**Title:** Agency and Morality

**Names:** Paulo Sousa and Christopher Manoharan

**Affiliation:** Institute of Cognition and Culture, Queen’s University, Belfast

**Email Addresses:** paulo.sousa@qub.ac.uk, cmanoharan22@qub.ac.uk

**Word Count:** 2,661

**Abstract:**

Cognitive traditions study how human beings utilize naïve concepts of agency and morality to make judgments about persons and their actions, and how these concepts and judgments are related. We discuss, in particular, how judgments of whether an action is intentional or not affect judgments of moral blame and moral wrongdoing. Cognitive approaches argue that judgments of intentionality constitute a fundamental factor in how humans calculate a person’s degree of blame for the occurrence of a bad event, even if some anthropologists have claimed that in many cultural contexts people leave aside questions of intentionality, adhering to a criterion of strict or absolute liability. However, cognitive approaches argue that judgments of intentionality constitute a much less important factor in how humans categorize an action as an instance of wrongdoing (i.e., as a normative transgression), since ordinary norms prohibiting an action often do not specify anything about intentionality.

**Main Text:**

The relationship between agency and morality can be approached in two basic ways. One may consider the extent to which human beings have agency and are moral beings, and how these properties are related. Or, one may consider how human beings utilize naïve concepts of agency and morality to make judgments about persons and their actions, and how these concepts and judgments are connected. While many philosophical, sociological and anthropological traditions have focused on the first approach, cognitive traditions have focused on the second.

The cognitive relationship between agency and morality has been discussed in two directions: the contribution of agency judgments to moral judgments; the contribution of moral judgments to agency judgments. While there is a long tradition of research focusing on the first direction, only during the past ten years has the second direction of the relationship been discussed in the cognitive sciences. Since this more recent research suggests simply that different concepts of agency are activated in moral contexts in contraposition to amoral contexts (see Sousa, Holbrook & Swiney, 2015), the focus here is placed on the first direction, the contribution of agency judgments to moral judgments. Before discussing research on this contribution, we characterize the relevant concepts and related judgments.

**1. Relevant concepts and judgments**

**1.1. Agency**

The naïve concept of agency considered here concerns the contrast between *intentional* and *unintentional* actions (Throughout our discussion, we use “action” in a broad sense that includes both actions and omissions of actions). A clear case of intentional action is an action that is intended and anticipated. For example, Peter may open the door abruptly in order to startle Roger, anticipating that he would succeed in doing so. If Peter succeeds, we would say that he startled Roger intentionally. An unintentional action is an action that is not anticipated, and hence not intended. For example, Peter may open the door abruptly without anticipating that he would startle Paul, hence without intending to startle Paul. In this case, if Peter startles Paul, we would say that he did so unintentionally.

**1.2. Morality**

Naïve morality concerns certain ordinary concepts utilized to evaluate persons and their actions. Two such evaluations are particularly prominent: the judgment that an action is morally wrong, and the judgment that a person is morally to blame for the occurrence of a bad event. In the literature, the second judgment here is often phrased in terms of “moral responsibility”, but, because the words “responsible” and “responsibility” are extremely polysemous and have caused much confusion both in anthropology and the cognitive sciences, we avoid this phrasing (see Sousa, 2009).

**1.2.1. Moral wrongdoing**

The judgment that an action is wrong supposes some normative expectation. Norms (obligations or prohibitions) contrast with what is discretionary. Doing what is obligatory (e.g., providing adequate care for one’s children) entails a sense of right, and doing what is prohibited (e.g., killing an innocent person) entails a sense of wrong. Doing what is discretionary (e.g., going to the cinema) entails a sense of neither right nor wrong. Because it is obligatory to omit what is prohibited (e.g., obligatory to not kill an innocent person) and it is prohibited to omit what is obligatory (e.g., prohibited to not provide adequate care for one’s children), both doing what is prohibited and omitting what is obligatory are seen as instances of wrongdoing. Thus, wrongdoings are actions that do not follow normative expectations—they are norm transgressions. But which actions are seen as *morally* wrong?

While most anthropologists use the word “morality” in the broad sense that includes all or almost all instances of wrongdoing, the cognitive tradition normally uses “morality” to refer to a specific type of wrongdoing. One important way of characterizing this specificity is to say that moral transgressions are seen as inherently wrong: their normative status is considered to be independent of authority or social consensus. Although there is an important dispute in the literature on whether only transgressions related to certain types of normative content lead to this sense of inherent wrongness (see Baumard & Sperber, 2012; Haidt, 2013; Sousa & Piazza, 2014), actions like killing an innocent person or committing incest, for example, are prototypical cases of moral transgressions in this sense.

**1.2.2. Moral blame**

The concept of *moral blame* involves two closely related aspects: the judgment that a person is *culpable* for the occurrence of a bad event and is therefore *liable* to some impositions, such as to suffer punishment and/or to make reparations. For example, someone may be judged culpable for the death of an innocent person, and therefore liable to receive punishment. These judgments evaluate persons in the sense that the person being judged is seen as acquiring a blemish in their record because of her contribution to the occurrence of the bad event, and as deserving punishment or having to make reparations. Both judgments presume a judgment of moral wrongdoing in that the person being judged is supposed to have done something morally wrong that contributed directly or indirectly to the occurrence of the bad event.

**2. The influence of agency judgments on moral judgments**

The bulk of the literature addressing the influence of agency judgments on moral judgments focuses on moral blame, rather than on moral wrongdoing, although these two judgments are not always clearly distinguished by many researchers. Accordingly, we shall reserve most of our discussion to the former influence.

**2.1. Agency and moral blame**

The concepts of culpability and liability have a graded structure. People may be deemed more/less culpable, liable to suffer more/less punishment, or to make more/fewer reparations for the occurrence of some bad event. The cognitive literature in developmental and social psychology has probed whether agency is one important variable that people take into account when determining degrees of culpability and liability.

The basic methodology utilized is a type of experimental methodology built into the context of scenario research (e.g., Zelazo, Helwig & Lau, 1996). The researcher constructs scenarios where “intentionality” is manipulated as the independent variable. For example, a researcher may create different scenarios about an agent harming an innocent person, with each of the scenarios varying only in terms of the level of intentionality of the causation of harm, from completely accidental to completely intentional. For each of these scenarios, participants are asked questions about the relevant dependent variables, namely, the culpability and liability of the agent. If participants increase the degree of culpability and liability of the agent according to the increase in the level of intentionality, this provides evidence that intentionality judgments influence culpability/liability judgments.

The results of this experimental research have shown consistently that children (from roughly the age of 5) and adults attribute more culpability and liability for intentional than unintentional harms. Moreover, people are sensitive to a distinction between situations of pure accidents, wherein one acts carefully and could not have foreseen the harmful consequences of their actions, and situations involving some sort of recklessness or negligence.Accordingly, people attribute more culpability and liability in cases of recklessness/negligence than in cases of pure accidents.

However, some anthropologists claim that in many cultural contexts there are prevalent norms proscribing people to speculate about others’ mental states (often called the folk doctrine of “the opacity of other minds”), which leads people to downplay the concept of agency, either in itself or as a variable relevant to judgments of moral blame (cf. Goldman, 1993; Rumsey & Robbins, 2008). In particular, some anthropologists claim that, because people in these contexts downplay mind-reading, they adhere to a concept of “strict or absolute liability”—they tend to calculate the amount of culpability and/or liability only in terms of the result (i.e., degree of damage caused by one’s action), and without taking into account the mental states that led to the result. For example, in the case of someone killing an innocent person, people in these contexts would calculate the amount of culpability and/or liability simply in terms of the fact of the death, independent of the type of mental state involved in the causation of death.

There are complications with the hypothesis that people in many cultural contexts adhere to strict or absolute liability. The hypothesis is not completely clear: is it that people disregard the general opposition between intentional and unintentional actions, in which case they would consider that intentional killing and unintentional killing involve the same amount of culpability and/or liability? Or is it that people disregard only the opposition between recklessness/negligence and pure accidents, in which case they would consider that death resulting from recklessness/negligence and death resulting from pure accident involve the same amount of culpability and/or liability? Additionally, the evidence provided to support the hypothesis is often underspecified; in particular, it does not include details on the different perspectives that normally exist concerning moral blame for the occurrence of bad events (see Goldman, 1993).

Consider Harry Walker’s illustration of the opacity doctrine and strict liability thesis, in a passage of his article on the incorporation of the idea of “Law” by the Urarina of Amazonian Peru (Walker, 2015, p. 53). Walker describes a mediated dispute regarding the case of the death of a young boy named Toribio, who died as a consequence of his friend having accidentally struck his hand with a machete. Toribio’s father demanded reparation in terms of “hefty compensation” for the death, and, according to Walker, the dispute focused only on the fact of the death and the degree of compensation, suggesting a complete disregard of the nature of the mental states of the perpetrator.

However, Walker states that everyone in the dispute implicitly acknowledged that the strike was unintentional; and he also suggests that the case would have been quite different if the strike had been interpreted as intentional (“negotiations may not otherwise have even begun”). Plausibly, if the action were seen as intentional, the case might have lead to the pursuit of bloody retribution (even if by mystical means) and/or to the intuition that Toribio’s father deserves even greater compensation. Either reaction would show that the mental states of the perpetrator are not disregarded and that the contrast between intentional and unintentional killing is indeed important for the Urarina.

Moreover, Walker’s passage does not give an account of the arguments put forward by the two sides of the dispute concerning the appropriate degree of compensation—in particular, it does not describe whether the two sides of the dispute had different interpretations of the accident. Did the victim’s side invoke something akin to recklessness/negligence (e.g., that there was some lack of due care on the part of the friend, in terms of the strike and/or what happened afterwards)? Did the perpetrator’s side interpret the death more like a pure accident or invoke something akin to contributory negligence on the part of the victim? A positive answer to either of these questions would show that intuitions about recklessness/negligence are deployed in calculating the amount liability in cases of accidents (similar to what one finds in Western tort law).

Thus, as this example illustrates, the evidence in support of the version of the strict-liability hypothesis that claims that in many cultural contexts people disregard the general opposition between intentional and unintentional actions in making judgments of moral blame is quite disputable. Indeed, cognitive and evolutionary approaches view this distinction as fundamental to basic psychological mechanisms establishing human cooperation—e.g., mechanisms detecting cheating or other kinds of selfish behavior.

Now, this example also illustrates that the evidence in support of the version of the strict-liability hypothesis that claims that in many cultural contexts people disregard the opposition between recklessness/negligence and pure accidents in making judgments of moral blame concerning accidents is weak (see also Goldman, 1993, for many other ethnographic discussions). Here, from a cognitive and evolutionary point of view, it is trivial that there are many situations of pure accidents that people have intuitions of strict liability (in particular, intuitions that there should be reparation), and this in *any* cultural context. For example, in a western context, if one were to break a friend’s vase by a complete accident while visiting her, wouldn’t one have the sense that one should make some reparation? However, the hypothesis that in many contexts people do not take into account *the difference* *between reckless/negligence and pure accident* in calculating the amount of blame related to accidents is much more questionable, for the same aforementioned cognitive-evolutionary reasons (Note that the often-quoted idea that Western tort law was initially based solely on a criterion of strict liability and only later incorporated considerations of negligence is disputable).

**2.2. Agency and moral wrongdoing**

Most of the literature concerning the relationship between agency and morality does not deal with the influence of agency judgments on judgments of moral wrongdoing because, in an important but somewhat obvious way, judgments of wrongdoing in general are independent of agency judgments.

Ordinary norms prohibiting an action, whether the prohibition is moral or not, often do not to specify that it is the action as intentional action that is prohibited; they specify simply that the action is prohibited. Thus, if the prohibited action is performed, even if unintentionally, this is sufficient for one to judge that the action is wrong—the norm is still in force and the action is a transgression of the norm. For example, etiquette norms do not normally specify anything about intentionality, and if one does not follow the etiquette norm, even if unintentionally, one is seen as transgressing the norm. Turning to a moral example, take the incest taboo. It is plausible to suppose that in most cultural contexts the way the incest taboo is conceptualized does not specify anything about intentionality in the prohibition of incest. Therefore, the fact that a couple commits incest unintentionally does not eliminate the fact that their action transgressed the norm, hence that their action is seen as wrong. Actually, there are cultural contexts in which the irrelevance of intentionality is built explicitly into the understanding of the transgression by the incorporation of an explicit rationale indicating why incest is wrong independently of intentionality—incest has catastrophic consequences that affect everyone regardless of whether it is committed intentionally or unintentionally (see Astuti & Bloch, 2015).

More broadly, this point concerns the detachment of judgments of wrongdoing from judgments of blame that may occur in ordinary cognition (i.e., an action can be seen as wrong but the agent is not seen as culpable because the action was completely unintentional). And it also indicates that it may be misleading to compare ordinary norms with western legal norms (e.g., crimes like premeditated or depraved-heart murder, voluntary or involuntary manslaughter), since the legal system has to build culpability qua levels of intentionality explicitly into the definition of crimes in order to establish different ranges of liability.

**SEE ALSO:**

Norms Psychology, Values

**References and Suggested Readings**

Astuti, R., & Bloch, M. (2015). The causal cognition of wrong doing: incest, intentionality, and morality. *Frontiers in psychology*, *6*.

Baumard, N., & Sperber, D. (2012). Evolutionary and cognitive issues in the

anthropology of morality, in Fassin, D. (Ed.), A Companion to Moral Anthropology, Wiley-Blackwell

Goldman, L. (1993). *The Culture of Coincidence: Accident and Absolute Liability in Huli*.

New York: Clarendon Press.

Haidt, J. (2013). *The Righteous Mind*. Allen Lane.

Rumsey, A., & Robbins, J. (2008) Social Thought and Commentary Section: Anthropology and the Opacity of Other Minds, *Anthropological Quarterly*, *81* (2), 407-494.

Sousa, P. (2009). A cognitive approach to moral responsibility—the case of a failed attempt to kill. *Journal of Cognition and Culture*, *9*, 171-194.

Sousa, P., Holbrook, C., & Swiney, L. (2015). Moral asymmetries in judgments of agency withstand ludicrous causal deviance. *Frontiers in Psychology*, 6: 1380.

Sousa, P., & Piazza, J. (2014). Harmful transgressions qua moral transgressions: A

deflationary view. *Thinking & Reasoning, 20(1),*99-128*.*

Walker, H. (2015). Justice and the Dark Arts: law and Shamanism in Amazonia. *American Anthropologist*, 117, 47-58.doi:10.1111/aman.12170

Zelazo, P., Helwig, C., & Lau, A. (1996). Intention, act and outcome in behavioral

prediction and moral judgment. *Child development*, 67, 2478-2492