, ISABELL, MÉRIAU, KATJA, SCHEIBE, CHRISTINA, GOODENOUGH, OLIVER R, VILLRINGER, ARNO, VAN DER MEER, ELKE; HEEKEREN, HAUKE. 2008. "Individual differences in moral judgment competence influence neural correlates of socio-normative judgments". *Social Cognitive & Affective Neuroscience*, 3:1, 33-46.
- PRINZ, JESSE. 2008. Is morality innate? In W. Sinnott-Armstrong (Ed.), *The evolution of morality* (Vol. 1, pp. 367–406). Cambridge: MIT Press.
- PRINZ, JESSE. 2007. *The emotional construction of morals*. New York: Oxford University Press.
- . 2004a. *Gut Reactions: A perceptual theory of emotion*. New York: Oxford University Press.
- . 2004b. Which emotions are basic? In D. Evans and P. Cruse (Eds.), *Emotion, evolution, and rationality* (pp. 69–87). New York: Oxford University Press. Traducido al español por Tovar, José. 2010. "¿Cuáles son las emociones básicas?". *Cuadernos de Crítica*, México: UNAM, Instituto de Investigaciones Filosóficas, No. 55.
- SALZMAN, DANIEL, and FUSI, STEFANO. 2010. "Emotion, cognition, and mental state representation in amygdala and prefrontal cortex". *Annu Rev Neurosci*, 33:173-202.
- TOVAR, JOSÉ, and OSTROSKY, FEGGY. 2013. *Mentes Criminales ¿Eligen el Mal? Estudios de Cómo se Genera el Juicio Moral*. México: Manual Moderno.
- WHEATLEY, THALIA, and HAIDT, JONATHAN. 2005. "Hypnotic disgust makes moral judgments more severe". *American Psychological Society*, 16:10, 780-784.