

## **The evolution of moral intuitions and their feeling of rightness**

**Christine Clavier & Chloë FitzGerald\***

### **Abstract**

Despite the widespread use of the notion of moral intuition, its psychological features remain a matter of debate and it is unclear why the capacity to experience moral intuitions evolved in humans. We first survey standard accounts of moral intuition, pointing out their interesting and problematic aspects. Drawing lessons from this analysis, we propose a novel account of moral intuitions which captures their phenomenological, mechanistic, and evolutionary features. Moral intuitions are composed of two elements: an evaluative mental state and a feeling of rightness (FOR). We illustrate the phenomenology of the FOR with examples of non-moral and moral cases, and provide a biological and mechanistic account: the emergence of human reasoning capacities created a need for the co-evolution of a psychological system producing the feeling of rightness (the FORs). This system is triggered when we experience conflicting evaluations. The FORs renders evaluations resulting from rational deliberation less compelling than the evaluations produced by simple evolved systems. It thus facilitates optimal decision-making, preventing excessive interference by rational deliberation. Our account sheds light on why moral intuitions are so frequently experienced and why they are so compelling and resistant to argument. In addition, the account fuels interesting speculations about common metaethical intuitions.

### **1. Introduction**

Humans have many kinds of responses that may be called “intuitions.” On a common understanding, an intuition is akin to a gut feeling, hunch or instinct, encountered frequently in both moral and non-moral contexts in everyday life. According to this picture, a moral intuition could be a response to a real life situation in which one is directly involved, such as deciding when to break it to one’s mother that she has a terminal illness. However, the kind of moral intuitions most investigated and examined by moral philosophers and psychologists consist in compelling responses to hypothetical scenarios, e.g., whether a doctor should sacrifice a healthy patient in order to save the lives of five other patients in an urgent organ donation (Foot 1967), or to abstract moral principles, e.g., the principle “It is wrong to lie.” Finally, philosophers are also interested in metaethical intuitions, which are concerned with the nature of morality. For instance, one commonly held intuition is that there are universally valid moral truths.

Moral intuitions are puzzling phenomena because they are very resistant even when we face good reasons for questioning them or discounting their force. We are typically convinced that the moral norms and judgments we intuitively endorse are right and should be universally held. However, anthropological and worldwide survey studies indicate that the moral norms intuitively endorsed by people can vary widely from one culture to another and even within a

---

\* Both authors contributed equally to the conception and writing of this article.

culture (Haidt & Graham 2007; Sachdeva, Singh, & Medin 2011). We routinely experience this diversity in moral opinions and observe the etiology of others' moral intuitions without this leading us to doubt our own deeply-felt moral intuitions. Let us illustrate.

Imagine that Aaron and Sabrina are engaged in a heated debate about immigration. Sabrina has a strong intuition that it is wrong to allow the country to be "swamped" by immigrants, while Aaron has a similarly strong intuition that it is wrong to ignore the suffering of fellow human beings. Each is aware that they cannot agree because they do not share each other's intuitions. Sabrina perceives that Aaron's liberal views are the result of his upbringing in a left-wing family that have inculcated him with "bleeding heart" values. She discounts his intuitions because she considers that their basis in his upbringing renders them arbitrary. However, Sabrina knows that her views stem from her own upbringing in a conservative household where patriotic values were prized. This knowledge does little to make her doubt her own moral intuitions because she simply feels fortunate to have grown up with the "right" values. This example illustrates how moral intuitions can remain compelling even when we reflect on information that leads us to question their objectivity. This aspect of moral intuition makes them attractive as an object of study.

The philosophical and psychological literature contains a variety of accounts of moral intuitions. However, a fully-fledged explanation of the phenomenological, evolutionary, and mechanistic aspects of moral intuitions is currently lacking. In particular, existing accounts provide only cursory descriptions of the compelling psychological experience of having a moral intuition, which distinguishes it from the experience involved in making a similar non-intuitive evaluation. Moreover, we still lack a general explanation of *why* humans have intuitions in given circumstances, but not in others; is there a general pattern or a mechanism that could explain or predict the occurrence of moral intuitions? We answer this question with a novel evolutionary account of the capacity to experience moral intuitions.

Our aim is to shed light on what moral intuitions are and why they are remarkably resistant in the face of evidence of their interpersonal variability. We provide a novel descriptive picture of moral intuitions within an evolutionary framework that avoids the major pitfalls encountered by other accounts of intuition. Section 2 is an overview of how moral intuitions are understood in the literature in psychology and philosophy. We explain why we find something lacking in these views, particularly when it comes to explaining the compelling phenomenology of moral intuitions. Section 3 is an account of moral intuitions as a two-component phenomenon, comprised of an evaluative mental state and a feeling of rightness. Section 4 provides an evolutionary and mechanistic account of the feeling of rightness and introduces a necessary eliciting condition for intuitions: the existence of conflicting evaluations in the subject's mind. Section 5 explains how this feeling of rightness applies to moral intuitions, including metaethical intuitions.

## **2. Intuitions as usually understood in the literature**

Moral intuitions are a popular topic, especially in philosophy and psychology. There is a wide variety of views and we can only provide a brief overview. Two common features of moral intuitions highlighted in the literature are that they are compelling from a first-person perspective and that they are not the result of conscious inference on the part of the intuiter.

Many regard just this latter feature as necessary and sufficient (e.g. Tersman 2007). When a more detailed characterization of intuitions is required, however, divergences arise. As we will see in this section, there are three broad categories of accounts: for some, intuitions are a particular form of judgment or belief; for others, intuitions are emotional or affective responses; and more recently, intuitions have been characterized as seemings, states that present the way things seem to us.

Normative ethicists tend to think of moral intuitions as a kind of judgment, belief, conviction, or proposition (Cohen 2008; DePaul 2006; Kamm 2006; McMahan 2013; Rawls 1971). Most ethicists do not give a particularly detailed descriptive account of intuition because they are more interested in the practice of normative ethics than in moral psychology. The moral judgments that are thought deserving of the label “moral intuition” are characterized in different ways, e.g., as strong (Sinnott-Armstrong 2008), self-evident (Audi 2009; Shafer-Landau 2005), or as the result of long and considered reflection (Kamm 2006). On these views, having the moral intuition that abortion is wrong involves having a strong, self-evident, or considered judgment that abortion is wrong.

One major difficulty with these accounts is that the way intuitions are distinguished from other judgments tends to be underdeveloped: characterizing intuitions as judgments that are “strong,” “self-evident,” or “considered” does not capture the phenomenology of intuitions. To illustrate, a Canadian, accustomed to the moral norm of leaving substantial tips in restaurants, is likely to feel intuitively that she must leave a tip while visiting Spain, even while aware that tipping is not an obligatory part of the local custom. In this case, she has the intuition that she should leave a tip, while simultaneously judging that there is no moral obligation to tip. In the reverse case, a Spaniard may have no moral intuition that she should leave a tip when in Canada, yet firmly judge that she should leave a tip because she knows that this is a local moral norm. Thus, a person may make a judgment while finding the opposite judgment intuitive. Furthermore, two different people may make the same judgment, one compelled by intuition, and the other deliberately going against her intuition. The phenomenological differences between the Canadian’s judgment that she should tip and her judgment that she should not cannot be convincingly equated with the strength or self-evidence of these judgments. Moreover, the example of the Canadian reveals how a person’s considered judgment can in fact be different from her intuition.

The phenomenology of moral intuitions has led some scholars to identify them, or a particular subclass of them, with emotions or affective responses (Greene 2007; Kauppinen 2013; Roeser 2011; Singer 2005). For instance, seeing a child being beaten by bullies is likely to make one feel indignant or compassionate, or imagining betraying one’s closest friend is likely to make one feel ashamed or disgusted. According to this kind of account, moral intuitions spring from these emotional experiences. For example, when the Canadian feels strongly that she should leave a tip, this springs from an emotional response, in contrast to her judgment that she is not required to tip.

The difficulty with this account, however, is that not all cases that are typically cited as moral intuitions seem to have an emotional component or to involve emotional responses. For instance, reflecting on the principle “It is wrong to steal for fun” may not arouse any emotion, even while it strikes one as intuitive. One can have a strong moral intuition about how to fairly distribute a cake among children at a party without feeling particularly emotional about

the matter. It is possible to avoid this problem by developing a more nuanced view of the link between emotions and intuitions, arguing that intuitions involve dispositions to feel emotions or judgments of the appropriateness of feeling certain emotions (Kauppinen 2013; Roeser 2011). Or one can argue in the opposite direction that moral intuitions that do not spring from emotional responses are superior to those that do (Greene 2007; Singer 2005).

Nevertheless, these views struggle to explain what makes the phenomenology of intuitions distinctive in those cases where emotions are not directly involved. A view of moral intuitions as *feelings* rather than emotions could also circumvent this difficulty, while retaining the advantage of explaining the phenomenology of intuitions via their affectivity (Haidt & Bjorklund 2008). Feelings are one possible element of fully-blown emotions, but they are also involved in non-emotional mental and physical states. However, “feeling” is a vague term that is used to describe a wide range of affective mental and bodily phenomena, e.g., “feeling of tiredness,” “feeling of uncertainty,” “feeling of hunger,” “feeling of dread,” “feeling of exclusion,” “feeling of entitlement,” “feeling of familiarity.” More detail is needed to individuate the kind of feeling that is involved in moral intuitions, a point we will return to in the next section.

A third kind of account classes moral intuitions as seemings, which are neither judgments nor emotions, but a particular type of state that represents how the world appears to us (Bedke 2008; Huemer 2008; Tucker 2013). A classic example of a seeming in the non-moral domain involves perceiving a stick as bent when it is half-submerged in water; we have the perceptual seeming that it is bent even while we judge that it is straight. If moral intuitions are seemings, it is easy to see how one can have a moral intuition that conflicts with one’s actual judgment. For instance, the Canadian in Spain has the seeming that she should tip, but she judges that she does not need to tip. Her seeming is not an emotional response; it simply presents the appearance to her that tipping is the morally appropriate action. In a way that is somewhat similar to the perceptual seeming of the stick, her intuition about tipping can be remarkably resistant despite her considered judgment to the contrary.

This view has the advantage of identifying a specific phenomenology that differentiates intuitions from judgments. Matthew Bedke argues that the seemings involved in moral intuitions are constituted by feelings such as felt appropriateness or veridicality (Bedke 2008: 261-3). Unfortunately, Bedke does not specify further the nature of the feeling that constitutes a moral seeming. This is a general problem in the literature on moral intuitions: the notion of “feeling” is found relevant for defining intuitions but remains underspecified. It includes no information about precisely what kind of feeling is involved, how it arises, and why it appears in certain situations but not in others. In this article, we provide such a contribution.

### **3. A definition of moral intuitions and a phenomenological account of the feeling of rightness**

We argue that moral intuitions are a two-component phenomenon involving a *challenged evaluation* and a *feeling of rightness* (FOR). Thus when the Canadian has the moral intuition that she should tip the waiter in Spain, her intuition is composed of both (a) an evaluation that it is right to tip the waiter, which is challenged by the evaluation that she is not morally required to tip, and (b) an FOR that reinforces her first evaluation. It is not necessary for our view to take a stance on the controversial matter of which kinds of mental states involve

evaluations. “Evaluation” here is a placeholder for a judgment, belief, emotion, decision, or any other judgment-like state that the reader considers involves evaluation. However, as will be explained in the next section, for an evaluation to feel intuitive, it must be challenged by a contradictory evaluation. For now, let us concentrate on the second and most original individuation criterion for moral intuitions: the FOR. This section provides a phenomenological description of this feeling.

The FOR belongs to a particular class of feelings that is not usually mentioned in the literature on moral intuition. Accounts that liken moral intuition to emotion or feeling are correct to recognize an affective component to the experience of finding something intuitive. This affective component is what produces the particular phenomenology and makes a moral intuition *feel* different from a judgment. The Canadian who has the intuition to leave a tip in Spain, is likely to say: “I know that I don’t have to leave a tip here, but I *feel* that I should nonetheless.” However, we argue that such feeling episode is not necessarily linked to bodily or mental states involved in emotions. It pertains to a class of feelings that psychologists call “metacognitive feelings,” which are feelings *about our own thought processes and cognitive capacities*.

There are various forms of metacognitive feeling. In the non-moral domain, one much-discussed feeling is the tip-of-the-tongue experience, a vivid experience of being very close to accessing something from memory, such as someone’s name, but not being quite able to do so (Schwartz & Metcalfe 2011). A different form of metacognitive feeling which has been identified by Valerie Thompson (2009) is the “feeling of rightness” or FOR, which is about the correctness of a belief or evaluation that arises quickly and fluently in one’s mind. It is a vivid experience of having the right answer which feels so compelling that it tends to shut down or bias further reflection, even if the first answer is not necessarily correct. As an example, Thompson provides a simple mathematical problem from an IQ test:

*If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?* (Thompson 2009: 172)

Most of us have the intuition that the answer is 100 minutes and do not bother resorting to effortful calculation because the answer is accompanied by an FOR. However, the right answer is 5 minutes.

Thompson does not discuss the FOR in moral cases, but we think that this psychological phenomenon is present whenever moral intuitions are felt. When we experience a moral intuition, the FOR is involved in a similar way as in the example above. It makes us feel that we have the right answer and renders us less likely to challenge it. For instance, if the Canadian in our tipping example forces herself to reason that Spanish moral norms do not require her to tip, the FOR renders this reasoning unconvincing and easy to ignore or to question. The Canadian may end up not leaving a tip, but not without a mental struggle to repress her intuition. In addition, she is likely to feel guilty about not leaving a tip, even while she judges that she behaved correctly according to local norms.

Our descriptive account of moral intuition encompasses elements from all three kinds of account discussed in the preceding section. Firstly, it recognizes that moral intuitions involve moral evaluations, often in the form of judgments, an aspect highlighted by the judgment

accounts. Secondly, it includes the FOR to account for the particular phenomenology of intuitions. With the FOR, we acknowledge the affective component of moral intuitions emphasized by emotion accounts, without assimilating moral intuitions to emotional reactions. Indeed, the FOR is distinct from emotional feelings. In highly emotional situations, e.g., when we condemn the Nazis' crimes against humanity, it may be difficult to discern the presence of the FOR because it is phenomenologically overshadowed by emotional feelings. The FOR is easiest to identify when it occurs in non-emotional contexts, e.g., when we evaluate the relevance of an abstract moral principle. Thirdly, our account is sympathetic to the idea that moral intuitions are seemings, distinct from emotions or judgments. We go one step further by elaborating on the composition of seemings as a two-component phenomenon: a challenged evaluation and the FOR. In the next section, we will add an evolutionary and mechanistic explanation to this description. This will allow us to characterize the FOR more closely and to explain why all intuitive evaluations are *challenged*.

#### **4. An evolutionary understanding of the feeling of rightness**

Human cognition is a complex phenomenon composed of numerous task-specific systems and sub-systems organized in a web of interactions that are largely unknown. Some simple systems have been studied and their biological function singled out. For instance, the food aversion system serves to avoid the consumption of toxic food (Bernstein 1999), or the mimicry of other's movements system serves to facilitate joint action with other individuals (Knoblich, Butterfill, & Sebanz 2011). Mechanistically, these evolved systems process a given type of stimuli, e.g., unpleasant experiences with a food item, and are made of simple and distinct task-specific operations, e.g., felt aversion towards foods that have been associated with an unpleasant experience. These task-specific operations are "information-frugal" in the sense that they are activated by very specific cues. They are also "processing frugal" in the sense that they use simple algorithms or rules tailored specifically to the task demands, and during the execution of these operations a small subset of the total available information in the mind is recruited (Carruthers 2007). Here we argue that the FOR is the expression of a simple task-specific system of this sort, whose biological function and mechanistic functioning can be singled out.

The FOR plays a role at the interface between Type 1 and Type 2 thinking. According to dual process theories (Evans & Stanovich 2013), human cognitive abilities can be divided into two kinds of thinking that can be roughly characterized as follows: Type 1 thinking involves fast, unconscious, autonomous processing, and is typically induced by the activity of relatively independent and primitive systems dedicated to particular decision problems. It does not make demands on working memory and is thus relatively fluent and effortless. Type 2 thinking, however, involves slow and conscious processing that gives rise to reasoning able to manipulate abstract concepts and rules. This type of thinking needs working memory resources, making it effortful and expensive. As Thompson (2009) originally argued, part of the function of the FOR is to constrain Type 2 thinking. The FOR favors Type 1 responses, which come easily and fluently to people's mind, and thus discourages or biases Type 2 thinking. Here we build on this idea and embed it in an evolutionary and mechanistic explanation.

On our view, the evolution of the capacity for Type 2 thinking created a problem for early humans. Abstract thinking and reasoning allowed humans to communicate with each other, to cooperate in large groups and to develop sophisticated tools and techniques (Barrett, Cosmides, & Tooby 2010). However, there is another side to the coin. As Gert Gigerenzer and others have highlighted, it is sometimes more adaptive to make a decision based on one's "gut feelings," or Type 1 processes, rather than based on careful deliberation (Gigerenzer, Todd, & ABC Research Group 1999). Humans often need to make quick decisions about complex problems based on incomplete knowledge. In these circumstances, they may sometimes be better off abandoning rational weighting of all relevant cues, relying instead on evolved task-specific systems that react to the most important information and ignore the rest. Daniel Dennett (1995) makes a similar point when he compares basic moral principles as useful "conversation-stoppers" which alleviate the need for "endlessly philosophizing, endlessly calling us back to first principles and demanding a justification". Thus, with the development of Type 2 thinking, a need to prevent the excessive interference of rational deliberation in the generally adaptive operation of evolved Type 1 task-solving systems emerged.

We hypothesize that this need emerging from the conflict of influence between Type 1 and 2 processing led to the co-evolution of a further system: the feeling of rightness *system* (hereafter FORs). The evolutionary function of the FORs is to facilitate humans' decision making. The FORs helps people to select optimally among competing evaluative responses that are simultaneously present in their minds. The FOR makes the responses produced by evolved task-specific systems more appealing. It thereby facilitates decision-making and biases it in favor of systems that have been selected for solving the task at hand, or a similar task. At the mechanistic level, as with any evolved mental system, the FORs responds to certain stimuli and produces distinct task-specific operations: the stimuli processed are conflicting evaluative responses; the operation performed is to give priority to the response that arises most quickly and fluently in the subject's mind because it is a reliable indicator for the activity of an adaptive task-solving system. Phenomenologically, the FORs expresses itself as an FOR, which pushes us to think that the rapidly-produced evaluation is correct.

Let us illustrate this phenomenon with the case of limerence, a term given to the heady experience of being-in-love (Tennov 1979). People in limerence experience intense feelings of love towards their partner, sexual attraction, obsessive thinking about the loved one, and a tendency to idealize and see only their good traits. They are likely to judge that their partner is the best possible match for them and that their love will last forever, despite their awareness that few couples remain happy in monogamous relationships for their whole lives, and despite previous disappointments. Limerence contributes to the forming of long-term pair bonds between two partners, an important condition for rearing human children given the length of gestation and the vulnerability of the human neonate (Fletcher et al. 2015). If a lover pauses and engages in a rational evaluation of her partner's traits, she may be forced to recognize that there are many potential partners out there who have similar traits, possibly in even greater measure. Why should she not question her judgment that her partner is the best possible match? In face of conflicting evaluation produced by the primitive limerence system and Type 2 reflective thoughts of this kind, the FORs is activated and renders the evaluation that is produced more quickly and fluently more compelling. This explains why people in the grip of

limerence, including philosophers, scientists, and relationship counsellors, tend to be intuitively convinced that their present lover is the best possible match for them, even when their rational mind tells them otherwise.

The FORs may reinforce evaluations that are adaptive but incorrect. Limerence is one example: the likelihood that a person has fallen in love with her best possible match is very slim, but believing this is an important bonding factor. A mother's distorted view of her child is another example. From an evolutionary point of view, mother-infant bonding is important because it secures maternal protection and caregiving, thereby ensuring children's survival, brain maturation, and social-emotional growth. Bonding in the mother expresses itself as a set of inclinations including seeking proximity and physical contact with her child, worry about her child's safety and welfare, and extensive mental focus on her child (Feldman et al. 2007). Bonding also includes a mother's inclination to think that her child is the most beautiful baby ever born (Feldman et al. 1999). Such a belief is both adaptive and objectively false in the vast majority of cases. In such situations, the FORs compels a mother to keep believing that her baby is the most beautiful, even when her analytic reasoning questions the reasonableness of this belief.

Once the FORs has evolved, it can also be recruited in support of rationally or culturally acquired operations. For instance, young children in the process of learning basic mathematics may not initially find it compelling that  $2 + 2 = 4$ . But after having become familiar with the calculation and understood the underlying logic of it, usually by physically adding objects together until the pattern "clicks," this type of calculation is processed in a Type 1 manner; whenever one asks them the sum of  $2 + 2$ , the answer comes quickly and fluently to their mind. At this point, if somebody tried to convince them that  $2 + 2 = 3$ , the FORs would be elicited, making them feel that 4 is the right answer. The same phenomenon can be observed with the example of the Canadian who feels intuitively that she should leave substantial tips even while visiting countries where this is not the custom. These examples illustrate a well-known phenomenon in evolutionary biology (Futuyma 2013): a system that has been selected for one task, here to facilitate optimal decision-making by preserving the adaptive responses of evolved task-solving systems, may be later recruited for other tasks, here to select among competing learned responses.

The FORs may also support incorrect and non-adaptive responses. In section 3, we mentioned the example of the incorrect answer to the mathematical widget question prompting a strong FOR in most people. These sorts of errors may occasionally induce non-adaptive decisions in humans. However, occasional failures in generating an adaptive answer is the fate of evolved traits. Indeed, for a trait to evolve it is sufficient that it produces adaptive responses in the most fitness-relevant contexts (Nettle 2006).

Our explanation of moral intuitions is more complete than alternative existing accounts. We make an important distinction between the phenomenological aspect (the FOR) and the mechanistic aspect (FORs) of intuitions. This distinction enables us to single out the eliciting factor for intuitions: they are generated by conflicting evaluations. Thus, any evaluation that feels intuitive has been *challenged* by a conflicting evaluation while a Type 2 process took place. Thanks to this account of intuitions, it is possible to explain why a quickly and fluently produced Type 1 evaluation sometimes feels intuitive and sometimes not. Imagine a situation (inspired from Greenspan 1988: Chapter 1) where a householder judges a salesman to be



trustworthy on the basis of the evidence available to her (Type 2 process), but stops short of signing a contract because she senses that the salesman is not to be trusted (Type 1 process). In such a situation, the householder experiences her suspicions as “intuitive” only because she is also tempted to evaluate the salesman as trustworthy; her suspicions are challenged by her reasoning that he has the correct identification documents and that her neighbor has recommended him. If all relevant information had concurred in confirming the impression of untrustworthiness, the householder would have quickly dismissed the salesman as untrustworthy, sent him away, and experienced no intuition about him. Similarly, we can routinely and quickly make the calculation  $2 + 2 = 4$  without experiencing it as intuitive. The characteristic phenomenology of intuition occurs only when the answer is challenged, e.g., by someone claiming that 4 is not the right answer.

Finally, our account highlights that the human faculty to experience intuitions is of general scope. The FORs is a blind task-specific system which may apply to non-moral domains, such as conflicting aesthetic and mathematical evaluations, in addition to moral domains, such as conflicting evaluations about how to behave in a social context. From an empirical point of view, there is nothing special about moral intuitions compared to other types of intuitions, aside from their subject matter. However, since this article focuses on moral intuitions, our next section is devoted to how the FOR applies to the moral domain.

## **5. The feeling of rightness applied to moral intuitions**

In sections 3 and 4, we argued that moral intuitions are composed of a challenged moral evaluation reinforced by a feeling of rightness (FOR). This feeling is produced by a psychological system (the FORs) whose eliciting factors are conflicting evaluations provided by different task-solving systems. The FORs prioritizes the response that comes quickly and fluently to the subject’s mind. In this section, we explain how the FORs functions in the case of moral intuitions, why moral intuitions are a frequent occurrence, and speculate on why humans are biased in favor of metaethical internalism and against moral relativism.

Much moral evaluation involves Type 1 thinking, which is influenced by genetically and culturally shaped systems, yet varies greatly from one person to another. There is empirical evidence that heritable systems are involved in determining which domains of interaction are considered suitable for moral evaluation, e.g., domains involving matters related to harming, fairness, loyalty, authority, and spiritual purity (Graham et al. 2011). To some extent, heritable systems even seem to influence the content of moral evaluations, e.g., in the direction of incest avoidance (Lieberman et al. 2003), or in the condemnation of violations of social exchange rules (Cosmides et al. 2010). However, despite their partial genetic heritability and the automaticity of their operations, the psychological systems involved in human moral evaluations are developed and shaped during human ontogenesis in various cultural environments (for a review, see Sunstein 2005). Consequently, significant variation can be observed among the Type 1 moral evaluations that are produced around the world and even between people from the same culture (Haidt & Graham 2007; Sachdeva et al. 2011).

In moral contexts, Type 2 thinking typically leads us to challenge our Type 1 evaluations and thus to experience intuitions. Type 2 processing can take more information into account than Type 1 processing, rendering people aware of challenges to their evaluations, which is a

trigger for intuitive experiences. For instance, to refer to our example from Section 1, when she is confronted by Aaron's criticism of her anti-immigrant stance, Sabrina may become aware that her views seem harsh and unkind from the immigrants' point of view; her FORs will be elicited, and since this system favors evaluations that are processed quickly and fluently, Sabrina will experience her Type 1 anti-immigration view as intuitive. At the end of the discussion, two scenarios are possible. Either Sabrina is thoroughly convinced that she is right, or she may rationally agree that Aaron is right, while still finding it intuitive to condemn immigration. In the latter case, her considered judgment conflicts with her intuition and she may have difficulties standing by her considered judgment in future circumstances unless she reinforces it in other ways. We can thus see how moral reflection and debate triggers intuitions. Since in contemporary pluralistic societies we often experience differing moral points of view, this explains why moral intuitions are so frequently experienced.

The FORs can favor learned and rational evaluations. As a result of life experiences, social pressure, or extensive private moral reasoning, people often come to modify their Type 1 evaluations. For instance, if Sabrina falls in love with an immigrant, she is likely to become able to understand his perspective and thus to sympathize with all immigrants, which may, over time, modify her Type 1 moral evaluations. Following her discussion with Aaron, she could also train herself to evaluate immigration positively, just as a convert to a particular brand of consequentialism may train herself to give equal moral weight to certain non-human animals and to mentally-impaired humans, in spite of her initial intuitions. Once these evaluative habits are firmly established as Type 1 processes, they are favored by the FORs whenever conflicting evaluations are revealed by a slower Type 2 form of thinking. This illustrates the fact that the FORs is a simple task-specific system that does not discriminate between evaluations according to how they have been produced, i.e. genetically, rationally, under the influence of life experiences. The FORs blindly tracks the rapidity and fluency with which evaluations are processed.

To conclude, we will briefly elaborate in a more speculative manner on how the FORs appears to bias our metaethical beliefs. Moral disagreement is common, thus the FORs is regularly activated and moral intuitions pervade moral discourse. Since the experience of an FOR tends to make people believe strongly in the correctness of their evaluation, frequent experience of the FOR in moral contexts renders it more likely that people believe that there are universal moral truths, which is a metaethical stance. Indeed, empirical data attest that when ordinary people make moral evaluations within their own culture, i.e. on matters they are closely concerned with, they tend to be very optimistic about the universal validity of their evaluations (Sarkissian et al. 2011). Moreover, as Joyce has pointed out, it may be practically useful to believe in universal moral validity because this belief silences selfish calculations and bolsters self-control against practical irrationality (Joyce 2005). However, as mentioned above, there is also significant variation among Type 1 evaluations and people regularly experience moral disagreements. This fact can tempt people to adopt a more relativist view of morality. Thus, facing the contradicting evaluations that moral beliefs are universally valid and relative, the FORs may be activated and is likely to render the universalist view intuitive because it is processed more fluently. Note, however, that this analysis is a description of what may happen in people's minds. It provides no direct argument in favor of moral relativism as theory.

Similarly, the FORs may bias people's metaethical beliefs about moral motivation. As demonstrated in empirical literature, people are usually very optimistic about the motivational power of moral evaluations (Björnsson et al. 2015), an attitude conducive to the view called "motivational internalism." This attitude is likely to result from Type 1 thinking. However, empirical evidence also shows that in the absence of additional external motivating factors, people are only mildly motivated to comply with the moral norms they openly advocate. Whenever their own interests are at stake, they tend to prefer strategic over moral behavior (Aarøe & Petersen 2013; Batson 2008; Dhimi, 2003; Talwar et al. 2004). They try to appear moral to others, while, if possible, avoiding the cost of being moral, a tendency that Daniel Batson calls "moral hypocrisy" (2008). These failures to follow moral norms are routinely experienced in our everyday life. Awareness of these facts while engaging in Type 2 reasoning may lead to disturbing conclusions: people, including oneself, may not be genuinely motivated to comply with the norms they openly hold! These contradictory Type 1 and 2 thoughts prompt the activation of the FORs which favors the Type 1 internalist belief.

People are not typically aware of the psychological mechanisms involved in their thoughts, such as whether their evaluation has been processed in a Type 1 or Type 2 manner, or whether it is reinforced by the FORs. Moreover, by restricting and biasing rational and objective assessment of quickly and fluently produced responses, the FORs enables people to resolve or overlook their mental inconsistencies with minimal discomfort. There is empirical evidence that people happily confabulate and produce post-hoc rationalizations for their moral intuitions (Haidt et al. 1993). It is thus possible that the FORs is one of the crucial mechanisms underlying the behavior reported by cognitive dissonance theorists (Festinger 1957).

## 6. Conclusion

We claim that there are two components to moral intuitions: a challenged evaluation and a metacognitive feeling of rightness (FOR). The emergence of the psychological mechanism responsible for the FOR can be explained by its clear evolutionary function in core cases: to facilitate optimal decision-making by preserving the adaptive responses of evolved task-solving systems.

Our account helps to bridge the gap between different conceptions of moral intuition in psychology and philosophy. On one hand, it recognizes the role of moral judgments, the most common form of moral evaluation, and of Type 2 thinking. On the other hand, it emphasizes the affective component of intuition: the FOR. Moreover, our account builds on the useful conception of moral intuitions as seemings and provides additional evolutionary and psychological information about the workings of this mental phenomenon. In particular, it highlights that the experience of intuition arises during Type 2 mental activity and only against a background of contradictory evaluations produced by different task-solving systems.

An interesting consequence of our account is that there is not much special about moral intuitions compared to other types of intuitions. All are produced by the same blind task-specific system, the FORs. Intuitions qualify as "moral" when the FORs applies to conflicting moral evaluations.

Finally, our account explains why intuitions are a frequent occurrence in the moral domain, why they remain so compelling and resistant in the face of reasons to discount them, and why humans are biased towards the ideas of universal moral truth and intrinsic motivation to follow moral norms.

## Acknowledgements

We thank Bernard Baertschi, Richard Dub, Benoît Dubreuil, Richard Joyce, and Fabrice Teroni for their useful comments, and the audience at *The Moral Domain: Conceptual Issues in Moral Psychology* conference in Vilnius, Lithuania.

## References

- Aarøe, L. & Petersen, M. 2013. "Hunger games fluctuations in blood glucose levels influence support for social welfare." *Psychological Science* 24: 2550–2556.
- Audi, R. 2009. *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton University Press).
- Barrett, H., Cosmides, L., & Tooby, J. 2010. "Coevolution of cooperation, causal cognition and mindreading." *Communicative & Integrative Biology* 3: 522-524.
- Batson, C. 2008. "Moral masquerades: Experimental exploration of the nature of moral motivation." *Phenomenology and the Cognitive Sciences* 7: 51–66.
- Bedke, M. 2008. "Ethical intuitions: What they are, what they are not, and how they justify." *American Philosophical Quarterly* 45: 253–269.
- Björnsson, G., Eriksson, J., Strandberg, C., Olinder, R., & Björklund, F. 2015. "Motivational internalism and folk intuitions." *Philosophical Psychology* 28: 715–734.
- Bernstein, I. 1999. "Taste aversion learning: A contemporary perspective." *Nutrition* 15: 229-234.
- Carruthers, P. 2007. "Simple heuristics meet massive modularity." In P. Carruthers, S. Laurence and S. Stich (eds.), *The Innate Mind: Culture and Cognition* (Oxford University Press) 181-198.
- Cohen, G. 2008. *Rescuing Justice and Equality* (Harvard University Press).
- Dennett, D. 1995. *Darwin's dangerous idea. Evolution and the meanings of life* (Penguin Press).
- DePaul, M. 2006. "Intuitions in moral inquiry." In D. Copp (ed.), *The Oxford Handbook of Ethical Theory* (Oxford University Press) 595-623.
- Dhimi, M. 2003. "Psychological models of professional decision making." *Psychological Science* 14: 175–180.
- Evans, J. & Stanovich, K. 2013. "Dual-process theories of higher cognition advancing the debate." *Perspectives on Psychological Science* 8: 223–241.
- Feldman, R., Weller, A., Leckman, J., Kuint, J., & Eidelman, A. 1999. "The nature of the mother's tie to her infant: Maternal bonding under conditions of proximity, separation,

- and potential loss.” *Journal of Child Psychology and Psychiatry and Allied Disciplines* 40: 929–939.
- Feldman, R., Weller, A., Zagoory-Sharon, O., & Levine, A. 2007. “Evidence for a neuroendocrinological foundation of human affiliation: Plasma oxytocin levels across pregnancy and the postpartum period predict mother-infant bonding.” *Psychological Science* 18: 965-970.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance* (Evanston: Row, Peterson).
- Fletcher, G., Simpson, J., Campbell, L., & Overall, N. 2015. “Pair-bonding, romantic love, and evolution: The curious case of *Homo sapiens*.” *Perspectives on Psychological Science* 10: 20-36.
- Foot, P. 1967. “The problem of abortion and the doctrine of double effect.” *Oxford Review* 5: 5-15.
- Futuyma, D. 2013. *Evolution* (3 edition) (Sinauer Associates, Inc.).
- Gigerenzer, G., Todd, P., & ABC Research Group. 1999. *Simple Heuristics that Make Us Smart* (Oxford University Press).
- Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. 2011. “Mapping the moral domain.” *Journal of Personality and Social Psychology* 101: 366–385.
- Greene, J. 2007. “The secret joke of Kant’s soul.” In W. Sinnott-Armstrong (ed.), *Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (MIT Press) 359-372.
- Greenspan, P. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification* (Routledge).
- Haidt, J. & Bjorklund, F. 2008. “Social intuitionists answer six questions about moral psychology.” In W. Sinnott-Armstrong (ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* (MIT Press) 181-217.
- Haidt, J. & Graham, J. 2007. “When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize.” *Social Justice Research* 20: 98-116.
- Haidt, J., Koller, S., & Dias, M. 1993. “Affect, culture, and morality, or is it wrong to eat your dog?” *Journal of Personality and Social Psychology* 65: 613–628.
- Huemer, M. 2008. *Ethical Intuitionism* (Palgrave Macmillan).
- Joyce, R. 2005. “Moral fictionalism.” In M. Kalderon (ed.), *Fictionalism in Metaphysics* (Oxford University Press) 287-313.
- Kamm, F. 2006. *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (Oxford University Press).
- Kauppinen, A. 2013. “A Humean theory of moral intuition.” *Canadian Journal of Philosophy* 43: 360–381.
- Knoblich, G., Butterfill, S., & Sebanz, N. 2011. “Psychological research on joint action: Theory and data.” In B. Ross (ed.), *The Psychology of Learning and Motivation* (Elsevier) 59-101.
- Lieberman, D., Tooby, J., & Cosmides, L. 2003. “Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest.” *Proceedings of the Royal Society B: Biological Sciences* 270: 819–826.

- McMahan, J. 2013. "Moral intuition." In H. LaFollette & I. Persson (eds.), *The Blackwell Guide to Ethical Theory* (2 edition) (Wiley-Blackwell) 103-120.
- Nettle, D. 2006. "The evolution of personality variation in humans and other animals." *American Psychologist* 61: 622-631.
- Rawls, J. 1971. *A Theory of Justice* (Harvard University Press).
- Roeser, S. 2011. *Moral Emotions and Intuitions* (Palgrave MacMillan).
- Sachdeva, S., Singh, P., & Medin, D. 2011. "Culture and the quest for universal principles in moral reasoning." *International Journal of Psychology* 46: 161-176.
- Sarkissian, H., Park, J., Tien, D., Wright, J., & Knobe, J. 2011. "Folk moral relativism." *Mind & Language* 26: 482-505.
- Schwartz, B. & Metcalfe, J. 2011. "Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience." *Memory & Cognition* 39: 737-749.
- Shafer-Landau, R. 2005. *Moral Realism: A Defense* (Clarendon Press).
- Singer, P. 2005. "Ethics and intuitions." *Journal of Ethics* 9: 331-352.
- Sinnott-Armstrong, W. 2008. "Framing moral intuitions." In W. Sinnott-Armstrong (ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* (MIT Press) 47-76.
- Sunstein, C. 2005. "Moral heuristics." *Behavioral and Brain Sciences* 28: 531-542.
- Tennov, D. 1979. *Love and Limerence: The Experience of Being in Love* (2 edition 1998) (Scarborough House).
- Tersman F. 2007. "Fresh air?" In T. Rønnow-Rasmussen (ed.), *Hommage à Wlodek Philosophical papers dedicated to Wlodek Rabinowicz*. [www.fil.lu.se/hommageawlodek](http://www.fil.lu.se/hommageawlodek)
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. 2004. "Children's lie-telling to conceal a parent's transgression: Legal implications." *Law and Human Behavior* 28: 411-435.
- Thompson, V. 2009. "Dual process theories: A metacognitive perspective." In J. Evans & K. Frankish (eds.), *Two Minds: Dual Processes and Beyond* (Oxford University Press) 171-195.
- Tucker, C. 2013. "Seemings and justification: An introduction." In C. Tucker (ed.), *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism* (Oxford University Press) 1-29.