



# An Expected Error: An Essay in Defence of Moral Emotionism

Justin J. Bartlett<sup>1</sup>

Received: 1 July 2020 / Accepted: 12 October 2020 / Published online: 17 October 2020  
© Springer Nature B.V. 2020

## Abstract

This work draws an analogical defence of *strong emotionism*—the metaethical claim that moral properties and concepts consist in the propensity of actions to elicit emotional responses from divergent *emotional perspectives*. I offer a theory that is in line with that of Prinz (The emotional construction of morals. Oxford University Press, Oxford, 2007). I build an analogy between moral properties and what I call *emotion-dispositional properties*. These properties are picked out by predicates such as ‘annoying’, ‘frightening’ or ‘deplorable’ and appear to be uncontroversial and frequent cases of *attribution error*—the attributing of subjective emotional states as mind-independent properties. I present a linguistic analysis supporting the claim that moral properties and their related concepts are reducible to a subset of emotion-dispositional properties and concepts. This is grounded in the observation that utterances featuring moral predicates function linguistically and conceptually in analogous ways to emotion-dispositional predicates. It follows from this view that asserted moral utterances are truth-apt relative to ethical communities, but that speakers misconceive the extensions of predicates. I show how the framework of Cognitive Linguistics allows us to explain this error. Further analysis of moral and non-moral utterances exposes the deeper conceptual schemas structuring language through cognitive construal processes. An understanding of these processes, coupled with an emotionist elucidation of moral properties and concepts, makes the attribution error an expected upshot of the emotionist thesis, rather than an uncomfortable consequence.

**Keywords** Metaethics · Moral psychology · Cognitive linguistics · Emotion · Moral concepts

---

✉ Justin J. Bartlett  
justinbartlettswu@gmail.com

<sup>1</sup> Srinakarinwirot University, Bangkok, Thailand

## 1 A Difficult Relationship

Much has been said about the relationship between emotion and morality. Aside from the academic work of philosophers, psychologists and sociologists, the relationship between the two is even a well-established notion in folk intuition. Quotidian experience tells us that the transgressions of ethical rules and moral disagreements are often followed with emotionally charged reactions and outbursts, thus making the connection between the moral and the emotional particularly salient. But is the relationship between morality and emotion simply a reactionary one? Do moral properties give rise to emotional responses or do emotional responses give rise to moral properties? This question is not simply a case of chiasmus, a cheap rhetorical device. Its implications have great philosophical weight. Exactly *how* our emotions interact with, or constitute, morality is still a topic of fierce debate among philosophers and psychologists alike; a wide range of metaethical and psychological theories attempt to explain the relationship between moral judgement and emotion. The theoretical melee rages, and discerning the stronger opponent is not straightforward. One observation, however, that is fairly easily made, is how rarely empirical findings from the field of moral psychology, and its related disciplines, enter into the metaethical arena. This has been noted previously by other authors such as Prinz and Nichols who write that:

[E]ven where moral philosophers have invoked emotions, they seldom attend carefully to the psychological characteristics of the emotions to which they appeal. Indeed, it would be hard to exaggerate the extent to which philosophers, even self-described sentimentalists, have neglected psychological research on the moral emotions (Prinz and Nichols 2010).

Our question, here, is an empirical one. This is not to say, however, that a priori deduction or phenomenological analysis are not effective methods of investigation, but rather to point out that answering such a question fully and accurately will inevitably need empirical support as well as philosophical insight. It is my aim in this paper, therefore, to answer the question of whether moral properties give rise to emotional responses, or vice versa by offering an analysis of emotion's necessary role in the formation of moral properties, concepts and judgements. This work is interdisciplinary in nature and makes use of work from psychology and Cognitive Linguistics in supporting the philosophical claim that emotional properties are necessarily conceptually prior to, and constitutive of, moral properties. This is the so-called *emotionist*<sup>1</sup> claim, or rather, a version of it. In the following paragraphs, I assume a strong emotionist stance towards morality which makes the following claim and related assumptions:

*Moral judgements* are those judgements referring to any conduct  $\phi$  perceived as having the propensity to elicit certain emotional responses in themselves and observers, in certain situations. Such that, a judgement  $\alpha$  is a moral

---

<sup>1</sup> I follow Jesse Prinz's terminology. See Prinz (2007).

judgement *iff* the conduct referred to by  $\alpha$  is perceived as having the propensity to elicit certain emotional states in agent  $x$  and observer  $y$  in  $S$ .

This elucidation of moral judgements makes two implicit assumptions:

- i. **Metaphysical Claim:** Moral properties can be understood as consisting in the propensity of certain actions to elicit moral emotional reactions in agents and observers.
- ii. **Epistemic Claim:** Grasping moral concepts consist in having the predisposition to experience the emotional reactions alluded to in the metaphysical assertion.

I will proceed to build a defence for the strong emotionist elucidation of moral properties, concepts and judgements as outlined above by presenting an analogical argument highlighting the similarities between moral language and more general emotional language which, based on the assumption that language structure mirrors conceptual structure, supports the conclusion that moral properties can be understood as a subset of emotional properties.

Firstly, I will lay the philosophical foundations of such a theory (Sect. 2) and offer an overview of the psychological evidence which supports a necessary link between emotion and morality (Sect. 3). I will then give further support for this reading of moral judgements by offering an analysis of language which draws strong parallels between moral language and non-moral emotional language. It will be made clear from an analysis of everyday utterances about emotional dispositions, that humans frequently make systematic *attribution errors*<sup>2</sup> when predicating *emotion-dispositional*<sup>3</sup> properties of extramental entities. I will endeavour to show how widespread and pervasive this phenomenon is in ordinary emotional language, and, furthermore, that this is not merely a linguistic mistake but is also, at its foundation, a conceptual one. An analogy will be made such that there can be no *in-principle* objections to systematic attribution error when predicating emotion-dispositional properties of extramental entities. This emotional language use will be compared to moral language and an argument by analogy will be employed supporting the conclusion that moral properties and concepts can be reduced to emotional properties and concepts (Sects. 4 and 5). Lastly, I will briefly offer a plausible explanation as to why we are prone to making this kind of error by employing the conceptual framework of Cognitive Linguistics (Sect. 6). The view I offer here shows that the root of this linguistic and conceptual error is easily explained by making reference to the influence of *construal phenomena*, which have been proposed and explicated by linguists such as Langacker (1987), Talmy (1978)

<sup>2</sup> The term ‘fundamental attribution error’ is normally used to refer to the tendency of observers to attribute an agent’s wrong behavior as resulting from their fundamental character traits. Here, I re-purpose this term for the present argument. However, as will be made clear, it still retains much of its original character.

<sup>3</sup> Here I will use the term ‘emotion-dispositional’ in order for the concept to stand out against other properties, such as red and bitter which are considered to be ‘response-dependent’ or secondary properties. See (D’Arms and Jacobson (2007)). Emotion-dispositional properties are response-dependent but have the propensity to elicit emotional reactions whereas standard response-dependent properties, such as ‘red’ typically do not.

and Croft and Wood (2000). It will be shown that an understanding of construal phenomena and how they affect language is instrumental in accounting for the prevalence of attribution error in emotional and moral language.

## 2 Philosophical Foundations

The view of moral judgments defended here can be referred to as a strong emotionist account of moral judgements. A strong emotionist theory is one which makes both the metaphysical claim that ‘moral properties are essentially related to emotions’ and the epistemic claim that ‘moral concepts are essentially related to emotions’ (Prinz 2007; p 14–16). Claims of the former metaphysical kind can be seen in theories which liken moral properties to response-dependent properties such as colour (McDowell 1985). The latter epistemic kind can be seen in Allan Gibbard’s work (Gibbard 1990, 2006) where he defines moral concepts in emotional terms. Understanding such concepts, therefore, consists in recognizing the potential emotional responses caused by certain acts. Gibbard (2006) states that moral concepts of ‘wrong’, for example, can be defined as those which lead to the elicitation of such emotions as guilt on the part of the transgressing agent, and which will be seen as *reprehensible* on the part of the observer; that for an action to be considered wrong, it must be one for which, at least, a guilty response is *warranted* by the agent. Hence, we have an account of morality which holds that moral concepts can be defined in terms of the actual and potential emotional dispositions elicited:

I say that the basic narrowly moral concept is being blameworthy or reprehensible. That an action is reprehensible just means that reprehension over it on the part of others and guilt over it on one’s own part are warranted. (Gibbard 2006).

A strong emotionist theory is one which combines both metaphysical and epistemic claims outlined in (i) and (ii) above. The work of Jesse Prinz (2007; 2006) provides an example of a strong emotionist theory, but other similar theories are to be found in sensibility theories (Darwall et al. 1992; McDowell et al. 1985). Sensibility theories, make the claim that moral properties are ‘response-dependant’. An analogy is often made with colour, the perception of which is supposed to be dependent upon the agent experiencing the property (D’Arms and Jacobson 2007). In McDowell’s words, moral properties are to be understood as ‘qualities not adequately conceivable except in terms of certain subjective states’ (Mc Dowell 1985, p.136). Hence, we can see moral properties as defined by, their propensity to elicit certain moral emotional reactions such as guilt, shame, remorse or compassion. Moral concepts, therefore, consist in the predisposition to experience these emotional reactions, and it follows from this that moral judgements can be defined as judgements about the propensity of an action to elicit these emotional responses.

Consequently, I claim, moral utterances of the form ‘*x* is *P*’ express belief-like states and are truth-apt propositions. We avoid an error theory as moral predicates,

thus understood, have as their extension certain forms of conduct. We could say that conduct perceived as having a certain ‘character’—a particular character recognized as having the propensity to elicit certain emotional responses in agents and observers—is the extension of moral predicates and, accordingly, propositions containing such predicates, do refer properly to states of affairs. The truth conditions of such judgements, however, are relative to a particular moral community—a group that shares the same values. The moral community is the set of all people who share a particular moral judgement.<sup>4</sup> The truth of a moral proposition will, therefore, be relative. There may, however, be some judgements which are universally accepted if shared by all humans, but this would be a contingent fact. Such a theory, therefore, can be said to be *minimally* realist (Van Roojen 2015, p 116). However, although we avoid a semantic error theory, we do not avoid a conceptual one, as I claim that speakers mistakenly perceive moral properties as objective mind-independent and non-reducible properties.

Emotionist or sensibility theories have, of course, been challenged and one such objection comes from Harman in response to Gibbard (2006) who, as seen above, proposes a guilt-focussed account of moral concepts. Harman (2009) warns against positing a definitional explication of moral concepts, especially in the case of guilt, on the grounds that it is possible for an agent to commit a morally reprehensible action and yet not feel guilty. Harman argues that a necessary link cannot be true a priori and that moral concepts cannot be defined in terms of guilt feelings due to the fact that it is possible to be motivated to act morally by other feelings such as empathy or compassion. Harman’s objection works against guilt-focussed accounts of moral concepts but not with emotionist theories in general and I will claim, due to these considerations, that an emotionist account of morality must encompass all *moral* emotional states.

Prinz (2006, 2007) has defended a strong emotionist thesis by pointing to the wealth of empirical data from psychology and neuroscience as support, and it is my view that, in light of the empirical data, such an account should indeed be taken seriously.

### 3 Psychological Evidence

It appears, intuitively, that a connection between emotion and morality holds, at least as a causal relation. Generally, emotions can be seen as a psychological and physiological reaction to a personally-significant stimulus:

[Emotion is] a complex reaction pattern, involving experiential, behavioral, and physiological elements, by which an individual attempts to deal with a personally significant matter or event. The specific quality of the emotion (e.g., fear, shame) is determined by the specific significance of the event. (*Emotion—APA Dictionary of Psychology*).

<sup>4</sup> This definition needs to be formalised but this rough characterisation will suffice for our current purpose.

We can see that such a causal relationship between stimulus and reaction holds, *prima facie*, for moral emotions too. For instance, if I have acted in a way which you and I both judge to be wrong, then I expect you to respond with anger, and, furthermore, I expect to feel guilty for my transgression (if not for the action in itself, then at least for upsetting you.) We do, then, for the most part, appear to have predictable emotional responses to the transgression of moral rules from both agent and observer. Armchair speculation about responses to what we might broadly call *moral stimuli* do not yet, however, constitute the grounds for a compelling emotionist argument, but they do provide us with a jump-off point; one which is strongly supported by empirical evidence.

The link between emotion and morality is widely accepted in the psychological literature, and we can even see anger, which is typically understood as being a non-moral emotion, profiled in vaguely moral terms; as an emotion which arises when, injustice or transgression is detected. As Prinz and Nichols (2010) point out:

In the recent moral psychological literature on anger, the familiar characterization of the profile of anger is that it's caused by a judgment of transgression and it generates an inclination for aggression and retributive punishment [...] Much of the evidence for this profile comes from work in social psychology over the last two decades. (Prinz and Nichols 2010).

Apart from emotional profiling and categorization, the psychological literature seems to give strong support for a link between morality and emotion which shows emotion to be more than simply an output or reaction. There are two prominent ways in which emotion has been connected to morality in the psychological and neuroscientific literature: (i) in interfering with, or guiding, moral decision making, and (ii) in being a constituent part of the moral judgement-making process. I will offer an overview of the work which documents these phenomena in turn.

An important and oft-cited example of how emotional centres of the brain affect moral decision making can be seen in the work of Greene and colleagues (2001) who observed neural activity in two distinct decision-making systems when subjects contemplated moral dilemmas. Greene et al. used fMRI scans to identify the areas of the brain used when making decisions in moral dilemmas and found clear activity from emotional centres of the brain when participants contemplated the scenarios. This was coupled with reduced activity in the pre-frontal lobe and structures associated with rational and logical thought, when participants were asked to contemplate more 'personal' moral dilemmas (Green 2001). This work is widely interpreted as showing that moral decision-making processes are, at least in part, influenced by emotional stimuli,<sup>5</sup> and seems to provide evidence for a connection between morality and emotion which is more than simply reactional.

In addition, there is evidence which suggests that emotional states actually guide or antecedently influence people's moral judgements. Isen and Levin (1972), for instance, have shown that inducing positive emotional states in subjects, makes

---

<sup>5</sup> Greene et al.(2001) even suggests that this evidence show that classic consequentialist and utilitarian ethics require rational thought processes whereas deontological thinking is driven by the non-rational emotional structures of the brain.

them more likely to help others, suggesting that emotional priming encourages pro-social behaviour. In another study, where people were asked to evaluate the wrongness of certain actions, participants who were sitting at a dirty desk gave consistently higher ‘wrongness ratings’ as compared with their counterparts who were sitting at clean desks (Schnall et al. 2008). Schnall et al.’s observations show the corollary of Isen and Levin’s; that the moral decision-making process can also be affected negatively when emotions such as disgust are present. This has also been corroborated by similar observations made by Wheatley and Haidt who induced disgust in subjects hypnotically (Wheatley and Haidt 2005). Of course, these findings do not provide conclusive evidence but do give compelling empirical support for a link between emotion and morality at the level of decision-making, such that emotion is shown to be more than just a reaction to moral stimuli, but also being actively involved in the moral decision-making process antecedently influencing moral judgement.

In addition to the work mentioned above, there are further empirical studies which give more direct support to the current argument about moral concepts, and these studies examine moral judgement in subjects without prior emotional priming. Before I move on, it is important to note that the examination of moral *judgement*, as opposed to moral *decision-making processes*, at the neurophysiological level, has a more direct bearing on the emotionist claim—that moral concepts can be reduced to emotional concepts. This, of course, is due to it being imperative to the making of a judgement that one understands the concepts employed by the judgement; that to be able to make a moral judgement, one must first grasp moral concepts. It can be concluded, then, that if making moral judgements requires the use of emotional structures in the brain, then moral concepts are likely to be emotional in nature. Evidence for this could come in two forms: evidence showing the involvement of emotional brain structures in the process of moral judgement making, and, conversely, evidence showing that individuals with emotional impairment fail to make moral judgements or grasp moral concepts. Both types of data are, in fact, available and I will briefly mention some examples here.

Studies by Moll et al. (2001, 2002) found that when participants made judgements about moral statements such as ‘we break the law when necessary’, distinct regions of the brain were activated including the frontopolar cortex, medial frontal gyrus, right anterior temporal cortex and the cerebellum, which is normally associated with emotional regulation (Schmahmann and Caplan 2006), that were not active when processing non-moral statements such as ‘stones are made of water’.

Even more promising, perhaps, is the data gathered from investigations into psychopathy, and individuals with antisocial personality disorder, which has shed considerable light on moral judgement-making. The results of such studies are indeed pertinent to the current argument. Antisocial personality disorder is characterized by the APA as ‘the presence of a chronic and pervasive disposition to disregard and violate the rights of others. Manifestations include repeated violations of the law, exploitation of others, deceitfulness, impulsivity, aggressiveness, reckless disregard for the safety of self and others, and irresponsibility, accompanied by lack of guilt, remorse, and empathy’ (*Antisocial Personality Disorder – APA Dictionary of Psychology*). As can be seen from this delineation, a

lack of prosocial moral emotions such as guilt, remorse and empathy are outstanding features of psychopathy and so we might expect such individuals, following our claims about the emotional nature of moral concepts, to lack competence with moral concepts; that their emotional impairment would mean moral impairment. Indeed, it has been shown, in line with this thought, that individuals with antisocial personality disorder fail to grasp the so-called moral/conventional distinction, meaning that they are not able to distinguish moral transgressions, such as stealing, from conventional non-moral transgressions such as talking in class or not waiting in line (Blair et al. 2016; Blair and Cipolotti 2000; Liao 2016). Interestingly, further research into the neurotransmitter serotonin, which has been found to be lacking in individuals who display characteristics typical of psychopathy such as callous-unemotional (CU) traits (Dolan and Anderson 2003; Soderstrom et al. 2001, 2003), has shown that serotonin function covaries positively with prosocial behaviours and negatively with antisocial behaviours, suggesting that emotional neurochemistry might play an intrinsic role in moral appraisal and action.

In sum, there is a significant body of empirical evidence which shows the involvement of emotional regions of the brain in various moral thinking processes and, more importantly for the current argument, that emotional states are not simply reactions in response to moral transgressions but also aid and influence moral judgements antecedently. Furthermore, in the case of psychopathy, emotional impairment correlates with an inability to make the moral/conventional distinction which is indicative of the inability to grasp certain moral concepts. Such evidence, although not conclusive, I believe licenses much optimism for a strong emotionist reading of morality.

#### 4 Moral Emotions and Perspectives

I have, so far, been using the term ‘emotion’ in a rather broad sense. Before building my defence of emotionism, it is necessary to give a more detailed explication of the emotional import that the current theory proposes is present in moral judgments. In order to do this, we need to make reference to three distinct groups of *moral* emotions. Firstly, there are *pro-social* emotions such as sympathy, empathy and compassion, which are assumed to promote positive moral behaviour. Secondly, there are *self-blame* emotions such as guilt and shame which are directed inwardly and suffered by the agent upon transgression. And, thirdly, *other-blame* emotions which include contempt, anger and disgust which are experienced by the observer and are directed outwardly at the transgressing agent. Understanding these distinct emotions and the divergent *perspectives* from which they arise is important for the foregoing argument. The three categories of pro-social, self-blame and other-blame form a triad of possible perspectives from which a moral action can be appraised: observer (agent-focussed), agent (self-focused) and observer (victim-focussed). Pro-social emotions emerge from an observer’s victim-focussed appraisal and are assumed to motivate morally good actions. Self-blame emotions represent negative self-focussed appraisal of the agent from the perspective of the agent themselves. These are assumed to motivate the avoidance of, and are occurrent from, actions



perceived by the agent as being morally wrong. And, other-blame emotions represent the perspective of the agent-focussed observer and are assumed to motivate retaliation, retribution and blame of the agent. With the distinct moral emotions, we have three divergent perspectives from which moral emotion can be elicited and these, in turn, provide separate emotional reactions.<sup>6</sup>

With an understanding of the moral emotions and their divergent perspectives in-hand, we can now define moral judgements in virtue of their propensity to elicit these specific emotional responses. We can, therefore, define moral judgements as:

*Moral judgements* are those judgements referring to any conduct  $\phi$  perceived as having the propensity to elicit moral emotional responses in themselves and observers, in certain situations. Such that, a judgement  $\alpha$  is a moral judgement *iff* the conduct referred to by  $\alpha$  is perceived as having the propensity to elicit moral emotional states in agent  $x$  and observer  $y$  in  $S$ .

This formulation of moral judgements follows, as shown above, from the metaphysical claim that moral properties consist in the propensity of an action to elicit moral emotional states, and the epistemic claim that the disposition to experience such moral emotions is to grasp moral concepts. Actions, therefore, will be perceived as having ‘moral’ properties (rightness/wrongness) if they are understood as having the propensity to elicit the appropriate moral emotional responses in agents and observers. Hence, sincerely assenting to a moral judgement consists in a recognition that a particular act has the propensity to elicit moral emotions.

In the following section, I will argue that moral properties and concepts, understood thusly, are analogous to myriad other dispositional properties and their corresponding concepts which we ascribe in error, and that moral properties appear to be a subset of such properties.

## 5 Emotion-Dispositional Properties

The wealth of empirical evidence garnered from the work of moral psychologists does indeed improve the soundness of our claims, but although it suggests heavy influence from emotional structures on higher-level cognitive functioning, we cannot use it to directly support either the metaphysical or the epistemic claims posited here. It is, then, in addition to the empirical data, necessary to provide observations which will bolster our claims about moral properties and concepts more directly.

As briefly alluded to above, the emotionist takes a view of moral properties to be broadly in line with sensibility theories (D’Arms and Jacobson 2007). Sensibility theories come in many flavours but have their origins in the work of David Hume.<sup>7</sup>

<sup>6</sup> It should also be noted that moral transgression might produce collateral non-moral emotions such as fear of repercussions and sadness in the agent, thus providing further motivation to avoid transgressive behaviour.

<sup>7</sup> I invoke Hume simply as a historical reference point and am not claiming that the current understanding of dispositional properties is ‘Humean’ at all.

Such readings often make use of an analogy between moral properties and colour properties. Colour, it is said, is not a mind-independent objective property but is rather a perceptual one—merely the subjective experience of the way human perceptual systems interpret light in the optic array. Redness, then, is a ‘dispositional’ property, and not a mind-independent one, like mass for instance. It is held that colour concepts, therefore, consist in being able to experience these dispositions. This, it is often suggested, is how we should understand moral properties too—as being dispositional in the way that colours are. We do indeed have a helpful analogy here with colour, and it has clear pedagogical merit. But this, I fear, is the full extent of its efficacy. This is because colour properties are in fact very distinct from the moral properties of rightness and wrongness, in some important respects. Colour is visually perceived; moral properties are not. Colour is emotionally inert; moral properties are not. The analogy with colour, then, is a useful tool for grasping the concept of dispositional properties but has little argumentative force when explaining moral values.<sup>8</sup> There are, however, alternative analogies to be drawn from other groups of dispositional properties that mimic more closely our understanding of moral properties. These, I claim, provide not only a more effective analogy but can also exert some explanatory and argumentative leverage on the phenomenon currently under analysis. One such example can be seen in the case of ‘annoyingness’. As with colour, we routinely ascribe annoyingness to extramental objects—most commonly people—and states of affairs. We are, in fact, very prone to saying things such as ‘he is annoying’, ‘this song is annoying’ or ‘the situation is annoying’ and in doing so we attribute the property of ‘annoyingness’ to people, objects and situations as if it were an objective and intrinsic part of their physical existence; as if it were ‘out there’. However, this is incorrect. Being annoyed, of course, is an emotional disposition; it consists solely in our being annoyed. Hence, being annoyed is a subjective emotional state which we misinterpret and wrongly attribute as being a property of extramental entities when uttering such things as ‘Smith is annoying’, for instance. I will call this an *attribution error*. Smith’s actions or mannerisms *do* have a particular character which induces an annoyed state in us but this character is not objective and is dependent upon the reaction this character elicits in the observer; it is a response-dependent property. In fact, what we really describe when we utter ‘Smith is annoying’ is ‘Smith makes me feel annoyed’ which is felicitous and, despite its uncommon and awkward form, correctly assigns ‘annoyingness’ to the speaker as a subjective emotional response. This makes our utterance of ‘Smith is annoying’ a mistake because being annoying is not something that anything can be, in and of itself. This mistake, however, is not simply a linguistic one, but also a sincere conceptual one and it shows through in the surface grammar of our utterance.<sup>9</sup>

<sup>8</sup> See Railton (2003) for a further objection to the analogy from colour.

<sup>9</sup> In *Inventing Right and Wrong* (1977: p 42) Mackie puts forward a similar example with the property of being disgusting. Mackie, unfortunately, spends little time on this line of thought and his example of disgust is also subtly different from annoyingness as being disgusted by the sight of vomit, the smell of rotting flesh or the taste of a fungus may well be an automatic or hard-wired reflex.

As per the emotionist view, we understand such utterances not as statements describing the properties of extramental objects, but as statements about one's own emotional state or disposition in response to those entities or actions. However, although it takes just a moment's reflection to realize that annoyingness is in the eye of the beholder, I claim that, under normal circumstances, people *do* in fact believe what their utterances pretend to be; that Smith *is* annoying; Smith *has* that property. If this were not the case, why would we not more commonly say 'I am annoyed' at witnessing Smith's actions or 'Smith's actions annoy me?' I believe that we make this mistake sincerely and that 'Smith is annoying' is not simply a more convenient grammatical form. Such utterances should be taken as indicative of the way people perceive the world. Evidence in support of this claim can be found when looking at normal conversation patterns. Upon declaring that someone is annoying, we normally expect our friends to agree. If they do not, we attempt to convince them 'you don't think so? But he *really is!* He is annoying because...'. But, no amount of justifying will convince you that he is annoying, if, that is, you don't already think so, because no amount of explaining is likely to change your feelings and dispositions. Such examples, uncover a systematic attribution error that is both linguistic and conceptual in nature.

It is indeed plausible that moral predicates function in the same way and I propose that this is how they are best understood. When comparing the conceptual features and linguistic functioning of moral discourse with emotion-dispositional discourse we find some striking similarities. For example, when one says 'that is wrong' the surface grammar of the utterance suggests that one intends to ascribe a mind-independent property to an extramental entity which one believes to be true—as is the case with annoyingness. And, furthermore, we are surprised if others do not agree with our moral judgement and are likely to make an attempt at convincing them that our view is correct—as with annoyingness. It is also the case that—as with annoyingness—no amount of convincing will sway people if they are not already similarly predisposed.

If my analysis is right and moral properties do turn out to be emotion-dispositional properties like annoyingness, this all amounts to a systematic error in our language and with our understanding of these properties. This is the sort of conclusion that many want to avoid. But, however displeasing this conclusion might be, the case of annoyingness gives us reason to believe that it is, at least, possible for us to be mistaken in such a way. Moreover, due to the clear links between morals and emotions, moral properties can plausibly be understood as being a special case of the general emotion-dispositional properties just described. To put this characterization more succinctly, in the case of annoyingness, I think it fairly uncontroversial to claim that:

1. Annoyingness is an emotion-dispositional property and is thus constituted solely by the relevant emotional response to a stimulus.
2. Having the disposition to be annoyed is a necessary condition for grasping the concept of annoyingness.

3. Annoyingness judgements of the form ‘ $x$  is annoying’ are those judgements referring to entities perceived as having the propensity to elicit the relevant emotional response of annoyance in observers.

The analogy of annoyingness might only be slightly better than the classic colour analogy if it were not for the fact that this is only one example in a wide range of emotion-dispositional properties which we routinely attribute in error. Taking a moment to reflect, it becomes clear that such cases are frequent in English.<sup>10</sup> When we use participle adjectives to say that something, or someone, is annoying, frustrating, boring, depressing, scary, confusing, sexy, disgusting, deplorable, disappointing, stressful, relaxing, nerve-wracking, frightening, awesome, inspiring, shocking or infuriating<sup>11</sup> we make the same mistake; we are falsely, but sincerely, attributing a subjective emotional disposition to an extramental entity. In light of this, I claim, there can be no in-principle objection to systematic attribution error in the case of moral judgements and, moreover, that the principle of charity starts to seem far too generous. The cases mentioned above are not simply perceptual errors in the way that colour is supposed to be, but they are also all cases of ascribing an emotion-dispositional property as if it were a mind-independent one.

So far, I have shown that various emotion-dispositional properties exist which are not purely extramental features of the world. Such properties consist in emotional reactions to relevant stimuli, and being predisposed to experience these emotional reactions is a necessary condition for grasping emotion-dispositional concepts. I have further claimed that moral properties and concepts mimic the conceptual and linguistic behavior of emotion-dispositional properties and concepts, giving us reason for optimism about an emotionist reading of morality, and supplying us with a structure on which to model moral properties, concepts and judgements. If moral properties are emotion-dispositional properties then from claims 1, 2 and 3 above, 1\*, 2\* and 3\* follow

- 1\* As emotion-dispositional properties, ‘rightness’ and ‘wrongness’ are constituted by the propensity of an act to elicit moral emotional responses in agent and observer.
- 2\* Being predisposed to experience the relevant emotional responses is a necessary condition for grasping the concepts of rightness and wrongness.
- 3\* Moral judgements of the form ‘ $\varphi$  is right/wrong’ are those judgements referring to conduct  $\varphi$  perceived as having the propensity to elicit moral emotional responses in agents and observers, in certain situations.

I have claimed that moral properties, concepts and judgements can be understood as being a subset of the wider set of emotion-dispositional properties and can be labelled as *moral emotion-dispositional* properties. There is, however, an immediate issue with our theory of moral emotion-dispositional properties. Given that this

<sup>10</sup> Other languages need to be considered. It appears that in Thai language emotion-dispositional properties are attributed in error in a similar manner.

<sup>11</sup> This is not an exhaustive list but I think it makes my point.

reading is true, we are convicting humans of routinely making systematic conceptual and linguistic errors. It remains to be explained just *why* we do this.

## 6 Through the Lens of Cognitive Linguistics

We are obliged, *ex hypothesi*, to tackle the now looming question of ‘*why?*’. Given that the above account is true, why is it that we make these sorts of systematic attribution errors? Why do we act *as if* emotion-dispositional properties were objective, when they are not? A more instructive way of phrasing this question would be “What is the difference between uttering ‘Smith is annoying’ and ‘I feel annoyed with Smith’?”. Being able to give a plausible response to this question will be needed to bolster our theory, and this is the task I will now undertake. In order to do this, it will be necessary to take a short but enlightening diversion into the realm of Cognitive Linguistics.

One of the foundational assumptions underlying the cognitive linguistic framework is that linguistic structures are the upshots of, or mirror, more fundamental conceptual structures, as Croft and Cruse (2004) put it:

[L]anguage is not an autonomous cognitive faculty. The basic corollaries of this hypothesis are that the representation of linguistic knowledge is essentially the same as the representation of other conceptual structures, and that the processes in which that knowledge is used are not fundamentally different from cognitive abilities that human beings use outside the domain of language. (Croft and Cruse 2004).

It is asserted that language is emergent; it is a function of more general cognitive processes and conceptual structures. Such an assertion is grounded in the fact that no specific structure in the brain has been found to be responsible for language (Anderson and Lightfoot 2002). Cognitive linguists argue that general cognitive and conceptualization processes underlie not *only* semantic representation but that syntax, morphology, and phonology are also generated by and grounded in general cognitive functioning.

Some theories in Cognitive Linguistics have taken inspiration from, and made use of, theories from other areas of cognitive psychology. One such example is borrowed from Gestalt psychology. The Gestalt theory of perception was first concerned with identifying and understanding the various ways in which the brain orders visual input (Wertheimer 1923). This theorizing led the Gestalt psychologists to posit various ‘principles’ of perception. One prominent Gestalt principle is that of *figure and ground*. It is apparent that visual perception is structured by underlying and unconscious cognitive processes. Figure and ground organization is one such process. This is the process whereby attention can be selectively focused on an object (the figure) allowing it to exist in the perceptual foreground and stand out, or appear separated from, its surroundings (the ground). Although first elucidated in the context of visual perception, linguists have found similar principles to be evident in *construal processes* which appear to structure language in an analogous way (Talmy 1975). From visual organization to grammatical organization, the same

construal principles seem to be at work. For instance, we can see clear cases of figure and ground construal in the passive and active voices. Take the following examples:

- a. ‘Jones cleaned the windows’
- b. ‘The windows were cleaned by Jones’

Both represent the same state of affairs but there is a difference at a psychological level which can be explained by reference to a figure and ground construal process. In *a* ‘Jones’ is the figure, or the focus of attention, whereas in *b* ‘the windows’ become the figure against the background of Jones cleaning, and this represents a change of focus. The figure is the more salient part of the scene being described and so the grammatical form chosen by the speaker is indicative of the way they perceive the situation. We can also recognize the figure and ground construal at work in the case of annoyingness mentioned above. Take our now familiar phrases:

- c. Smith is annoying.
- d. I am annoyed with Smith.

In *c* the figure is ‘Smith’—‘Smith’ appears as the focus of the utterance. In *b* we see the opposite arrangement where ‘I’, the speaker, is profiled as the figure and ‘Smith’ as the ground. When analyzing the difference between the two phrasings in terms of figure and ground construal, we see how the focus of each utterance is shifted, but also note that the attribution error disappears from *c* to *d*. The fact that *c* – the phrase which makes the attribution error – is a common phrasing is telling and provides us with further reason for optimism. It is possible to show that phrases of the form in *c* are more frequently found in English than the paraphrase *d* with corpus data. Several English web-based corpora show that the phrase ‘am annoyed’ appears far less frequently than the phrase ‘is annoying’ in a range of contexts. For instance, the I-Web corpus showed ‘am annoyed’ at a frequency of 572 as compared with ‘is annoying’ at 11,676. And, in the Corpus of Contemporary American English ‘am annoyed’ appears 52 times versus ‘is annoying’ which appears 390 times (Davies 2008, 2018). Such results are indeed to be expected if my present analysis is correct. The fact that the form used in *c* appears more frequently is significant for two important reasons. Firstly, it vindicates our claim that attribution error occurs and is widespread. Secondly, it gives us further insight into the way in which humans generally perceive the world, as per the Cognitive linguistic framework. In uttering *c*, the speaker places ‘Smith’ as the figure—the more salient concept. I claim that this happens naturally as ‘Smith’ is perceived as the stimulus which gives rise to my annoyed state—recall how emotions are reactions to stimuli. So, although Smith is not intrinsically annoying—as the surface grammar suggests—Smith *did* cause me to be annoyed. Smith is therefore the stimulus and, hence, appears as more cognitively salient. This salience is reflected in the surface grammar of the utterance by construing the stimulus (in this case ‘Smith’) as the figure of the utterance. This fix of attention is shown in the common phrasing ‘*x* is *P*’ and gives a plausible explanation as to why we naturally and intuitively utter sentences of this form more frequently than alternatives. From an anthropological

perspective, it is possible that humans could have evolved a propensity to focus on and attend to the triggers of emotional reactions before introspecting. The foregrounding of the emotional stimulus provides possible support for this assertion and gives us an explanation as to why we more commonly utter phrases of the form ‘Smith is annoying’, although it is in error. It is easy to see how a propensity to exhibit a cognitive bias of this sort would have been a potentially effective survival tool.

Cognitive construal operations such as figure and ground structure human thought and language. Sentence grammar is therefore indicative of the way we conceptualize states of affairs. The same analysis from above can be run on moral utterances. In moral utterances of the form ‘ $\phi$  is wrong/right’,  $\phi$  is foregrounded as the more salient figure of the utterance. This suggests that it is the focus of the speaker’s attention and, therefore, is seen as an emotional stimulus. Hence, the figure and ground construal gives us reason to believe that  $\phi$ -like moral acts are identified (possibly subconsciously) as emotional stimuli and are, therefore, deserving of our attention. When viewed in such a light, attribution error is not an anomaly that needs to be explained away, but rather an expected outcome grounded in underlying human conceptualization processes.

## 7 Emotional Import

As defined here, moral properties have the propensity to elicit certain moral emotions in the agent and the observer. This makes them a special case of emotion-dispositional properties as they elicit emotions from multiple perspectives. Hence, we need to explain, now, the differences between emotion-dispositional properties and moral emotion-dispositional properties (rightness/wrongness).

Recall that with emotion-dispositional properties, there is an available paraphrase which represents an alternative construal and that has the upshot of not committing the attribution error as seen below:

- e. Smith is annoying. (Common phrasing with attribution error)
- f. I am annoyed with Smith. (Less common paraphrase without attribution error)

In the case of moral judgements, there seems not to be a similar paraphrase available:

- g.  $\phi$  is wrong
- h. #I feel wrong about  $\phi$

Our analogy seems as if it might break down here. However, lack of a paraphrase is to be expected if our current definition of moral emotion-dispositional properties is correct. Standard emotion-dispositional predicates contain only one emotional perspective and so can be used in the first person, as seen in *f* above. Remember, however, that moral concepts such as ‘rightness’ and ‘wrongness’ include, as part of their definitions, more than one emotional perspective and so cannot be used in the first person (*h.*). Hence, judgements of the form ‘*x* is wrong’ can be used easily as

‘blanket’ or universal judgements as they allude to the potential emotional responses from divergent perspectives (observer (agent-focused), observer (victim-focused) and agent (self-focused)). In contrast ‘Smith is annoying’ describes one single emotion-dispositional property, from the one perspective of the observer, and so can be felicitously construed in the first person as a subjective judgement about one’s own emotional state. It would seem, then, that given this dynamic, and given our definition of moral emotion-dispositional properties, that a first-person error-free construal of moral utterances might not be possible. However, moral utterances *can* be construed in the first person, thereby avoiding the attribution error, but this construal will be limited to representing only one perspective, and each perspective will make clear the emotional import contained within moral judgements. The judgement, ‘ $\phi$  is wrong’ can be reconstrued in a number of ways, each way representing a unique emotional perspective:

- i. I am disgusted with  $\phi$  (Observer’s perspective—agent-focused)
- j. I feel sorry for  $p$  because of  $\phi$  (Observer’s perspective—victim-focused)
- k. I feel guilty about  $\phi$ , (Agent’s perspective—self-focused)

We can see from *i*, *j* and *k* that moral judgements containing moral emotion-dispositional properties can be reduced to emotion-dispositional properties, as seen in the above examples, and such utterances do not commit the attribution error or use moral predicates. The fact that this is possible gives further support for the thesis that moral properties are reducible to emotion-dispositional properties. We can even see this more explicitly in common parlance where moral opinions are often couched in explicitly emotional terms; we can, and do, talk about moral acts without using moral predicates. For example, it is often said of immoral actions that that they are heinous, deplorable, disgusting, contemptible, loathsome, hateful or detestable.<sup>12</sup> The phrase ‘ $\phi$  is deplorable’, for example, has a clearly moral tone, despite the use of an emotional predicate, as opposed to a moral one. It is easily inferred pragmatically that this expresses a moral judgement. When analyzing these examples, we find an even tighter analogy with emotion-dispositional properties where a non-moral emotional predicate is being used to convey moral indignation. Hence, in such cases, we have no need to draw a mere *analogy* between moral properties and emotion-dispositional properties, as the former *is reducible to* the latter. In using adjectives like deplorable, heinous or detestable to describe morally wrong acts, we clearly paint morality with an emotional pallet.

<sup>12</sup> Such terms also appear to be used more frequently in sentences of the form ‘ $x$  is  $P$ ’ which commit the attribution error. For instance, in the I-Web online corpus, ‘is contemptible’ appeared 180 times vs ‘feel contempt’ at 115; ‘is disgusting’ 4832 vs ‘am disgusted’ at 1124 and ‘is deplorable’ appeared 801 times as compared with ‘I deplore’ at only 410. The relative frequency of such forms shows the stimulus as psychologically salient.



## 8 In Conclusion

Above, I have argued for a strong emotionist reading of moral properties, concepts and judgements. I, firstly, presented psychological evidence which shows a clear connection between moral judgement making and emotion, implying that a strong link between emotion and morality is evident. Secondly, I claimed that such evidence is best explained by understanding moral properties, and therefore moral concepts and judgments, as emotionally constituted. I, then, offered an analogical argument of non-moral emotion-dispositional properties drawing clear parallels between these two kinds of properties in showing how they function in language and discourse. Moral emotion-dispositional properties (rightness/wrongness) are a subset of the wider set of emotion-dispositional properties which contain properties such as ‘annoyingness’. Accordingly, moral judgements of the form ‘ $\phi$  is wrong/right’ are defined as:

those judgements referring to any conduct  $\phi$  perceived as having the propensity to elicit moral emotional responses in themselves and observers, in certain situations. Such that, a judgement  $\alpha$  is a moral judgement *iff* the conduct referred to by  $\alpha$  is perceived as having the propensity to elicit moral emotional states in agent  $x$  and observer  $y$  in  $S$ .

I claim that moral judgements, like emotion-dispositional judgements, are best understood as sincerely intended truth-apt propositions which express belief-like states. Thus understood, their referents are certain forms of conduct perceived as having a moral ‘character’ having the propensity to elicit moral emotional responses in agents and observers. Judgements, therefore, are relative to a particular moral community and have as part of their make-up subjective emotional import. Such judgements, however, commit an attribution error. They admit of a perceptual mistake—the mistake of attributing a response-dependent or emotion-dispositional property as an objective mind-independent one.

Finally, I claimed that our natural propensity to focus attention on emotional stimuli possibly leads humans to make such errors; to wrongly attribute such dispositional properties to, emotional stimuli. In support of this claim I offered an analysis of language from the perspective of Cognitive Linguistics. Analyzing construal operations apparent in moral utterances shows that the most common grammatical form when using emotion-dispositional predicates,  $x$  is  $P$ , foregrounds  $x$ —the emotional stimulus. This suggests that the emotional stimulus is more psychologically salient. This salience is conceptually significant and is apparent in the grammar of the utterance through the figure and ground construal process. Moral and emotional judgements of the form ‘ $x$  is  $P$ ’ are relatively frequent in comparison to passive grammatical forms which represent the emotion (as opposed to the stimulus) as the salient ‘figure’ of the utterance, and which do not commit the attribution error. This suggests a natural predisposition to focus attention on the stimulus, labelling it with an emotional tag, rather than to immediately focus on one’s emotional state.

When viewed this way, our utterances, although mistaken, can be seen as reflective of the way we perceive the world. Underlying conceptual structures augment and determine linguistic structures and do so with the aim of communicating socially significant and useful concepts and not merely at expressing true propositions and, thus, in this process we find an expected error.

**Acknowledgements** I am grateful for the comments made by anonymous reviewers which have helped to improve the arguments presented here.

## References

- Anderson SR, Lightfoot DW (2002) *The language organ*. Cambridge University Press, Cambridge.
- Antisocial personality disorder – APA Dictionary of Psychology. (n.d.). Retrieved September 19, 2020, from <https://dictionary.apa.org/antisocial-personality-disorder>
- Blair RJR, Cipolotti L (2000) Impaired social response reversal. A case of “acquired sociopathy.” *Brain* 123(6):1122–1141. <https://doi.org/10.1093/brain/123.6.1122>
- Blair RJR, Hwang S, White SF, Meffert H (2016) Emotional learning, psychopathy, and norm development. In: *Moral brains*. Oxford University Press, Oxford, pp 185–202. <https://doi.org/10.1093/acprof:oso/9780199357666.003.0009>
- Croft W, Cruse DA (2004) *Cognitive linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>
- Croft W, Wood E (2000) Construal operations in linguistics and artificial intelligence. In: Albertazzi L (ed) *Meaning and cognition: amultidisciplinary approach*. John Benjamins, Amsterdam, pp 51–78
- D’Arms J, Jacobson D (2007). Sensibility Theory Projectivism. <https://doi.org/10.1093/OXFORDHB/9780195325911.003.0008>
- Darwall S, Gibbard A, Railton P (1992) Toward Fin de siecle ethics: some trends. *Philos Rev* 101(1):115. <https://doi.org/10.2307/2185045>
- Davies M (2008) The corpus of contemporary American English (COCA): million words, 1990–present. <https://www.english-corpora.org/coca/>
- Davies M (2018) The 14 Billion word iWeb Corpus. <https://www.englishcorpora.org/iWeb/>
- Dolan MC, Anderson IM (2003) The relationship between serotonergic function and the psychopathy checklist: screening version. *J Psychopharmacol* 17(2):216–222. <https://doi.org/10.1177/0269881103017002011>
- Emotion – APA Dictionary of Psychology. (n.d.). Retrieved September 19, 2020, from <https://dictionary.apa.org/emotion>
- Gibbard A (1990) *Wise choices, apt feelings: a theory of normative judgment*. Harvard Universit Press, Harvard.
- Gibbard A (2006) Moral feelings and moral concepts. In: Shafer-Landau R (ed) *Oxford studies in metaethics*, vol 1. Oxford University Press, Oxford.
- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537):2105–2108. <https://doi.org/10.1126/science.1062872>
- Harman G (2009) Guilt-free Morality. *Oxford Stud Metaethics* 4:203–214
- Isen AM, Levin PF (1972) Effect of feeling good on helping: cookies and kindness. *J Pers Soc Psychol* 21(3):384–388. <https://doi.org/10.1037/h0032317>
- Langacker R (1987) *Foundations of cognitive grammar*, vol. 1: theoretical prerequisites. Stanford University Press, Palo Alto.
- Liao SM (2016) *Moral brains: the neuroscience of morality*—Google books. Oxford University Press, Oxford.
- Lyons J (1977) *Semantics*. In: *Semantics*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/cbo9780511620614>
- Mackie JL (1977) *Ethics: inventing right and wrong*. Penguin Books, London.
- McDowell J (1985) Values and secondary qualities. In: Honderich T (ed) *Morality and objectivity*. Routledge & Kegan Paul, London.

- Moll J, De Oliveira-Souza R, Bramati IE, Grafman J (2001) Functional networks in emotional moral and nonmoral social judgments. *Neuroimage*. <https://doi.org/10.1006/nimg.2002.1118>
- Moll J, De Oliveira-Souza R, Eslinger PJ, Bramati IE, Mourão-Miranda J, Andreiuolo PA, Pessoa L (2002) The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci* 22(7):2730–2736. <https://doi.org/10.1523/jneurosci.22-07-02730.2002>
- Prinz JJ (2007) *The emotional construction of morals*. Oxford University Press, Oxford.
- Prinz J, Nichols S (2010) Moral emotions. In: Doris JM (ed) *The moral psychology handbook*. Oxford University Press, Oxford, pp 111–146.
- Prinz J (2006) The emotional basis of moral judgments. *Philos Explorations* 9(1):29–43. <https://doi.org/10.1080/13869790500492466>
- Railton P (2003) Red, Bitter, Good (1998). In: *Facts, values, and norms: essays toward a morality of consequence* (Cambridge Studies in Philosophy, pp. 131–148). Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511613982.006>
- Schmahmann JD, Caplan D (2006) Cognition, emotion and the cerebellum. In: *Brain: a journal of neurology* (vol 129, Issue Pt 2, pp. 290–292). Oxford Academic, Oxford. <https://doi.org/10.1093/brain/awh729>
- Schnall S, Haidt J, Clore GL, Jordan AH (2008) Disgust as embodied moral judgment. *Pers Soc Psychol Bull* 34(8):1096–1109. <https://doi.org/10.1177/0146167208317771>
- Soderstrom H, Blennow K, Manhem A, Forsman A (2001) CSF studies in violent offenders: I. 5-HIAA as a negative and HVA as a positive predictor of psychopathy. *J Neural Trans* 108(7):869–878. <https://doi.org/10.1007/s007020170036>
- Soderstrom H, Blennow K, Sjodin AK, Forsman A (2003) New evidence for an association between the CSF HVA:5-HIAA ratio and psychopathic traits. *J Neurol Neurosurg Psychiatry* 74(7):918–921. <https://doi.org/10.1136/jnnp.74.7.918>
- Talmy L (1975) Figure and Ground in Complex Sentences. *Ann Meeting Berkeley Linguistics Soc* 1:419–430. <https://doi.org/10.3765/bls.v1i0.2322>
- Talmy L (1978) The relation of grammar to cognition: a synopsis. In *proceedings of the 1978 workshop on theoretical issues in natural language processing (TINLAP '78)*. Association for computational linguistics, USA, 14–24. <https://doi.org/10.3115/980262.980266>
- Van Roojen M (2015) *Metaethics: A contemporary introduction*. a Contemporary Introduction. In: *Metaethics*. Taylor and Francis Inc, Milton Park
- Wertheimer M (1923) [1950] *Laws of organization in perceptual forms* (abridged and translated by Willis D. Ellis). In: Willis D. Ellis (ed) *A source book of Gestalt psychology*. Humanities Press, New York, pp 71–88.
- Wheatley T, Haidt J (2005) Hypnotic disgust makes moral judgments more severe. *Psychol Sci* 16(10):780–784

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.