MORAL ILLUSIONS

Maria Waggoner

Forthcoming in *Philosophical Psychology* (please cite published version if possible)

Abstract: When one fails to have an accurate moral percept, a defect in character is often invoked. Little attention has been given to a certain kind of inaccurate moral percept – namely moral illusions – and what this would mean for virtue. This paper aims to show that moral illusions may in fact be a manifestation of virtue. By examining how an optimal perceptual system can give rise to perceptual illusions, a similar argument is made when it comes to moral perception: an optimal moral perceptual system, or that which a virtuous person would have, plausibly operates by using stored probabilistic information and can likewise give rise to moral illusions. Lastly, in light of this, I consider how we are to proceed. First, I argue that we should rethink our conception of the ideal virtuous person. Second, I suggest alternative moral resources which a virtuous person could use to navigate cases of moral illusions.

I. Introduction

Many parents describe sleep training methods that involve some amount of 'crying it out' as cruel, neglectful, and horrifying. One Reddit poster recounts that they "tried it out because of peer pressure on my first born and I have never felt more ashamed of myself… the sounds haunted me for months" (Zeusjordie, 2024) and another: "It seems so harsh and cruel to not comfort your baby, and the thought of doing so…makes my stomach turn" (Anonymous, 2022). Plausibly, even if the data was overwhelmingly in favor of its effectiveness in improving sleep quality and posed no risk of psychological harm – even if it was all things considered *good* for the infant – it's likely that many parents would not perceive it as so. For even for those who have decided that using the cry-it-out (CIO) method was the right thing to do, they report it often runs counter to

their intuitions and parental instincts.[1] The reasoning in this post might help explain why this is so:

> If you did CIO with any other group it would be seen as cruel…If you left your disabled or elderly relative to just cry locked in their room you would be reported to social services...If you left your dog to just wail in distress all night you'd be reported for animal neglect...[so w]hy is it ok to leave a baby to cry alone? (VegetableWorry1492, 2022)

Let us assume that we have a case where one has an inaccurate moral perceptual experience – while it seems that letting one's baby cry it out is cruel, neglectful, or otherwise morally wrong, it actually is not. In many such cases, one might misperceive *precisely because* of one's virtue.[2] It is because one is broadly attuned towards responding to and alleviating another's suffering that one misperceives in this particular case. Call this kind of inaccurate moral perceptual experience a *moral illusion.* In this paper, I'll give an account of moral illusions, or misperceptions that are manifestations of the person's virtue.

[1] This parent, for instance, details that "every time [my baby] awakens and cries, I have to suppress and ignore all of my maternal instincts that tell me to go comfort my child" (Myrighthandwoman, 2019).

[2] Unlike sensory illusions, whether a case is indeed that of a moral illusion or an accurate moral perception will often be debatable, since there is no way to verify whether one is having an inaccurate moral perception or not. If the reader is resistant in granting that the case given here is plausibly a moral illusion, I encourage them to consider further cases I give below in Section *III.D.*

While much ink has been spilled on the possibility and nature of moral perception,[3] little has been said about the phenomenon of moral illusions and how they relate to virtue and vice. While it is plausible that a virtuous person can have inaccurate moral perceptual experiences from time to time, the virtue ethics tradition seems to rule out that one's virtue can *manifest* itself in inaccurate moral perceptual experiences. Failures can occur, for "virtue allows for fallibility…virtue is consistent with some degree of 'unavoidable' err" (Audi, 1995, p. 456). But the "manifestation of virtue cannot be wrong" (Jordan, 2013, p. 267). This is because "virtues are supposed to always yield [what is] correct" (Bloomfield, 2020, p. 40). When it comes to action, "moral virtue itself is never the cause of wrongdoing" (Drefcinski, 1996 p. 143). Drawing from Aristotle, Drefcinski goes on to further explain: "[B]ecause a moral virtue does not produce its contrary action (NE, 1129aII-17, 1140b21-24), in order for virtuous people to commit wrongdoing we must say that they are not utilizing their moral virtue but acting from some other capacity" (ibid, p. 145).[4] Hursthouse (2006) explains that the virtuous person "gets things right" (p. 103): The virtuous person not only acts rightly but also morally perceives accurately.[5]

---

[3] Audi, 2013; Blum, 1994; Cowan, 2015; Cullison, 2010; Goldie, 2007; Jacobson, 2005; McBrayer, 2010; McDowell, 1998; McGrath, 2018; Roberts, 2003, 2013; Werner, 2016, 2020. The notion of a 'moral illusion' has been largely neglected in this literature (although see Moss, 2009, and Tappolet, 2016).

[4] Jordan (2013, p. 253) refers to this view as The Normative Invariance of Virtues of Character.

[5] Given that there are cases of *accurate* moral perception as well as *inaccurate* moral perceptual experiences, this means that there are success conditions for moral perception. This paper assumes that morality is objective, in at least some respects. One has an accurate moral perception when it accurately reflects moral

So, the virtuous person has reliable, yet not infallible moral perceptual faculties. But, according to the virtue ethics tradition, when she does misperceive, it is not a manifestation of her virtue, but some other capacity. It is this precise claim that will be questioned in this paper.

Just as we experience illusions in sensory perception, we should wonder – assuming there is such a thing as moral perception – if there are also moral illusions and whether we should expect the virtuous person to have them. In order illustrate this, I first consider how sensory perception works: according to a Bayesian computationalist framework, illusions are manifestations of a well-functioning and finely tuned sensory system.[6] So, I suggest, one also might experience moral illusions because of their well-functioning and finely tuned moral perceptual system. And – for the virtuous – their moral perceptual system is finely tuned to moral reality, having engaged in virtuous moral learning. For the virtuous person, moral illusions might very well be manifestations of their virtue.

I proceed as follows: In *Section II*, I give a Bayesian, computationalist account for sensory perception and illusions, which will provide the framework *Section III*, where I use this Bayesian story to explain how moral perception might work, if it indeed does

---

reality; one has inaccurate moral perceptual experiences (including moral illusions) when there is some sort of mismatch between morality reality and the perceptual content.

[6] For instance, Warren (2005, p. 350) explains that sensory illusions "are manifestations of the proper function of a perceptual system that is tuned to information for environmental surfaces, edges, and objects." Likewise, Ernst (2010, R357) posits that "illusions explained by this [Bayesian statistical] framework may therefore be considered manifestations of optimal perceptual performance."

exist.[7] Given this account, I explain we should also expect moral illusions to occur. In *Section IV*, I explore how virtuous and vicious persons might undergo moral perceptual learning, such that the former would have accurate moral perceptions and the latter inaccurate experiences. I also examine how one's own agency would be involved in forming the kind of moral perceptual system reflective of one's character. In *Section V*, I close by considering the implications of my proposal for our conception of the ideal virtuous person as well as how a virtuous person who is under a moral illusion should proceed.

II. A Bayesian Explanation for Sensory Illusions

To help illustrate how moral illusions might occur – and so how they can be manifestations of virtue – I will first take a brief detour into a common explanation for the occurrence of sensory illusions. In this section, I will give a Bayesian account of sensory illusions, which show how illusions arise from an optimal sensory system. In the next section, I'll then apply this account to moral perception.

Consider the dots below in Figure 1a: The reader will likely see some circles as convex and others as concave. But if we flip this image 180 degrees, as depicted in Figure 1b, the circles that were perceived as convex are now seen as concave, and vice versa. This image is actually consistent with any of the circles being seen as either convex or concave.

[7] To clarify, I do not intend to argue *for* the existence of moral perception. Rather, I suggest that *if* moral perception is a real phenomenon – and is in some way analogous to sensory perception, as some have suggested – *then* we should expect it to operate in a similar manner, thus giving rise to illusions.

FIGURE 1

One reason why you view a particular circle as convex (or concave) is that your visual system makes assumptions about where the light source is, and so from what direction shadows are being cast. It is thought that we have a light-from-above prior, which results from our visual system picking up on the fact that light generally does shine down from above (Adams et al., 2004).

This example of seeing circles as convex or concave illustrates how our perceptual system makes use of stored information that is learned over past experiences. According to a computationalist Bayesian framework, there are two basic components of perception: Incoming sensory information and assumptions about that incoming information.[8] Oftentimes, incoming sensory information is vague, uncertain, or ambiguous, and so our perceptual system uses stored information gathered from past experiences to form an inference about what is the most likely percept. The combination of these two components results in the percepts that we see, like convex or concave circles.

[8] The reader may wonder why a Bayesian computationalist framework of perception is being privileged over alternatives. One reason is that alternatives – such as direct realist accounts of perception, including the Gibsonian ecological approach – face difficulties when accounting for genuine perceptual illusions (Smith, 2002, 2010; Millar, 2015; McLaughlin, 2010). Gregory (1997) explains that "[t]o maintain that perception is direct, without need of inference or knowledge, Gibson generally denied the phenomena of illusion" (p. 1122). While Gibson did discuss illusions, his explanation appealed to 'inadequate' information (Gibson, 1966, p. 288) and that we can rid ourselves of such illusions by employing a "very special kind of selective attention" (ibid, p. 313). Since I wish to give an account of illusions – including moral illusions – I need to employ a model of perception that can adequately account for such phenomena, and this is why I make use of a Bayesian computationalist framework in this paper.

But oftentimes, the incoming stimuli consists of multiple features that are bound together, resulting in rich content. In such cases, stored information is used to make predictions about how the individual features relate to each other. During perception, we use an "internal representation that includes single-element and co-occurrence statistics, as well as information about the *predictivity between elements*" (Avarguès-Weber et al. 2020, p. 25923, italics mine). As Fiser (2011) explains, our perceptual system forms the "most likely…grouping of its previous experience into independent representational units" (p. 141). Our perceptual system learns the statistics of our environment to make predictions about what elements tend to hang together, resulting in the rich perceptual content that we experience.

Learning from past experiences, storing information, and making use of it later on is thought to be optimal, meaning that it is the best possible strategy for increasing reliability or maximizing accuracy over the long run. Shams & Kim (2010) explain that "In carrying out basic perceptual tasks, the human perceptual system performs causal inference and multisensory integration, and it does so in a fashion highly consistent with the Bayesian observer. This strategy is statistically optimal as it leads to minimizing…error" (p. 280). In a similar vein, Seriès and Seitz (2013) explain that when one's priors reflect one's past experiences with their world, they are said to be optimal. Seriès and Seitz review various studies, showing how these priors can be updated across

a lifetime, matching the statistics of one's environment and maximizing accuracy in the long run.

Optimality can be defined as follows: A perceptual system is optimal when it maximizes reliability or minimizes error in the long run. This will require effectively learning the statistics of one's environment, so that the percepts approximate the physical world around them. In addition to a perceptual system operating optimally, the percepts that proceed from such a system are also said to be optimal: A percept is optimal insofar as it is produced via an optimal perceptual system.

While an optimal perceptual system minimizes error in the long run, this does not necessarily mean it is infallible. Rather, it is precisely because it learns environmental statistics and makes use of this information that a certain kind of inaccurate perceptual experience – namely perceptual illusion – arises. An example of this occurs in cases of motion illusions: We live in a world where objects are usually static or move very slowly. As a result, we acquire a slow-motion prior. When incoming sensory information is vague or uncertain, we rely more heavily on our priors, meaning in low-contrast visual environments, more weight will be given to the slow-motion prior. This maximizes reliability in the long run, since we live in a world where objects tend to move slow. Nonetheless, this can result in an illusion – such as the Stepping Foot Illusion, depicted in Figure 2. Figure 2a shows that when the contrast is low – such as a when yellow block moves over a white line or a blue block moves over a black line – the blocks are perceived

as moving more slowly than they actually are, since one's slow-motion prior will be more heavily relied on. And when the contrast is high, our perception is less vague and fuzzy, so our slow-motion priors will be weighted less. Thus, the blocks are perceived as moving faster. Given that blocks move across an alternating background that produces high and low contrast visual experiences, the perceiver will see the blocks as moving in a stepping-like fashion. However, as seen in Figure 2b, when the contrast differences are reduced by turning the background to grey, the illusion disappears – the blocks now move in a smooth fashion.

FIGURE 2

Weiss and colleagues (2002) explain that "many motion 'illusions' are not the result of sloppy computation by various components in the visual system, but rather a result of a coherent computational strategy that is optimal under reasonable assumptions" (p. 603). Stocker and Simoncelli (2006) give a similar explanation:

> [For an] observer who lives in a world in which slower motions are more likely to occur than faster ones and whose judgments are based on noisy measurements…the perceived speed and direction of a moving visual stimulus depends significantly on attributes other than its physical motion…[but such] behavior can be seen as optimal (p. 578).

Although the outcome might be an occasional illusion, this occurs precisely because the perceptual system is optimal, having learned and made use of probabilistic information that match environmental regularities, maximizing accuracy in the long run.

This section summarized the Bayesian computationalist account of perception, and how an optimal perceptual system – precisely because of its optimality – can give rise to occasional illusions. In what follows, I'll apply this Bayesian computational account to moral perception and flesh out the implications it has for virtue and vice.

III. Moral Perception and Moral Illusions

This section fleshes out three possible interpretations of moral perception. The first interpretation is one where moral perceptual experiences are actually perceptual; the second and third are ones that describe moral perceptual experiences as perception-like intuition and perception-like emotions or affective states, respectively. For all three interpretations, I argue that we should expect them to operate in an optimal manner, producing optimal moral percepts. But if this is so, then we should also expect moral illusions to arise.

Before beginning, I should clarify that I am not giving these arguments to imply that moral perception, as a phenomenon, exists. Rather, I am giving a conditional argument: *If* moral perception exists, and plausibly works in the way suggested, *then* we should expect it to operate in an optimal manner and so result in moral illusions.

III.A. *Moral perception as perception – The argument from rich content*

One argument given for moral perception begins with the observation that we seem to be able to perceive complex content – like dogs, tables, and trees. As mentioned above, we do this by making use of stored statistical information about the relations

between the components, binding them together into complex units. Our perceptual system learns statistical information about what is the most probable complex percept, given the individual elements. Priors are made use of, resulting in the rich content that we end up perceiving.

Some have argued that we might expect moral perception to work in a similar way: If we can bind simple content into more complex content so to perceive morally relevant perceptual stimuli – like perceiving another's pain, perceiving intentionality in action, or perceiving another's emotions – then why couldn't that content get bound together further to create a more complex perception that has moral content – such as wrongness, badness, cruelness, or kindness. Audi (2013) compares perceiving moral properties to perceiving high-level properties in aesthetics (p. 35-37). Cowan (2015) suggests that "if perceivers can represent, e.g. natural kinds, in experiences, then it is perhaps less incredible that agents can have experiences of wrongness" (p. 668). Similarly, McGrath (2018) argues that if we can perceive morally relevant content, like another's pain or intentional action, then why couldn't those morally relevant contents be further unified together, forming even more complex moral content, such as moral wrongness? (p. 165).

Furthermore, empirical evidence supports the claim that we do perceive things like another's pain or intentionality in action: Singer and colleagues (2004) found that when one watches another individual in a pain-inducing situation – such as when another's finger is being sliced by a knife – this activates brain regions in the observer

which are also activated when one experiences the painful stimuli for oneself. Insofar as we perceive and sense our own pain, so too it seems that we can perceive and sense another's pain.[9]

Additionally, there is evidence that we perceive intentionality in action. Scholl & Gao (2013) imaged subjects' brains under two conditions. In the Wolfpack condition, subjects observed moving arrows pointed towards a moving disc. In the Control condition, subjects observed moving arrows perpendicular to the moving disc. Subjects in the Wolfpack condition – a condition where subjects often report the arrows to be 'chasing' the disc – showed activation in brain regions typically involved in visual perception, which was over and above the activation observed in those in the Control condition. Given that the lower-level features, like the basic movements of the disc and arrows, were the same across both conditions, the researchers concluded that subjects in the Wolfpack condition were seeing something in addition – i.e. they were perceiving intentionality.[10]

Like McGrath (2018) suggests, we might think *pain* and *intentionality* to be the basic building blocks of at least some moral perceptual content. If that's right, then it isn't so strange to think that there might be something such as moral perception, and that it is a result of binding simpler, morally relevant, content together with the aid of stored

[9] This perception is sensory, although perhaps more somatosensory than visual.
[10] See also Carruthers (2015), who likewise argues that "our awareness of the mental states of other agents is often perceptual in character" (p. 498).

statistical information that's reflective of one's moral environment. But, if moral perception works in this way by employing priors based on the most probable grouping of simple elements, then we should likewise expect it to operate in an optimal manner, and so sometimes manifest in moral illusions. I will go on to further illustrate how moral illusions are optimal percepts, resulting from an optimal moral perceptual system below in *Section III.D*. But before that, I'd like to turn to a few other ways that moral 'perception' might operate

III.B. *Moral intuitions as moral 'perception'*

Perhaps we should not think of moral 'perception' as actual sensory perception but interpret 'perception' in a more metaphorical way, as something like moral intuition. Intuition has often been said to be akin to perception, insofar as both are impressions or presentations of the world (Bengson, 2015; Chudnoff, 2020). So, if moral 'perception' is actually just moral intuition, then we should look at what mechanisms are at work in moral intuition to see if it, too, is optimal, and so might also give rise to moral illusions.

There is good reason to think that moral intuition does operate in an optimal manner and so will also result in moral illusions: It's recently been suggested that intuitions are 'smart' and 'rational' – "the result of learning complex statistical relationships" (Railton 2017, p. 182), "guiding behavioral selection via the balancing of costs, benefits, and risk" (Railton, 2014, p. 833), in a way that "approximates Bayesian updating" (ibid, p. 835). Woodward and Allan (2007) suggest that intuitions, including

moral intuitions, are the result of not just the current environmental input, but involve "a complex repeated game of some kind" (p. 185) and are "based on experience-dependent probabilistic models" (p. 186). Nichols and colleagues (2016) propose that moral intuitions involve general moral rules, learned through tracking environmental regularities and statistical updating that is optimal, or approximate Bayesian learning (p. 549). Kleiman-Weiner and colleagues (2017) advance an account of moral learning where noisy observations, along with innate priors, are used to build more complex moral models that are later used in Bayesian inferencing. Innate priors are updated, and the stored information begins to match environmental regularities, given the observed behavior of others and feedback of one's own behaviors.

To illustrate how moral intuitions are the result of 'learning complex statistical relationships' that involve a 'repeated game of some kind', consider Railton's interpretation of Haidt's JULIE AND MARK case. Haidt gives subjects a vignette describing two siblings, Julie and Mark, who have incestual sex just once, taking birth control and experiencing no negative psychological consequences. In fact, it "makes them feel even closer to each other" (Haidt, 2001, p. 814). After reading the vignette, subjects are then asked if it was okay for Mark and Julie to make love. Subjects often say 'no' but have trouble providing reasons that are applicable to the case. They will often cite potential genetic defects to the offspring or emotional trauma of Julie and Mark. But these

consequences are ruled out by the stipulations of the case. When pressed, subjects are dumbfounded; Haidt concludes that our moral intuitions are not rational.

Railton (2014), however, argues that our moral intuitions are smart and rational, especially over the long run. Railton gives an analogous case to illustrate this, showing why the subjects' intuitions about the wrongness of Mark and Julie's actions aren't irrational. This case is of Jane and Matthew: Similar to Mark and Julie, Jane and Matthew are brother and sister and decide to engage in playing Russian Roulette on one particular occasion:

> As it happens, the gun does not go off, and neither suffers any lasting trauma from the experience. They both enjoyed the game but decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other (p. 849).

Railton thinks that it is obvious why it was wrong for Jane and Matthew to play Russian Roulette: "Jane and Matthew carelessly put all this at risk for the sake of a potentially 'interesting and fun' evening. [This is] not OK, despite the fortunate outcome" (ibid). Even though Jane and Matthew come out unscathed in this instance, if they were to keep playing Russian Roulette, it would not end well. In entering this social arrangement, both Jane and Matthew put each other at risk, and this risk they pose on one another is morally wrong.

Our moral intuitions are based on running the probabilities and registering the risks, over the long run. In this way, our moral intuitions involve a 'repeated game', based

on 'learning complex statistical relationships,' and so are optimal. However, insofar our moral intuitions are optimal in this way, we should also expect moral illusions to arise.

III.C. *Moral emotions as moral 'perception'*

A third way to understand the nature of moral perception is to turn to emotions and affective experiences. For instance, Christina Tappolet (2016) holds that emotions are evaluative perceptual experiences, whereby one has a perceptual experience of values or evaluative properties. Tappolet argues that emotions are analogous to sensory perception in all the relevant ways, and so should be thought of has a kind of perception too.[11] When one feels fear, one has a perceptual experience of the value of the fearsome; feeling shame or admiration are perceptual experiences of the shameful and the admirable, respectively. Roberts (2003, 2013) has also proposed that "emotions are a kind of perception" (2003, p. 87), meaning they are "impressions, [or] ways things appear to the subject." (ibid, p. 75). One way to further flesh out the basic idea that emotions are evaluative perceptual

[11] Tappolet's argument consists of pointing out that both sensory perceptions emotions (1) are automatic and involuntary, (2) are representational states, (3) have correctness conditions, (4) have a certain phenomenology, (5) need not be conceptual, and (6) are encapsulated from cognitive states (Tappolet, 2016, pp. 19-24). However, by Tappolet's own lights, it seems that intuitions would also qualify as perceptual experiences, and so perhaps we might think the line between moral intuitions and moral emotions is quite fuzzy. As Tappolet observes with emotions, many others have pointed out that intuitions also (more or less accurately) represent the world (Bedke, 2008, Bengson, 2015 and Chudnoff, 2020; the latter two call intuition "quasiperceptual", which Tappolet also is happy to refer to emotions as such, p. 30). Intuitions are likewise automatic and involuntary (Haidt, 2001; Topolinski & Reber, 2010), have a certain phenomenology (Audi, 2022; Clavian & Fitzgerald, 2017; Huemer, 2005), need not be conceptual (Gigerenzer, 2007; Railton 2014; Mamin, unpublished manuscript), and can also be encapsulated from other cognitive states (Lyons, 2018). It is not surprising that many suggest emotions to be a kind of intuition, or equate the two (Haidt, 2001, Roeser, 2012; Keltner et al., 2012). Thus, it may very well be the case that distinguishing emotions from intuitions, as Tappolet does, is unsupported. Nonetheless, I grant for the sake of argument that the two are separate, and worth exploring independently rather than collapsing one into the other.

experiences is to turn to a recent proposal given by Peter Carruthers (2023, 2018) which provides an evaluative account of affective valence. Emotions are a kind of affective state, and it's commonly thought that affective states have two components: Valence (positive or negative) and arousal, or intensity (Russell, 1980; Feldman-Barrett & Russell, 1999; Reisenzein, 1994; Rolls, 1999). According to the evaluative account of affective valence, "the two valences are analogue-magnitude representations of value (non-conceptual) goodness and badness, respectively" (Carruthers, 2023, p. 534). What this means is that in having a given emotion like fear or admiration, one experiences affective states with a certain positive or negative valence. And when accurate, this experience of positive or negative valence is a 1-to-1 representation of the goodness or badness 'out there' in the world.[12] Carruthers (2018) explains that positive and negative valences are "perception-like representation[s]" of seeming badness and seeming goodness (p. 6). Just as our sensory perceptual system mirrors – with varying accuracy – physical reality, our affective states mirror evaluative reality, or basic goodness and badness.

One important thing to note here is that emotions, and affective experiences more generally, oftentimes represent badness and goodness as it matters to the person. This may or may not involve an explicitly and consciously held concern, desire, or goal. For instance, Roberts' (2003, 2013) perceptualist account of emotions holds that emotions are

[12] Beck (2018) and Maley (2021) describe this as a 'mirroring' of the thing represented and the representation in the head

concern-based construals; when a parent feels anger at another who slighted his child, it is because this parent cares about his child, and that his child be treated with fairness and kindness. In this way, emotions and affective states are quite flexible, arising in relation to whatever cares and concerns one has. Cares and concerns might not always be so explicit though and can also consist of how one is doing relative to evolutionarily hard-wired fitness goals, such as feeling disgust towards contamination and disease with the implicit aim of staying healthy and surviving. Even though emotion and affective states might be a way of understanding moral perception, there is one obvious way that affective states seem to be unlike the other interpretations of moral perception discussed thus far: emotions and affective states might seem to be quite unreliable and far from optimal. Unlike sensory perceptions and intuitions which operate in an optimal manner, a quick personal reflection of our own emotional lives might have us thinking that our emotions often get it wrong: "more often than not, emotions misfire – we are afraid of a great number of innocuous things, and we are angry at a great many non-offensive people" (Tappolet, 2016, p. 42).

In response, a few things must be said. First, we might wonder how accurate our personal reflections of our emotional and affective lives are. It's quite plausible that the emotional episodes that are salient to us are the ones where emotions are especially strong and cases where we got it wrong. When our emotions are operating as they should and with reduced intensity, they might just move to the background; we are likely not

aware of all the successful instances. This seems especially true when including affective states more generally. Affective states not only constitute our conscious emotional states, but also involved in basic interoceptive states like hunger and thirst, intuitions, decision making, and even spatial navigation. This means when we are evaluating the reliability and optimality of our affective states, it's a much bigger landscape that we might have thought. It doesn't just include the sudden anger I feel when my child has been slighted, but also the twinge of anxiety when trying to find my way home at night, the confidence I feel when making a purchase, and the slight uncomfortability that makes me reach for a sip of water. When taken all together, it's actually quite dubious that 'more often than not, emotions misfire.'[13]

Secondly, recent empirical literature points to the fact that emotional valence seems to operate according to optimal Bayesian inferencing, making use of priors or of stored information and weighting this relative to the precision of incoming information (Hesp et al., 2021; Majumdar et al., 2023; McGovern, et al. 2021). Drawing from the observation that accurately estimating uncertainty is a crucial aspect of our optimal perception, Majumdar and colleagues (2023) go onto show that emotions similarly involve an accurate estimation of uncertainty, or the expected changes in our environment. In virtue of registering how volatile our environment has been in the past, our affective system incorporates statistical predictions about how volatile it'll be in the

[13] I am thankful to Evan Westra for bringing this response to my attention.

future, and what we might expect the evaluative landscape to be like. Our emotions and affective states involve registering evaluative change, given prior information of the change that was most likely to occur. They also note that if we optimally represent value, resulting in optimal action policies (as some have found or suggested: Behrens, et al. 2007; Nassar et al., 2010; Piray & Daw, 2020, 2021), then this requires optimal uncertainty representations as well (Majumdar et al., 2023, p. 3768). Likewise, Hesp and colleagues (2021) explain that emotional valence operates in a similar Bayesian way as it does in sensory perception, weighing probabilistic expectations against the precision of incoming sensory information. With emotions and affective states, though, the story is a bit more complicated, as it also includes anticipatory information about what sorts of events one expects to happen and how one plans to act or respond considering these anticipated events. Hesp and colleagues show how emotional valence representations can be optimized over time, through "Bayesian optimal updating…track[ing] changes… lend[ing] a sign to otherwise unsigned divergences between predictions and outcomes" (p. 399). When one's predictions or model of anticipated situations and expected or preferred actions in those situations end up greatly diverging from what actually happens, priors are updated. Information about uncertainty and volatility is registered through affectively experienced valence. The more uncertainty, the more negatively valanced affect experienced; the more confidence, the more positively affect experienced (ibid, p. 431- 2). Lastly, McGovern and colleagues (2022) have put forth an account of

anxiety which results from optimally learning uncertainty in one's environment. If one is in a very volatile and uncertain environment, "anxious priors form" (p. 7). Such statistical information about environmental uncertainty is learned, whereby "this occurs in a Bayesian optimal way" (ibid). McGovern and colleagues (2022) seem to actually describe something like an illusion taking place, as a result of this optimality: "Because [learning environmental uncertainty] occurs in a Bayesian optimal way, model precision increases, despite clearly not being adaptive under novel environmental conditions where certainty may now exist" (p. 7). In a sudden change of context, where more certainty exists, the anxious priors will still be made use of, giving rise to an illusory experience.[14]

This research focuses on the way our emotional system operates in an optimal manner when it comes to registering uncertainty, both in the environment and how we expect to act and respond to the environment. If this is right, then as I'll go on to argue, we might likewise expect moral illusions to arise.

---

[14] It should be noted that McGovern and colleagues are putting forth an account of clinical or pathological anxiety disorders. They use this model to further argue that clinical anxiety will also involve this learned "inherent uncertainty…persist[ing] across time" (p. 7), failing to eventually adapt or learn the new statistical information of one's environment. In clinical cases, we can think of the illusion persisting and new priors are not learned, even after being in an environment with greater levels of certainty for a long period of time.

Insofar as emotions consists of optimally detecting the volatility and uncertainty in one's environment, an interesting implication may arise: While detecting uncertainty in one's environment is a feature of all kinds of emotions (Majumdar et al., 2023), uncertainty may be more central to some emotions, like anxiety, than others. What this might mean is that moral illusions should be more commonplace in situations of high uncertainty, where a consistent and large degree of anxiety is appropriate. Insofar as uncertainty is a constitutive part of emotions, in general, we can expect moral illusions to arise. But they might be more commonplace in circumstances that call for certain emotions over others.

III.D: Moral Illusions as Optimal Percepts

Now I'll give two examples of moral illusions that we could expect to occur within a Bayesian computational model of moral perception. In other words, these moral illusions would be ones that arise from optimal moral perceptual systems, or from moral perceptual learning that maximizes reliability in the long run. Consider, first, a variant of Railton's case:

> JESSICA AND MICHAEL: Jessica and Michael are hiking together in the middle of Alaska. One evening while hiking, they come across a gun with ~7 million chambers but only one is loaded. They decide it would be interesting and fun if they tried playing Russian Roulette. As it happens, no bullet fires, and neither suffers any lasting trauma from the experience. They both enjoyed the game but decided not to do it again. They keep that night as a special secret, which makes them feel even closer to each other.[15]

In firing the gun at themselves, Jessica and Michael have approximately the same risk of dying as if they were to go on a joyride. If your intuition in this case is that it is still too risky for Jessica and Michael to fire the gun, then you are likely experiencing a moral illusion. But plausibly, this illusion is optimal. This illusion arises because the moral norms, rules, and statistics that we use to get around in our world are reliable and maximize accuracy in the long run. But, when we employ these learned norms, rules, and statistics in this hypothetical situation – which is quite unlike our everyday experiences – we are led astray, and an illusion arises.

Note the following facts about our own actual world:

[15] I am grateful to Allan Hazlett for suggesting these modifications to Railton's original case.

1) We live in a world where 7 million chamber guns do not yet exist. Our intuitions that have been formed around firing guns are not reflective of these kinds of guns.

2) We have been taught to never point a gun at someone, even if it is unloaded.

3) We have been taught to never fire a gun unless you want to fire a bullet.

Given the great risk of harm in firing a gun, acquiring these sorts of intuitions about gun shooting is plausibly optimal. But these intuitions may nonetheless also result in a moral illusion in the case of JESSICA AND MICHAEL.

Now consider a second case – one in which the reader doesn't experience a moral illusion, but a person in the vignette (Arnold) does.

> ARNOLD: Arnold and his friend, George, find a baby squirrel that fell out of a tree in a recent windstorm. The squirrel is alive but seems to have major neurological damage; both Arnold and George are confident that the squirrel will not survive. George suggests using his camping axe to put it out of its misery. Arnold sees this as cruel and instead feels that the compassionate thing to do is to make the squirrel as comfortable as possible as it dies on its own. George takes matters into his hands and kills the squirrel with a quick strike. Afterwards, Arnold feels that what George did was wrong.

This case illustrates one where a virtuous person's (e.g. Arnold's) moral perceptions and intuitions – despite perhaps being inaccurate in this particular case – might nonetheless result from an optimal moral perceptual system, or a moral perceptual system that gets things right in the long run.

Plausibly, we can see Arnold, and his well-tuned moral perceptual system, as virtuous. Over the years, Arnold has rightfully learned that intentionally killing innocent

creatures who pose no danger as simply cruel. And navigating his moral world has taught Arnold that intentional killing is worse than allowing death (Nichols et al., 2016), and that physical harm and bloodshed is what good people avoid. Perhaps for those like George, who may have had different experiences than Arnold, are not subject to such a moral illusion.[16] But, given Arnold's past experiences and virtuous moral learning, he experiences a moral illusion. Yet, this inaccurate moral perceptual experience is plausibly one that is a manifestation of his virtue.

One way to understand the occurrence of these moral illusions is to look at parallel cases of perceptual illusions. Consider, again, motion illusions. It was noted above that they arise due to (i) acquiring a slow-motion prior from living in a world where objects tend to move very slowly, and (ii) being in a context where the incoming information is vague and uncertain. In such contexts, we have an illusory experience of the stimuli moving more slowly than it actually is, for we rely on priors that were learned in a slow-moving world but are now being employed in a context where things move much faster. Additionally, the incoming information is vague and so priors are assigned more weight (If the incoming information was very precise, the priors would not be weighted as

[16] Suppose George grew up in a family of hunters, or experienced violence in his youth.

See Cheng et al., 2007, for this sort of evidence found within physicians and their decreased sensitivity in pain perception.

highly, since the incoming sensory evidence would be more trustworthy. In these cases, illusions would be less likely to arise).

So too, then, with moral perception: We can understand cases like JESSICA AND MICHAEL and ARNOLD as ones that fit this story. Both scenarios involve new and novel contexts, different from what we, and Arnold, have encountered before. Furthermore, the incoming evidence is vague, uncertain, and untrustworthy: it might be quite hard for us to grasp what a 1/7 million chance looks like; a gun with that many chambers would have to look quite different than any gun we have previously encountered. As for Arnold, the incoming evidence he is receiving is probably mixed, and it is far from clear what evidence to trust: on the one hand, the squirrel seems to be dying and in pain. And his friend, George, is providing him with some evidence that killing the squirrel with an axe is the right thing to do. On the other hand, Arnold has past experiences where the compassionate and morally right actions were providing comfort and companionship was called for (rather than taking an axe to another's throat), as when his grandmother was dying.

For Arnold, we can see that it's not that his moral perceptual faculties are completely context insensitive. But rather, when he enters into this new context, the incoming information he receives is vague, uncertain or untrustworthy. As a result, his moral perceptual system makes use of priors which were learned in previous contexts. Despite switching into this new context, the previously learned priors are highly

weighted and relied on in his moral perceptual processing, which gives rise to illusory experiences. This is because his moral perceptual system has not yet learned the statistics of his new environment, and so has not yet adjusted its priors to reflect it. If it's right these sorts of cases involve (1) novel and new contexts and (2) the incoming evidence is uncertain, vague, untrustworthy, and hard to make sense of, then we should expect that (3) a prior hypothesis, acquired and suitable for a different context, would be given much weight even in this new context. This could result in a moral illusion.

IV. Moral Illusions and Their Relationship to Virtue and Vice

Thus far, it's been argued that if moral perception is like other cases of sensory perception, we should expect an optimal moral perceptual system – the kind of moral perceptual system a virtuous person has – to occasionally manifest itself in moral illusions. While optimality is plausibly an excellence of the system, for a perceptual system that does not update in this way will be less reliable, it is not necessarily an excellence of a moral perceptual system that only a virtuous person would have. Rather, whether one is virtuous or vicious (or somewhere in between), one's moral perceptual system will likely learn from one's moral environment in this way, updating itself given the statistical information, and making use of this information in future occasions to arrive at the most probable percept.

In addition to having an optimal moral perceptional system, the virtuous person will also need to have the right virtue-conducive experiences to provide the virtue-

conducive statistical information to learn on. There might be some experiences, such as situations of violence or environments structured by racism and sexism, that when learning optimally from them, preclude acquiring reliable moral perceptual faculties or morally perceiving the world the way that the virtuous person would. Having the moral perceptual faculties of a virtuous person likely requires having certain previous experiences with particular environmental regularities (and not others).

At this point though, the reader might have a few worries: First, if our moral perceptual priors update based on the statistics of our moral environment and so influence future moral perceptions, does this mean that our environment completely determines what we morally perceive? In sensory perception, if one is placed in an environment where light shines up from below for a long enough time, one's priors begin to change to reflect this experience (Adams et al., 2004). This environment seems to determine the priors that one's visual system learns and makes use of in future perception. If moral perception is like this, then this would be problematic for a few reasons. First, moral perception would be largely sub-personal, operating not at a level of agency and character but at a level of mere computations. Second, if two people occupy the same environment for long enough, then we should expect the perceptual content of their moral perceptual systems to be more or less identical. But this seems dubious: two different people, with two very different characters could occupy the same environment for several years and continue to morally perceive differently. We expect our characters

and the moral perceptual systems they inform to persist through various situations and environments, rather than completely conform to them. We expect that a vicious and virtuous person could occupy the same environments and retain their virtues and vices, rather than converging to be identical in character. Call this issue the ***Sub-Personal Problem.***

In response, it must be stressed how our agency and character is often manifested in how we subjectively construe our objective situations. Social psychologists Ross and Nisbitt (1991) explain that "[t]here is significant variability in a given person's construal of events, enough to lead us…to expect that there will be nontrivial variation in behavior across two objectively almost identical situations" (p. 68) The formation of one's subjective construal involves a focusing of one's attention on certain features of the situation and not others. And, what one directs their attention to is often connected to one's character and agency. Empirical research show we can consciously choose to change our subjective construal of a situation by taking a 'fly on the wall' perspective (Kross & Ayduk, 2011) or a more abstract appraisal (Trope & Liberman, 2010). By choosing to shift one's subjective construal, one changes how the objective situation appears, having a cascading effect on one's emotions, decision making and actions.

Philosophers have picked up on this idea that people may differ in how they subjectively construe the same objective situation, carving out a place for our character to make a difference. Roberts (2013) holds that our subjective construal is often influenced

by our cares and concerns, or what we take to be important. Vigani (2018) explains that "one's conception of how one ought to live…can direct an individual's attention and therefore affect that individual's subjective construals of situations." (p. 240). Sherman (1991) also observes the relationship between agency, character, and subjective construal: "The agent will be responsible for how the situation appears as well as for omissions and distortions. Accordingly, much of the work of virtue will rest in knowing how to construe the case" (p. 29). Thus, a virtuous and vicious person may subjectively construe their objective situations quite differently. The environmental statistics that one's moral perceptual system is trained on might best be thought of as the statistics of the subjectively construed environment. And so, the regularities that one experiences and learns from will themselves be, in part, a result of one's character. Given this, one's subjectively construed environment might determine the priors that are learned, stored, and made use of in one's moral perceptual system, but this subjectively construed environment is influenced by one's character and agency.[17]

[17] One additional reason why the same environment might not result in identical perceptual content across various people is that such people could have differences in some of their innate priors. Scholl (2006) explains that some of the priors in our perceptual system are like a "default setting or 'factory setting' of the relevant variable. But in no sense is that principle then written in stone, since its value can later be updated and tuned via interaction with the environment" (p. 49). Insofar as people may differ with respect to the innate priors they start out with – since these innate priors might be a result from one's genetic code (Gregory, 2006) and individuals vary with respect to their genetics – such priors will be updated according to their environment, but their updated priors may still differ, since they began at different starting points (or with different innate priors). These differences in innate priors might be a way, at least in part, describing natural virtues, or the kinds of tendencies one is born with before undergoing intentional character formation.

Furthermore, even if were true that one's objective environment completely determined one's moral perceptual system, this need not exclude agency and character from the picture. This is because what objective situation one chooses to enter into may be where agency and character are exercised. Rodgers & Warmke (2015) argue that being a "good situation-chooser" (p. 18) "is a disposition to act for certain kinds of reason: a virtue" (p. 22). The life of the virtuous person will involve an awareness of how one's environment might passively shape one's moral perceptual faculties and so will choose oneself in virtue-conducive environments while avoiding vice-conducive ones.

In addition to the Sub-personal Problem, the reader might also be concerned about what I'll call the ***Moral Feedback Problem***: In what way can one be said to learn the moral statistical information of one's moral environment? In sensory perception, our perceptual system updates and learns based on feedback. When our priors fail us, we don't effectively get around the physical world: we misperceive depth and bump into tables and chairs; we fail to register shadows and fall into potholes. We get clear feedback, and our perceptual system updates on this. But there doesn't seem to be similar moral objects that we "bump into" when we get things wrong. Moral feedback isn't clear and obvious in the same way that physical feedback is. So, in what way can our moral perceptual system be said to learn the statistics of our moral environment?

While moral feedback is almost always more implicit than physical feedback, consisting of things like subtle bodily gestures, a particular emoji sent over text, or the

tone in a friend's voice, that does not mean it isn't noticed and updated on. What we morally "bump into" and what we update on are things like social norms, hurt feelings, or acknowledgements of gratitude. The moral information we receive might also come from reading novels, learning about historical events, or adopting certain religious frameworks. This information may change our perception of the goodness or badness (and the relative degrees of these properties) of certain actions or states of affairs (e.g. death, suffering, etc.). And, as mentioned above, what feedback our moral perceptual system updates is in part a result of how we subjectively construe our environments and feedback. Moral praise given from one's church pastor might positively reinforce the pious churchgoer's moral perceptual priors but have a completely different effect on the apathetic teenager who has no respect for their parent's religion.

Another way that moral feedback is different than physical feedback is that while physical feedback provides us with information about what is actually out there, it is plausible that one can update their moral perceptual systems even when the feedback does not necessarily track actual moral reality. Just as one can update on social norms, even when those norms do not track moral truths, one can likewise engage in moral updating, picking up on the moral statistics of their environment. Yet, these "moral statistics" should be understood as environmental regularities that put forth morally relevant features, even when this departs from actual moral reality. For example, one could pick up on the moral statistics of one's environment which puts forth false moral

information that people of a certain race lack human dignity. Updating based on this 'information' can still be done in an optimal manner, but this moral feedback will not necessarily lead to updates which make one's moral perceptual system better at navigating moral reality. Rather, such feedback will just make one better at minimizing 'errors', relative to their moral environment. This means that vicious, as well as virtuous, people engage in moral updating, picking up on the moral statistics of their (subjectively construed) environment, whereby this is done in an optimal manner.

We might now have some rough sketch of what a moral perceptual system (if it indeed exists) of the virtuous person – and perhaps the vicious person – looks like: Both the virtuous and vicious person would have a moral perceptual system that will update, given moral feedback. Their moral perceptual systems will be optimal, maximizing the reliability of moral percepts in the long run. However, what this 'reliability' is relative to will differ: For the vicious, the moral feedback they will receive will lead them to navigate their environment in a way that leads to minimal disruption or "bumps," but this environment and effective navigation does not actually match up with moral reality (e.g. white slaveowners in the antebellum south who continued to operate business as usual).[18] In contrast, the virtuous person will have a moral perceptual system updates and learns

[18] There is a further question as to how one can know whether they are virtuous or vicious, or whether their moral perceptions are accurate or not – how would the slaveowner in antebellum south know he is having inaccurate moral perceptual experiences? This question is complicated and cannot be answered here. But it's worth nothing that this is not a unique problem for moral illusions. Rather, it is a general worry that can arise across various moral epistemological approaches. Insofar as morality is not verifiable, one may often fall short of the confidence they want.

on moral feedback which actually reflects what is morally true, rather than merely one's (morally deprived) environment. The virtuous person will both select and construe their environment so that it is one which promotes and sustains truth-tracking moral perceptual faculties. However, whether one is virtuous or vicious, insofar as their moral perceptual systems are optimal and update on the statistics of their moral environment, we should expect occasional moral illusions to arise. This means that for the virtuous person, they will occasionally experience inaccurate moral perceptions precisely because of the excellences and virtues they have. Given their virtues and the (objective as well as subjectively construed) environments they place themselves in to, along with their optimal moral perceptual system which updates based on this statistical information of their environment, they will be much more likely to morally perceive what is actually true. However, when they find themselves in an environment that is atypical, or unlike that which their moral perceptual system has been trained on – such as Arnold – their priors may lead them astray, resulting in an illusion.

V. How Should a Virtuous Person Proceed?

If it is true that the virtuous person will at least sometimes morally perceive incorrectly because of their excellences and virtues, then this leaves us with a few pressing questions: 1) how should this impact the way we conceive of the virtuous person? 2) how should the virtuous person proceed when they are under a moral illusion?

If it is right that moral illusions can arise out of one's virtue, then this should prompt us to reconsider whether the virtuous person will always "get things right" (Hursthouse, 2006, p. 103) in their moral perceptions; rather, virtue can manifest itself in getting things wrong. However, this could be resisted by arguing that the virtuous person is an ideal, not any one of us human beings (Annas, 2004, p. 67). While it might be true that human beings – with all of their Bayesian computational quirks – are subject to moral illusions, no human being is fully virtuous anyways.

However, assuming that there is such a thing as moral perception and that moral illusions are expected to arise, this should not just influence the way we that we conceive of human persons en route to virtue. It will also have implications for our notion of the ideal virtuous person.

There are various senses of an 'ideal' which are relevant for thinking about and pursing virtue. For instance, an ideal could be understood as something that we appreciate as excellent or deem as good; call this the exemplary sense. But there is another sense which I am concerned with here: the emulatory sense, where the ideal virtuous person provides us with a goalpost to strive towards and a psychology which we seek to mold our own around. Insofar as our moral perceptual system is optimal and subject to moral illusions, it is this sense – the emulatory sense – of the ideal that we ought to adjust, for it is an *illusive* or self-undermining ideal.

An *illusive ideal* is one where progress within one facet of the ideal undermines progress made in another facet (Siversten, 2019). Consider the ideal of an all-around excellent runner, where this means being both an excellent endurance runner and excellent sprinter. Given the way our muscular system works, we cannot be both, for pursuit in one area undermines progress made in another. When one engages in sprinting-related training, her intermediate muscle fibers turn into fast twitch fibers; engaging in endurance training turns these fibers into slow twitch. One's progress within the realm of endurance running undermines progress within the realm of sprinting, and vice versa. Thus, the ideal of an all-around excellent runner – at least insofar as it is an emulatory ideal – is an illusive one, for when we try to approximate it, our efforts are self-undermining. Insofar as our moral perceptual systems are subject to moral illusions, trying to mold our psychologies to be like the ideal virtuous person – one who morally perceives accurately in all environments – is on par with attempting the physical training necessary to be an all-around excellent runner. Our efforts in both realms will often be self-undermining. Suppose that one slowly changes their moral perceptual system through learning of new priors, given a new moral environment. This can at least sometimes involve changing the moral priors they previously had, or losing progress made in other moral environments. In visual perception, if one is placed in an environment where light is made to shine from below, one will learn the statistics of this environment, and adopt a light-from-below prior. But, as a result, one's light-from-above

prior will be 'unlearned' or greatly weakened. When placed in a situation where light shines down from above, one's learned light-from-below priors will result in an inaccurate perception. In short, one cannot have both light-from-above and light-from below priors. So, to have accurate perception in all environments is an illusive ideal. So, too then, with moral perception. Obviously, it would be great if one could accurately morally perceive across a wide variety of environments, instantaneously adjusting their priors. And, likewise, it would be great if one's intermediate muscle fibers could be trained to be both fast and slow twitch, whichever being activated at a moment's notice, given the race one is partaking in. But that's just not how the human muscular system – and plausibly, not the human moral perceptual system – works.

Thus, if we do not modify the ideal of the virtuous person and continue to strive to perceive accurately in all moral situations, we will engage in a self-undermining pursuit, making our moral lives into frustrating and counterproductive ones. One modification, then, is that we should not try to pursue accurate moral perception across all environments, but instead employ alternative moral resources. We can look to the ways we could and do navigate our sensory world when experiencing a sensory illusion as a guide to how we might navigate our moral world when under a moral illusion.

Consider the Müller-Lyer Illusion, where we inaccurately experience two lines as different lengths when they actually are not. There are two ways to acquire knowledge about the correct length of the lines: First, we could simply take out a ruler and measure

the lines. Alternatively, we could defer to others who are not susceptible to this illusion. Some empirical research indicates that those who live in "carpentered" environments – spaces structured by straight lines, right angles, and square corners, such as we often see in Western cities, for instance – are more susceptible to this illusion than those who live in "uncarpentered" environments, where dome-like structures are more typical. The Carpentered-world hypothesis posits that the difference in susceptibility to the Müller-Lyer Illusion is explained by differences in living environment, and the information that our visual system is updated on (Segall et al., 1966; Gregory 1968, 2009; Stewart, 1973). Those who inhabit carpentered environments learned the statistics of their physical environment, resulting in particular perceptual priors.

These two ways of navigating the Müller-Lyer Illusion can provide a framework for how to proceed when one is under a moral illusion: One might take out their 'moral ruler' whereby this may consist of comparing like cases to like and/or identifying general underlying moral principles at work. Second, one could rely on the moral testimony of another virtuous person who is prone to having different moral perceptual priors and so not subject to the same moral illusion.

It might not always be obvious whether one is experiencing a moral illusion,[19] and furthermore, it might be even harder to identify a virtuous person who has different moral priors than oneself, and so *not* subject to a moral illusion in this case. But one general feature which predicts when (moral) illusions are likely to occur is when one's priors are not suitable for their environment. Thus, a sudden change in circumstances could make one more likely to experience a moral illusion, and deferring to another person who has long occupied those kinds of circumstances might be a good strategy.[20]

One thing that may complicate this suggestion is the fact that the vicious person may also have an accurate moral perception in the precise situation where the virtuous person experiences a moral illusion. As noted above, the vicious person also has occupied different moral environments and adopted different moral perceptual priors. This might mean that one experiencing a moral illusion has no more reason to defer to a virtuous person (who has different moral priors than her own) over a vicious person, since both could have the correct moral percept in this case. Or, relatedly, even if one should defer the virtuous person rather than the vicious, how will one know this? The reader might

[19] This skepticism doesn't necessarily uniquely plague my proposal, as it arises in various other approaches of moral epistemology. One can almost always reasonably question whether they are having an accurate moral perception, whether they have arrived at the correct moral judgment, or whether they are carrying out the right action.

[20] Consider the case of the life-long pediatric nurse, who has spent her time caring 1-on-1 for rather healthy, young children, providing personalized care and support. Then suddenly, the Covid-19 pandemic hits and she is quickly transferred to a make-shift field hospital to care for dying, mostly adult and elderly, patients. Given the sudden change in contexts, and differences in moral features of her job, she should be on the lookout for moral illusions. Furthermore, this nurse might do better in navigating this new environment by deferring to other nurses who have occupied these sorts of circumstances for quite some time and have developed appropriate moral priors.

be worried that since the same (accurate) moral perceptions can be had by a virtuous and vicious person alike, which both differ from the moral illusion, there will be no way to figure out who the *virtuous* (and not the vicious) person, such that the person experiencing the illusion knows who they ought to defer to.

However, even if the virtuous and vicious persons both have accurate moral perceptions in a given scenario, it's plausible that there would be other ways for one to decipher the virtuous from the vicious. For one, their general traits – virtues versus vices – will vary; and relatedly their intentions will also likely differ. Their varying attitudes and reactions might convey underlying differences in character. Relatedly, one might do better in the long run if she were to defer to the virtuous, rather than vicious, person, for these differences in intentions and attitudes are likely not isolated from one's other moral perceptions and the ways that these other moral perceptions are linked up to actions and habits.

Nonetheless, much more needs to be done to thoroughly examine the potential role that the vicious person might serve in these sorts of situations. Given the unusual circumstances in which moral illusions plague the virtuous person, it may very well be that the vicious person – and their moral perception – has an important role to play. This would be unexpected. But plausibly so are moral illusions, or that inaccurate moral perceptions arise precisely out of one's virtue. And this in itself, is what I have tried to show in this paper: if there is such a thing as moral perception, then we should expect

that one's virtue doesn't just manifest itself in accurate moral perceptions but also might be at work in cases of moral illusions.[21]

Bibliography

Adams, Wendy J., Graf, Erich W., & Ernst, Marc. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience, 7*(10):1057-8.

Anonymous. (2022). "Now that I'm a mom, CIO makes me feel ill." Reddit. https://www.reddit.com/r/AttachmentParenting/comments/uuwsk6/now_that_im_a_mom_cio_makes_me_feel_ill/ . Accessed March 12, 2025.

Annas, Julia. (2004). Being virtuous and doing the right thing. *Proceedings and Addresses of the American Philosophical Association, 78*(2):61-75.

Aristotle. (1984). *The Complete Works of Aristotle*. J. Barnes (Ed.) Princeton, NJ: Princeton University Press.

Audi, Robert. (1995). Acting from virtue. *Mind, 104*(415): 449-471.

---------- (2013). *Moral Perception*. Princeton, NJ: Princeton University Press.

---------- (2022). The phenomenology of moral intuition. *Ethical Theory and Moral Practice, 25*: 53-69.

Avarguès-Weber, Aurore, Finke, Valerie, Nagy, Marton., Szabó, Tunde, d'Amaro, Daniele, Dyer, Adrian G., and Fiser, József. (2020). Different mechanisms underlie implicit visual statistical learning in honeybees and humans. *Proceedings of the National Academy of Sciences, 117*(41):25923-25934.

[21] I am grateful to Adam Waggoner, Allan Hazlett, Casey O'Callaghan, Matt McGrath, John Doris, Evan Westra, Jaqueline Mariña, Pat Kain, J.P. Messina, Peter Railton, members of the Purdue Mind and Cognition Research Working Group, and two anonymous reviewers for very helpful feedback on previous drafts of this paper.

Feldman Barret, Lisa, Russell, James, A. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*: 805–819.

Beck, Jacob. (2018) Analog mental representation. *Wiley Interdisciplinary Reviews: Cognitive Science,* 10(3), e1479-10.

Bedke, Matthew S. (2008). Ethical intuitions: What they are, what they are not, and how they justify. *American Philosophical Quarterly, 45* (3): 253-270.

Behrens, TEJ, Woolrich MW, Walton ME, Rushworth MFS. (2007). Learning the value of information in an uncertain world. *Nat Neurosci, 10*(9): 1214–1221.

Bengson, John. (2015). The intellectual given. *Mind, 124*(495): 707-760

Bloomfield, Paul. (2020). "Humility is not a virtue." In M. Alfano, M.P. Lynch, and A. Tanesini (eds.), *The Routledge Handbook of Philosophy of Humility*, pp. 36-46. London: Routledge.

Blum, Lawrence. (1994). Moral Perception and Particularity. New York: Cambridge University Press.

Carruthers, Peter. (2015). Perceiving mental states. *Consciousness & Cognition, 36*:498-507.

---------- (2018). Valence and value. *Philosophy and Phenomenological Research, 97*(3): 658-680.

---------- (2023). On valence: Imperative or representation of value? *British Journal for the Philosophy of Science, 74* (3): 533-554.

Cheng, Y., Lin, C. P., Liu, H. L., Hsu, Y. Y., Lim, K. E., Hung, D., et al. (2007). Expertise modulates the perception of pain in others. *Curr. Biol. 17*:1708–1713.

Chudnoff, Eli. (2020). *Forming Impressions: Expertise in Perception and Intuition.* Oxford: Oxford University Press.

Clavian, Christine & Fitzgerald, Chloe. (2017). "The evolution of moral intuitions and their feeling of rightness." In R. Joyce (ed.), *The Routledge Handbook of Evolution and Philosophy*, pp. 309-321. New York: Routledge

Cowan, Robert. (2015). Perceptual intuitionism. *Philosophy and Phenomenological Research, 90*:164-193.

Cullison, Andrew. (2010). Moral perception. *European Journal of Philosophy, 18*(2):159-175.

Drefcinski, Shane. (1996). Aristotle's fallible phronimos. *Ancient Philosophy, 16*(1):139- 154.

Ernst, Marc O. (2010). Eye movements: Illusions in slow motion. *Current Biology, 20*(8):R357-9.

Fiser, József. (2009). Perceptual learning and representational learning in humans and animals. *Learning & Behavior, 37*(2):141-53.

Gibson, James J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gigerenzer, Gerd. (2007) *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking Press

Goldie, Peter. (2007). Seeing the kind thing to do: Perception and emotion in morality. *Dialectica, 61*(3):347-361.

Gregory, Richard L. (1968). Visual illusions. *Scientific American*, 21966–76.

---------- (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B, 352:*1121–1128.

---------- (2006). Bayes Window (3): Where do prior probabilities come from? *Perception, 35*(5): 289-90.

---------- (2009). *Seeing Through Illusions*. Oxford: Oxford University Press.

Haidt, Jonathan. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol Rev, 108*(4): 814-34.

Harman, Gilbert. (1977). Ethics and observation. *The Nature of Morality*. Oxford: Oxford University Press, pp. 3–10.

Hesp, Caspar, Smith, Ryan, Parr, Thomas, Allen, Michael, Friston, Karl, J., Ramstead, Maxwell, J. D. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation, 33*(2): 398–446.

Huemer, Michael. (2005). *Ethical Intuitionism.* London: Palgrave Macmillan.

Hursthouse, Rosalind. (2006). "Are virtues the proper starting point for morality?" In *Contemporary Debates in Moral Theory*, Ed. J. Dreier. Blackwell Publishing, pp. 99-112.

Jacobson, Daniel. (2005). Seeing by feeling: Virtues, skills, and moral perception. *Ethical Theory and Moral Practice, 8*(4): 387-409.

Jordan, Andrew. (2013). Reasons, holism, and virtue theory. *The Philosophical Quarterly, 63*(251):248-268.

Keltner, Dacher, Horberg, E. J. & Oveis, Christopher (2012). "Emotions as moral intuitions." In J. P. Forgas (ed.), *Affect in Social Thinking and Behavior,* pp. 161-175. New York: Psychology Press.

Kleiman-Weiner, Max, Saxe, Rebecca, & Tenenbaum, Joshua B. (2017). Learning a commonsense moral theory. *Cognition, 167*:107-123.

Kross, Ethan & Ayduk, Ozlem. (2011). Making meaning out of negative experiences by self-distancing. *Current Directions in Psychological Science, 20*(3):187–191.

Lyons, Jack C. (2018). "Perception and intuition of evaluative properties." In A. Bergqvist & R. Cowan (eds.), *Evaluative Perception,* pp. 183-199. Oxford: Oxford University Press.

Maley, Corey. (2021). "The physicality of representation." *Synthese, 199*(5-6): 14725-14750.

Majumdar, Gargi, Yazin, Fahd, Banerjee, Arpan & Roy, Dipanjan. (2023). Emotion dynamics as hierarchical Bayesian inference in time. *Cerebral Cortex, 33*(7):3750–3772.

McBrayer, Justin. (2010). A limited defense of moral perception. *Philosophical Studies, 149*:305–320.

McGovern, H. T., De Foe, Alexander, Biddel, Hannah, Leptourgos, Pantelis, Corlette, Phillip, Bandara, Kavindu, & Hutchinson, Brendan, T. (2022). Learned uncertainty: The free energy principle in anxiety. *Frontiers in Psychology 13*: 943785.

McGrath, Sarah. (2018). "Moral perception and its rivals." In R. Cowan, A. Bergqvist (Eds.), *Evaluative Perception*. pp. 161-182.

Millar, Boyd. (2015). Naïve Realism and illusion. *Ergo,* 24(2):607-625.

Moss, Jessica. (2009). Akrasia and perceptual illusion. *Archiv für Geschichte der Philosophie, 91*(2):119-156.

Myrighthandwoman. (2019). "Sleep training my 1 year old- the worst part of parenting that I've ever experienced." Reddit. https://www.reddit.com/r/Parenting/comments/ad4sz7/sleep_training_my_1_year_old_the_worst_part_of/ Accessed March 18, 2025.

Nassar MR, Wilson RC, Heasly B, Gold JI. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J Neurosci. 30*(37):12366–12378.

Nichols, Shaun., Kumar, S., Lopez, T., Ayars, A., & Chan, H. Y. (2016). Rational learners and moral rules. *Mind & Language, 31*(5):530-554.

Piray, Payam. & Daw, Nathanaiel, D. (2020). A simple model for learning in volatile environments. *PLoS Comput. Biol, 16*: e1007963.

Piray, Payam &, Daw, Nathaniel, D. (2021) A model for learning based on the joint estimation of stochasticity and volatility. *Nature Communications, 12*(1): 658

Rabinoff, Eve. (2018). *Perception in Aristotle's Ethics.* Evanston, IL: Northwestern University Press.

Railton, Peter. (2014). The affective dog and its rational tail: Intuition and attunement. *Ethics, 124*(4):813-859.

---------- (2017). Moral Learning: Conceptual foundations and normative relevance. *Cognition, 167*:172-190.

Reisenzein, Ranier. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*: 525–539.

Roberts, Roberts. C. (2003). *Emotions: An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press.

---------- (2013). *Emotions in the Moral Life.* Cambridge: Cambridge University Press.

Rodgers, Travis. J., and Warmke, Brandon. (2015). Situationism against Situationism. *Ethical Theory and Moral Practice, 18*(1):9-26.

Roeser, Sabine. (2012). *Moral Emotions and Intuitions.* London: Palgrave Macmillan.

Rolls, Edmund, T. (1999). *The Brain and Emotion*. Oxford: Oxford University Press.

Ross, Lee, & Nisbitt, Ross E. (1991). *The Person and the Situation: Perspectives of Social Psychology*. Mcgraw-Hill.

Russell, James, A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6): 1161–1178.

Scholl, Brian (2006). "Innateness and (Bayesian) visual perception reconciling nativism and development." In. P. Carruthers, S. Laurence, & S. Stich (eds.), *The Innate Mind: Structure and Contents,* pp. 34-52. Oxford: Oxford University Press.

Scholl, Brian J., & Gao, Tao. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford & V. A. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 197-230). Cambridge, MA: MIT Press.

Segall, Marshall. H., Campbell, Donald. T. & Herskovits, Melville. J., (1966). *Influence of Culture on Visual Perception* Indianapolis, IN: Bobbs-Merill.

Seriès, Peggy & Seitz, Aaron. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience, 7:*668.

Shams, L. and Kim, Robyn. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews, 7*:269–84.

Sherman, Nancy. (1991). *The Fabric of Character: Aristotle's Theory of Virtue.* Oxford: Clarendon Press.

Siversten, Sveinung. S. (2019). On the practical impossibility of being both well-informed and impartial. *Erasmus Journal for Philosophy and Economics, 12*(1):52-72.

Smith, A.D., (2002). *The Problem of Perception*. Cambridge: Cambridge University Press

---------- (2010). Disjunctivism and illusion. *Philosophy and Phenomenological Research, 80*(2):384–410.

Stewart, V. Mary. (1973). Tests of the "carpentered world": Hypothesis by race and environment in America and Zambia. *International Journal of Psychology, 8*(2), 83–94.

Stocker, Alan A. & Simoncelli, Eero P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*:578–585.

Tappolet, Christine. (2016). *Emotions, Values and Agency*. Oxford: Oxford University Press.

Topolinski, Sascha & Reber, Rolf. (2010). Gaining insight into the "aha" experience. *Current Directions in Psychological Sciences, 19*, 402–405.

Trope, Yaacov & Liberman, Nira. (2010). Construal-level theory of psychological distance. *Psychological Review, 117*(2):440–463.

VegetablyWorry1492 (2024). Comment on "Why do some people treat parents who support the cry it out method as cruel?" Reddit. https://www.reddit.com/r/NewParents/comments/xo8a7a/comment/ipy92uh/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button. Accessed March 6, 2025.

Vigani, Denise. (2018). Virtuous construal: In defense of silencing. The American Philosophical Association, 5(2):229-245.

Warren, William, H. (2005). Direct perception: The view from here. *Philosophical Topics, 33*(1):335-361.

Weiss, Yair, Simoncelli, Eero P. & Adelson, Edward H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6):598–604

Werner, Preston J. (2016). Moral perception and the contents of experience. *Journal of Moral Philosophy, 13*(3):294-317.

---------- (2020). Which moral properties are eligible for perceptual awareness? *Journal of Moral Philosophy, 17*(3):290-319.

Woodward, James & Allan, John. (2007) Moral intuition: Its neural substrates and normative significance. *Journal of Physiology, 101*:179-202.

Zeusjordie. (2024). Comment on "CIO breaks my heart." Reddit. https://www.reddit.com/r/AttachmentParenting/comments/171khz5/comment/k3s27j1/?utm_source=share&utm_medium=web3x&utm_name=web3xcss&utm_term=1&utm_content=share_button. Accessed March 18, 2025.