

Seeing Wrongness

Josefa Toribio | ORCID: 0000-0002-5180-6755

ICREA-Universitat de Barcelona; Department of Philosophy, University of
Barcelona, Barcelona, Spain

jtoribio@icrea.cat

Accepted 21 April 2024 | Published online 17 July 2024

Abstract

This paper examines the plausibility of an attention-based version of moral perceptualism (AMP). According to AMP, our perception of moral properties is characterized by perceptual attentional patterns that reflect a sensitivity to morally salient features. First, I argue that the explanation for the empirical evidence offered to support AMP primarily hinges on cognitive processes rather than perceptual ones. Second, while I acknowledge the critical importance of attention in recognizing moral properties, I contend that we must expand AMP's explanatory scope to address the question of what drives this attention. I propose an account of our (in)sensitivity to wrongness that builds on AMP's core statement. In this account, the notion of salience structure of information, defined by the varying accessibility of both perceptual and cognitive representations, plays a central explanatory role.

Keywords

moral perception – attention – implicit bias

Careful attention will not reveal colors to the color-blind

QUILL R. KUKLA

1 Introduction

It is a fact that some of our actions strike us as being morally charged, that we can be aware of the moral significance of some of the things that we say or do. It is also a plain fact that we often fail to do so. When we notice the wrongness of an action, whether the action is ours or someone else's, what is the nature of this recognition, and what does it tell us about how we come to know about such a moral property? A proposal that has regained some popularity in addressing this issue is moral perceptualism (e.g., Audi 2013; Cullison 2010; McBrayer 2010; McGrath 2011; Werner 2016). Moral perceptualism holds that we can perceive instantiations of at least some moral properties – e.g., descriptively ‘thin’ evaluative properties, such as being right/wrong or being just/unjust – in a way that is comparable to the way in which we perceive other complex properties.¹ According to this view, which following Vance and Werner (2022), I shall label Contentful Moral Perceptualism (CMP), agents can represent (some) moral properties as part of the content of their perceptual experience.

CMP has been challenged on several fronts. It has been argued that perception traffics in causal connections, and it is hence ill suited to represent moral properties, since moral properties are causally inert (see, e.g., Griffin 1996; Huemer 2005). Another point of criticism has been the absence of a characteristic perceptual appearance or “look” of moral properties (see, e.g., Reiland 2021). Relatedly, critics have also raised the issue of the explanatory redundancy of perception given that perceptual experience would not be a morally charged experience without prior moral beliefs or knowledge (see, e.g., Cowan 2015; Faraci 2015; Reiland 2021). Despite some recent rejoinders to these objections (see, e.g., Cullison 2010; McBrayer 2010), the claim that we can represent moral properties in perception remains a very contentious one.

More recently, the idea that we can perceive moral properties has been characterized as a type of perceptual attunement that consists in a heightened tendency to attend to morally salient features, without this entailing a commitment to the idea that moral properties can be part of the content of perception. Vance and Werner (2022) call this view *Attentional Moral Perceptualism* (AMP) (see also Clifton 2013; DesAustels 2012). The strategy here is to distinguish between two types of perceptual attunement. On the one hand, the perceptual attunement of the morally sensitive agent could be understood as the agent's perceptual ability to identify specific patterns based

¹ The central claim to be discussed here is about the perception of evaluative but still descriptive moral properties such as being right or wrong, not about the perception of moral norms, such as whether we *ought* to act in one way or another (see, e.g., Goldie 2007).

on their moral significance. This reading appears convivial to CMP. In this case, the patterns that the agent identifies are perceptually salient *as* patterns of moral properties. On the other hand, the perceptual attunement of the morally sensitive agent could be understood as a perceptual awareness of general saliency, as a pattern *of* moral salience. This is the reading that AMP favors. The morally sensitive agent need not identify the pattern *as* a specific pattern, and hence the agent need not represent those patterns as moral properties in perceptual experience. A pattern of salience “involves the perceptual system’s selection of certain features in an environment as relevant in such a way that the features are subject to attentional focus in perceptual experience” (Vance & Werner 2022, 7). AMP’s central claim is that the pre-theoretical intuition that some actions strike us as right or wrong can still be explained by the sensitivity of certain *perceptual*, albeit attentional, mechanisms to “moral difference-makers,” which “are the features within one’s perceptual environment that *make a moral difference in that situation*” (Vance & Werner 2022, 9). By shifting the explanation from content to attention, AMP appears to be a more empirically and philosophically plausible version of moral perceptualism – a version that is free of standard objections against CMP.

This paper discusses the plausibility of Attentional Moral Perceptualism (AMP). The version of AMP that I target is the following (Vance & Werner 2022, 7):

AMP (1) Perceptual, attentional mechanisms tend to be sensitive to moral difference-makers and this sensitivity is reflected in attentional patterns in perceptual experience. (2) Moral cognition is influenced by these attentional patterns such that changing patterns of perceptual attention can change moral judgments and decisions.

AMP is a novel view, and its core claim is both appealing and challenging. The central idea of a sensitivity to moral difference-makers that is reflected in attentional patterns in perceptual experience foregrounds processes of receiving and directing focus towards sensory information in a way that runs the risk of inviting an interpretation that understates the role of cognition in the explanation of such an attunement. In the first part of the paper, I argue that construing AMP as entailing that our sensitivity to morally salient situations is merely a matter of *perceptual* attention might be enticing. However, this perspective would render AMP an unlikely stance. For the empirical evidence offered to support AMP clearly illustrates that, even though at least some of the attentional mechanisms are perceptual, the relevant perceptual processes interact with others that are genuinely cognitive, and the weight of the

explanation for the offered evidence predominantly lies with these cognitive processes. Vance and Werner (2022) do gesture in this direction when they claim that the attentional mechanisms at play “could also be the result of feedback loops between moral judgments and perceptual processing over time” (Vance & Werner 2022, 7). Yet, the very idea of this feedback loop suggests an ongoing, dynamic interaction between cognitive processes, in the form of moral judgments, and perceptual processes, governed by attention. This is a rather intricate process that hints at the role of learning and experiential tuning. If perceptual, attentional mechanisms have developed or been honed through such loops over time, then cognition plays a vital and continuous role in shaping our perceptual, attentional mechanisms, thus giving such cognitive processes a dominant explanatory role.

In the second part of the paper, and more positively, I argue that attention does play a pivotal role in our recognition of moral properties, but that we need to broaden AMP’s explanatory scope by further exploring the question of what it is that drives our attention when facing morally charged situations. By putting together elements from the biased competition view of attention and a recent account of prejudice defended by Jessie Munton (2021), I suggest an account of our (in)sensitivity to wrongness that incorporates the important role that attention plays in AMP and builds on its core statement. Some of the attentional mechanisms at play will thus be perceptual, but the central explanatory role will be played by a notion of salience structure of information construed in terms of the different accessibility of both perceptual and cognitive representations.

The paper is organized as follows. Section 2 introduces CMP and briefly sketches the main objections against it. While CMP is not the primary target here, this Section aims to highlight the hurdles that AMP would encounter – as it remains a version of moral perceptualism – given the intertwined nature of perceptual, attentional mechanisms with other processes that extend beyond perception. In Section 3, I review and critically assess the empirical evidence that Vance and Werner (2022) offer to support AMP. Lastly, in Section 4, I bring together the two converging bodies of research mentioned above: the so-called biased competition view of attention (see, e.g., Desimone & Duncan 1995; Duncan 2006) and Munton’s (2021) recent account of prejudice. The key explanatory notion here is not attention itself, but rather what drives attention. And what drives attention, Munton argues, is the different accessibility of certain pieces of information: a salience structure. Prejudice is conceived as an ordering of information that is unduly organized around certain social categories. The strategy in this last Section is to focus on cases in which the wrongness of an action is overlooked as a result of prejudice to shed light on

how we come to know about such a moral property. The central claim is that the virtuous agent's sensitivity to morally charged situations is constituted by appropriately organized salience structures (around moral categories) in the same way in which at least some important forms of prejudicial attitudes are constituted by inappropriately organized salience structures around social categories.

2 Content Moral Perceptualism (CMP)

Pace Harman (1977), the moral perceptualist claims that, when we see a group of young hoodlums pour gasoline on a cat and ignite it, we can, in the same robust sense, *see* that such an action is wrong. CMP comes in many varieties, but there is an important difference between what Werner (2018) calls 'pure' and 'impure' versions of thesis. Pure moral perceptualists hold that our perception of moral properties is not causally or otherwise dependent on any moral beliefs (see, e.g., Werner 2018). Impure moral perceptualists hold, by contrast, that we do represent moral properties in perceptual experience, but only because background moral beliefs and moral knowledge have an influence on the perceptual processing that leads to the experience. In other words, impure moral perceptualists rely on the truth of the so-called cognitive penetrability thesis. Although there are different formulations of the cognitive penetrability thesis in the philosophical market, it is usually characterized as the nomological possibility that two subjects (or the same subject at different times or in different counterfactual circumstances) could have different perceptual experiences as the result of differences in other cognitive (including emotional) background states, while sharing the same proximal stimulus and attending to the same distal stimuli under the same external conditions (cf. Siegel 2012; Macpherson 2012).²

Viewed this way, CMP belongs within so-called rich theories of the content of perceptual experience, according to which we can perceive high-level properties, such as natural kind (Siegel 2010) or esthetic properties (Stokes 2018), and it is offered as an explanation in moral psychology of the etiology and phenomenology, i.e., the what-it-is-likeness, of moral judgments, intuitions, and feelings. Although this will be the main framework for the discussion in this paper, motivations for moral perceptualism, in all its varieties, come not just from moral psychology but also from epistemology, where it is often viewed as

² See, e.g., Stokes (2013) and Machery (2015) for a good discussion and for different accounts of the cognitive penetrability thesis.

providing a fundamental, non-inferential view of moral justification and moral knowledge (see, e.g., McGrath 2004). Here, I intend to remain neutral on the epistemological side of the debate. My interest lies in the moral psychology side of the discussion, especially on the issue of whether perception is rich enough to allow for the representation of moral properties. I thus take for granted representationalism as a general theory of perceptual experience. I assume, together with all major advocates of moral perceptualism, that perceptual experiences have content, that they can be accurate or inaccurate, and that the phenomenal character of such experiences supervenes on their content, i.e., that there could not be a difference in the what-it-is-likeness of two perceptual experiences without a difference in their content. I will focus on impure moral perceptualism, so as to strengthen the plausibility of the view, and I will also restrict the moral perceptualist claim to sense-based perceptual experience, as opposed to a notion of experience that also either involves or is defined by emotional factors – as characterized by, e.g., Martha Nussbaum (1990).

Very briefly, these are the main objections against CMP. First is the so-called causal objection. McBrayer (2010, 292) articulates it as follows:

1. For a subject *S* to perceive a property *F*, *S* must be in appropriate causal contact with *F* (the causal constraint on perception).
2. Human subjects are never in appropriate causal contact with moral properties.
3. So, human subjects cannot perceive moral properties.

To perceive a property *F* in premise 1 should be understood as perceiving something *as F*. No one would deny that the passer-by who witnesses the young hoodlums burning the cat perceives their action and that she may have independent reasons to think that their action is wrong. The crucial claim for moral perceptualism is rather that the passer-by can perceive the hoodlums' action *as* wrong.

Here is another formula for the same objection, labeled by Cullison (2010) *the morally blind objection*: if we can literally see that an action is wrong, then an agent's blindness to wrongness would appear to be the result of a defective perceptual faculty. But the passer-by who witness the hoodlums burning the cat and does not perceive their action to be wrong does not seem to have a defective perceptual faculty. The problem lies elsewhere. So, we cannot perceive moral properties (Cullison 2010, 161).

Not surprisingly, most of the rejoinders to this objection require that the agent already possesses some background information, without which the categorization needed for perceiving something *as* anything would not be possible. Yet, if the perception of moral properties depends on the influence of cognition – whether or not such an influence takes the form of

cognitive penetrability as standardly formulated – then all the explanatory and causal work will be based on the subjects' background moral knowledge and beliefs (see, e.g., Väyrynen 2018). Furthermore, from an epistemological point of view, such moral background beliefs could not, in turn, be justified by perception, on pains of circularity. And if they are not justified by perception, they would seem to be justified *a priori*, which goes against the empiricist slant that the moral perceptualism pursues (see, e.g., Crow 2014). This is, in fact, a version of the objection based on the explanatory redundancy of perception mentioned in the Introduction (see, e.g., Cowan 2015; Faraci 2015).

Finally, let me say a few words about the third objection mentioned in the Introduction: moral properties lack a characteristic perceptual appearance or look (Reiland 2021). Against Cowan (2015), Reiland argues that the putative perceptual experience of moral properties cannot be defended by an appeal to the notion amodal completion, i.e., it cannot be compared with our experiences of entire objects despite the occlusion of some of their parts, or with the experience of the three-dimensionality of a perceived object even though only the facing surface reflects light to our retina. The putative perceptual experience of moral properties cannot be defended, again against Cowan (2015), as a case of experiencing kind membership, as when we perceive that a tree is a pine tree. Both types of comparison rely on the idea of moral properties having a typical look or a typical appearance, but moral properties do not have typical looks.

The purpose of this section has been to briefly examine some of the problems that remain unsolved when considering the plausibility of CMP. Although CMP is not the target of this paper, it is certainly worth keeping these problems in mind because AMP remains a version of moral perceptualism, a version that is taken to be compatible with AMP, as Vance and Werner (2022, 9). acknowledge. We must thus be wary of the pitfalls that AMP could encounter to ensure that its proposed perceptual-attentional sensitivity to moral properties does not fall into the same traps as CMP, especially when it comes to the explanatory and causal role of background knowledge and beliefs. To examine whether it can do it, let us delve deeper into AMP's characterization and the empirical evidence supporting the view.

3 Attentional Moral Perceptualism (AMP) and Empirical Evidence

Attention plays a variety of roles in our mental economy. It is always useful to go back to James' (1890, Vol. 1, Ch. XI, 403–404) classic characterization:

Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called *distraction*, and *Zerstreuung* in German.

In the contemporary literature, we often find a distinction between two types of attention. Bottom-up or stimulus-driven attention is triggered solely by external stimuli that stand out due to their intrinsic characteristics in relation to the background. When you overhear your name being mentioned in someone else's conversation, it catches your attention inadvertently and automatically in this bottom-up manner. Top-down or goal-directed attention, by contrast, involves the internal allocation of attention to certain features, objects, situations or even thoughts or feelings, based on our background knowledge, our interests, goals, and desires. The internal guidance of top-down attention sometimes involves consciously directing your focus toward a specific task, object, or feature, and it is associated with goal-oriented behavior. If I want to know what the group next to me is talking about, I may voluntarily pay attention to their conversation. But top-down attention also works automatically. This is its most interesting function. For instance, if someone is repeatedly exposed to images of cats, and then they are shown a blurry picture that could either be a cat or a dog, their top-down attention, influenced by prior exposure to cat images, will often automatically lead them to perceive the ambiguous image as a cat. This automatic shift in attention is driven by their previous experiences without conscious effort or intention.

The distinction between top-down and bottom-up attention is not completely uncontroversial. The debate ranges from the appropriateness of the terminology (Carrasco 2011), to the adequacy of the dichotomy (Awh et al. 2012), and includes, significantly, whether these two types of attention share or not the same underlying mechanism (Katsuki & Constantinidis 2014). In what follows, I endorse the standard top-down vs. bottom-up distinction and remain neutral about these other issues. It is important to keep in mind though that, however attention is deployed, whether in a top-down or bottom-up manner, the effect seems to be the same. In both cases, what we attend to receives preferential processing, or, as James would say, what we attend to shapes our experience.

If AMP is true, attention is attuned to patterns of salience of perceptual features that, without being moral properties themselves, make a moral

difference in that situation. These are the moral difference-makers that lie at the heart of Vance and Werner's characterization of AMP. Someone's burning a cat's tail draws our attention to a morally relevant difference-maker – a set of non-moral features that allow for the indirect recognition of a moral property. Or to take one of Vance and Werner (2022, 9) examples: if we see a pedestrian being hit by a car that went through a red light, the traffic light being red becomes a moral difference-maker, even though being red is not morally significant. In claiming that our sensitivity to moral difference-makers is reflected in attentional patterns of salience, AMP aims to distance itself from the idea that moral properties are part of the content of our experience (CMP). The strategy here is analogous to the one involved in rejoinders to phenomenal contrast arguments that aim to reject the idea that we do represent high-level properties in perception, while conceding that, due to attentional shifts, there can be a difference in the phenomenology of a perceptual experience before and after we acquire certain recognitional capacities. My aim in this section is to clarify this role of attention in AMP and to do it from the point of view of the empirical evidence offered in its favor.

Vance and Werner (2022) rely on what Gantman and Van Bavel (2014) call *moral pop-out effect* as an important piece of empirical evidence in favor of AMP. What Gantman and Van Bavel show is that, when presented at the threshold for perceptual awareness, people can “correctly identify moral words more frequently than non-moral words” (p. 28). Such an increased frequency in the identification of morally charged stimuli takes place in conditions of perceptual ambiguity. Vance and Werner interpret these results as showing that attentional mechanisms are sensitive to words that refer to moral difference-makers.

Vance and Werner also report on a binocular rivalry study about the influence of gossip on our vision (Anderson et al. 2011). In binocular rivalry experiments, each eye sees a different image, but subjects only see one of them at a time. After a few seconds, the perceived image fades while the suppressed one becomes dominant. It has been known for a long time that images that are brighter or have sharper contrast win the rivalry, thus suggesting a bottom-up attentional mechanism driven by the physical properties of the stimulus. Top-down attention does not seem to play an important role in binocular rivalry. Changes in rivalry dynamics are typically driven by sensory input, with subjects being unable to focus on either stimulus. In this binocular rivalry experiment, however, Anderson and collaborators illustrate how social learning can influence in a top-down fashion the visual salience of a stimulus with the same physical properties. They show that neutral faces, i.e., faces with the same visual appearances, dominate longer in visual awareness when

they are first paired with descriptions of negative – as opposed to positive or neutral – social behavior. Since many of the negative social actions are patently immoral, e.g., cheating, lying, or stealing, Vance and Werner take these findings to show a shift in the subjects' top-down attentional mechanisms driven by their previous exposure to such moral difference-makers.

Vance and Werner survey several other studies to motivate the relevance of attention in morally charged situations: eye tracking in experiments involving moral dilemmas (Decety et al. 2012; Skulmowski et al. 2014; Kastner 2010), gaze tracking in scenarios that offer an incentive for cheating (Hochman et al. 2016), or in game scenarios testing for participants' levels of generosity (Teoh et al. 2020). In all those cases, Vance and Werner argue, it has been empirically demonstrated that “perceptual, attentional mechanisms are sensitive to moral difference-makers and this sensitivity is reflected in attentional patterns in perceptual experience” (Vance & Werner 2022, 17). In addition, such patterns of attention have a causal influence in shaping moral cognition, which, in turn, can influence how we perceive things – the feedback loops between moral judgments and perceptual processing mentioned in the Introduction.

I would like to take issue just with the first and most important part of the claim, i.e., with the idea that the sensitivity to moral difference-makers reflected in attentional patterns is a sensitivity that belongs to perceptual experience. I will focus on the moral pop-out effect and the binocular rivalry cases, but what I have to say also applies to the rest of the evidence that Vance and Werner provide as justification for their claim. Let us begin with the moral pop-out effect. The moral pop-out effect presupposes an understanding of moral concepts, as it is based precisely on the recognition of morally relevant words. So, if attentional mechanisms are at play here, it could only be as part of a more general top-down effect that reflects an influence on the perception of this kind of background knowledge. As previously noted, AMP should be seen as embracing a nuanced notion of perceptual attentional mechanisms influenced and shaped by cognition. Yet, we have grounds to believe that while the moral pop-out effect certainly involves both visual processing and visual experience, its impact on perception is more accurately attributed to a top-down effect on the speed at which memory retrieves certain representations, and only indirectly as an effect on perception. Therefore, unless we view AMP's perceptual attentional mechanisms as having intricate connections between memory retrieval and perceptual processing – as the moral pop-out effect explanation suggests – AMP might encounter criticisms similar to those directed at CMP in the previous section regarding the explanatory redundancy of perception. Here too it seems that the explanation of the effects on visual recognition piggybacks on pre-existing top-down effects on memory. As

Firestone and Scholl (2015, 410) argue, “many alleged top-down effects on perception are explicable as effects on memory rather than on perception.”

Top-down effects on perception are framed as effects on what we see, but many studies of such phenomena instead report effects on how we recognize various sorts of stimuli. By its nature, recognition involves not only visual processing *per se* but also memory retrieval: in order to recognize something, the mind must determine whether the presented stimulus matches some stored representation in memory. For this reason, any improvement in visual recognition could reflect either an influence on ‘front-end’ visual processing (in which case it would challenge claims of encapsulated perception) or merely an influence on ‘back-end’ memory (e.g., influencing how efficiently the relevant memory representations are retrieved).

The effect of priming in lexical decision tasks is a well-known phenomenon in the cognitive sciences. It shows that subjects are much faster at recognizing words like, e.g., “nurse” if they first read “doctor.” Of course, in these tasks, visual recognition necessarily involves visual awareness, but the key explanation of these effects are standard memory processes such as semantic priming. Reading the word “doctor” lowers the standard threshold for the activation of semantically related words, such as “nurse” (see, e.g., Collins & Olson 2014). In Gantman and Van Bavel’s (2014) study, morally relevant words were correctly identified more often than morally irrelevant words. Yet, the most plausible explanation is that they were thus identified because they were all semantically related – they all belonged to the moral realm. The moral pop-out effect undoubtedly involves some perceptual effects. However, the heightened attentional sensitivity to morality thus reflected in perception appears to be best explained as an instance of semantic priming. If that is the case, this effect would generalize to other categories containing semantically related words.

Firestone and Scholl (2015) tested this hypothesis by replicating Gantman and Van Bavel’s pop-out effect for two different categories of words: fashion and transportation. In both cases, they found that subjects were able to correctly identify fashion words such as “stiletto” or “pyjamas” and transportation words such as “car” or “route” more frequently than non-fashion or non-transportation words. They thus conclude that Gantman and Van Bavel’s moral pop-out effect does not show any especial visual sensitivity to morality *per se*, and “should be interpreted in terms of ‘back-end’ memory retrieval rather than ‘front-end’ visual processing” (Firestone & Scholl 2015, 412).

A similar argument could be run for the relevance of Anderson et al.'s (2011) binocular rivalry study on the alleged especial sensitivity of perceptual, attentional mechanisms to the negative moral status of agents' actions. As stated above, Anderson and collaborators presented their study as a case of top-down effect of (negative) gossip on perception via an enhanced attentional sensitivity to negative social actions, i.e., actions which could be considered immoral. The crucial message conveyed by this study is that "top-down affective information acquired through gossip influences vision, so that what we know about someone influences not only how we feel and think about them, but also whether or not we see them in the first place" (Anderson et al. 2011, 4).

Now, gossip is, as Anderson and collaborators acknowledge, a form of social learning, i.e., one of the ways in which we can gain information about other people. As with any other kind of learning, it is not surprising that this type of social learning boosts the deployment of attention to negative information. Negativity bias is, like semantic priming, a well-known phenomenon in the cognitive sciences. Vaish et al. (2008, 383) characterize it as "the propensity to attend to, learn from, and use negative information far more than positive information" (see also, e.g., Ito et al. 1998; Lupfer et al. 2000). However, negativity bias is hardly a perceptual phenomenon, even if attention does play an important role in helping to fix negative features in memory. We are just more likely to encode negative attributes because this allows us to use those memories to recognize the warning signs of a possible negative or threatening experience later. And it makes sense that we are more likely to remember negative attributes, including, e.g., the neutral faces previously associated with negative attributes, because we focus more on the negative than the positive (see, e.g., Huang et al. 2017).

Memory and, in particular, semantic memory, is the key explanatory factor in negativity bias, and so it is in the binocular rivalry experiments that Vance and Werner vindicate as a piece of empirical evidence in favor of their perceptual attention model. Semantic memory, which is our memory for general knowledge and concepts, can shape the perception of images during binocular rivalry by biasing the perception towards images that are more semantically meaningful or are associated with stronger memories. I thus grant that there are attentional effects in the rivalry cases discussed by Vance and Werner, and these effects can influence both perceptual processing and perceptual experience. Having acknowledged that the perceptual attentional mechanisms that AMP posits are deeply influenced by cognitive processes, the prioritized attention affecting these perceptual processes and experiences can be understood as resulting from a broad cognitive phenomenon like negativity

bias. Yet, this picture once again hints at a potential concern: AMP's explanation of our sensitivity to morally charged situations might just be building upon mechanisms typically associated with semantic memory.

Although to the best of my knowledge, there are no experiments aiming to replicate Anderson et al.'s (2011) results with pairings of neutral faces and non-socially relevant, but still negative information, the fact that faces with the same visual appearances dominate longer in visual awareness in a binocular rivalry paradigm when they are first paired with descriptions of negative social behavior seems to be better explained as a result of negativity bias in the retrieval of information from semantic memory. In general, there is plenty of evidence that emotionally charged stimuli, whether they are situations or words, are remembered more accurately than stimuli with no emotional connotation. Moreover, if the valence of the emotion is negative, then it is more likely that the event or the words will be encoded and recalled (see, e.g., Mickley & Kensinger 2008).

To sum up, attentional mechanisms play a significant role in the encoding, retrieval, and preferential processing of information at all levels. AMP's claim regarding perceptual attentional mechanisms' role in detecting morally charged situations holds weight only when we view such mechanisms as deeply rooted in a widespread mesh of shared processes ranging from the perceptual to the cognitive. Moral difference-makers have the attentional grip needed to make a moral difference only when their perception is a blend of perceptual and cognitive elements – shaped by our knowledge, our beliefs, our past experiences, expectations, and context. Consider the red traffic light in the pedestrian accident example: it is indeed a perceptual feature, but one that has a conventional meaning – a meaning that one must learn either through exposition to the norm or through some other kind of cognitive capacity. In this context, our perception of the red light indeed contributes to an evaluative judgment, and it does so by becoming the kind of morally salient feature to which our attention is sensitive. But the red traffic light becomes morally salient in perception because we already know what it *means* – because we allow the overall mental state that we call perceptual experience to include such background knowledge as well as the emotional and affective responses typically associated with it. Therefore, reconciling our analysis of the empirical data Vance and Werner rely on with AMP requires acknowledging a rather rich notion of perception and perceptual attention. Yet, as previously pointed out, this suggests a possible critique: that the perceptual attentional mechanisms at play might be overshadowing the explanatory work performed by background cognitive states and other, more properly cognitive processes and mechanisms.

4 Attention, Implicit Bias, and Salience Structures

If the considerations I have put forward so far are sound, the plausibility of AMP hinges on steering clear of a simplistic understanding of perceptual attentional mechanisms and acknowledging that, while attention indeed facilitates (and at times inhibits) perception, it involves a blend of both perceptual *and* cognitive processes. Attention is mediated by a complex interplay of bottom-up and top-down processing mechanisms that operate at different levels. We selectively process salient information, and salient information – either in the form of external stimuli or in the form of beliefs, expectations, and prior knowledge – captures our attention, thus shaping our resulting cognitive and behavioral responses. Virtuous agents appear notably attuned to morally charged situations. There is thus something interesting and appealing about attention playing an important role in the explanation of our sensitivity to morally salient situations. Rightness and wrongness stand out in our experiences, but to explain the virtuous agent's sensitivity to such properties requires delving deeper into the reasons for such a salience. In this concluding part of the paper, I aim to do just that, i.e., to sketch an account of such sensitivity that, like AMP, takes attentional patterns of salience to be central in recognizing moral properties, but that broadens AMP's explanatory scope by raising the question of what drives such attentional patterns, and does it by invoking processes rooted more deeply in cognition. The perspective outlined here thus aligns with Vance and Werner's foundational presentation of AMP. It can be seen as an evolution or expansion of their view, positioning it within a broader cognitive framework.

The strategy is somewhat indirect. I discuss scenarios in which certain moral properties of an action, such as the property of being wrong, are overlooked to shed light on how we come to know about such properties. The idea is to address the relevance of attention and salience in explaining our *insensitivity* to the wrongness of a particular type of actions: our insensitivity to the wrongness of implicitly biased behavior. Implicitly biased behavior is triggered by implicit attitudes or prejudices, i.e., by attitudes that are at odds with our unprejudiced, egalitarian values. Despite portraying ourselves as agents consciously committed to egalitarianism, i.e., as virtuous agents, we often betray such egalitarian, explicit beliefs by behaving in stereotypically biased, often discriminatory, prejudicial ways. If the egalitarian who behaves prejudicially as a consequence of holding an implicit attitude against someone due to, e.g., their race, gender, nationality, age, or sexual orientation fails to notice the wrongness of their behavior, what is the nature of this failure and what does it tell us about how we come to know about such a moral property?

In this section, I attempt to give an answer to these questions. I argue that attention does play an important role in our recognition of moral properties, but that AMP merely kicks the explanatory can down the road by opening up the question of what it is that drives our attention when facing morally charged situations. By bringing together two converging bodies of research: the so-called biased competition view of attention (see, e.g., Desimone & Duncan 1995; Duncan 2006) and an account of prejudice characterized as a salience structure of information which drives undue attentional patterns (Munton 2021), I suggest an account of our (in)sensitivity to wrongness that incorporates the important role that attention plays in AMP, but that develops it and places it within a larger cognitive economy.

Let us begin with the biased competition view of attention, also known as the biased competition hypothesis. The biased competition view of attention is a theory in cognitive neuroscience that suggests that attention operates by biasing the competition between different neural representations in the brain. The brain has a limited capacity to process all the perceptual stimuli that it receives from the environment. So, it has to optimize its processing resources. Such an optimization is achieved by biasing underlying competitive interactions between such stimuli so that only those which are more behaviourally salient are processed in the relevant cortical areas. The biased competition view suggests that attention can selectively modulate what features become represented at different stages of information processing, from early sensory areas, at the single-neuron level, to higher-level cognitive areas, depending on the task at hand. Both top-down and bottom-up mechanisms exert a biasing influence on any of the competitions going on at each time and such competitions are integrated across different cortical modules.

Bias in the context of this theory is value-neutral; it is just a technical term to refer to assigning an increased processing weight to some information, which thus *wins* the competition and becomes selectively represented. Yet, the principles of the biased competition view of attention offer a lens through which we can understand the mechanics of implicit prejudices; they provide a general framework for understanding how certain features of an individual, such as their race, gender, or age, can serve as powerful bottom-up cues that gain dominance in our cognitive and emotional processing, and how our prior beliefs, stereotypes, and experiences with members of certain social groups can bias our attentional processes in a top-down way. The stimuli that win the attentional competition as a result of these biases are processed more rapidly leading to judgments and behavior that are triggered automatically and without conscious deliberation, thus bypassing our self-declared, explicit egalitarian

commitments and making us blind to certain forms of discrimination, i.e., making us blind to the wrongness of our actions.

Bias, thus, can take on a negative valence when the attentional mechanisms that selectively enhance the representation of some inputs over others are driven by salience structures of information that are unduly organized around certain social categories, such as race, gender, sexual orientation, or age. This is the central claim of Jessie Munton's (2021) account of prejudice, for whom at least some important forms of prejudicial attitudes are constituted by inappropriate attentional patterns to such social categories. Undue attention, however, is a very heterogeneous class. It comprises a selectively enhanced representation of negative features, as when we focus on negatively valenced facts about some racial groups, but also an excessive attention to positive features in detriment of others, as when non-native speakers of English are complimented for their proficiency in the language by native speakers. Most importantly, inappropriate patterns of attention manifest themselves in failures to attend to valuable features of people of other races, religions, or sexual orientations. We should thus go beyond this mixed bag of undue attentional patterns to explain prejudice. We need to look for what drives and unifies them. Munton's proposal is that such a common element is the difference in *accessibility* of certain representations. This variance in the accessibility of information constitutes what Munton calls a *salience structure*.

To understand Munton's notion of salience structure, it would help to compare it with other, recently developed, similar notions, like Sebastian Watzl's (2017) priority structure or Ella Whiteley's (2022) cognitive salience perspective. According to Watzl (2017), attention is captured in terms of a structuring process that renders some parts of our stream of consciousness as central, while leaving others at the periphery. Priority structures are ways of arranging that of which we become consciously aware and are determined by the relative salience of different items or aspects of a situation. They are dynamic structures that can and do change, often rapidly, depending on both external stimuli and internal factors like our goals, interests, background beliefs, or the tasks that we are engaged in. Both top-down and stimulus-driven attention remains a subject's mental action that calls for the prioritization of the stimulus and cannot be identified with any specific sub-personal mechanism. Importantly, however, and this is crucial to understand the difference between Watzl's notion of priority structure and Munton's notion of salience structure, priority structures have content. In fact, Watzl argues, they have an imperative type of content. So, Watzl's notion of priority structure does not appear to be the best fit for developing a general understanding of salience that respects AMP's most distinctive feature against CMP, since priority structures are salient

as patterns of properties that are thus represented in our experience, albeit represented as a command.

Whiteley's (2022) notion of cognitive salience perspective is inspired by Watzl's notion of priority structure. According to Whiteley, a cognitive salience perspective is shaped by dispositions that determine which properties stand out and which are backgrounded in our experience as a result of attention ranking some as more noticeable than others. Whiteley is mainly concerned with what she calls 'harmful salience perspectives,' which are harmful precisely inasmuch as they involve paying attention to the properties of individuals or situations that should not be salient, as race or sexual orientation, or not paying enough attention to properties that deserve to be salient, such as personhood-related traits. An important difference with Watzl's notion of priority structure, however, is that Whiteley's cognitive salience perspectives are defined not by their content, but about the way in which they organize such content, i.e., about the way in which attention structures information. This approach is closer to both Munton's notion of salience structure and much more convivial to AMP's central idea of patterns of salience in the explanation of our sensitivity to morally charged situations. Yet, Munton's approach goes a step further in providing an answer to the question of what drives this important structuring role of attention. The key notion is accessibility.

As I mentioned earlier, Munton's notion of salience structure is defined as the variance in the accessibility of information. Accessibility, in turn, is understood as the ease (or difficulty) with which information is called into play by either perceptual or cognitive mechanisms. Importantly, again, what is salient is not the content of the representations themselves, but their accessibility. Prejudice is not conceived as a type of mental state, whether it is beliefs, associations, mental imagery, or imaginings, but as an ordering of information around certain categories. Prejudice does not become reified as an agent's identifiable attitude, but is "a rigid set of accessibility relations, such that recognizing that someone is a Muslim makes information about terrorism or female repression highly salient" (Munton 2021, 17). Furthermore, salient structures of information range not just over the agent's existing beliefs and background knowledge, but also over information and facts about the agent's physical and social environment.³ Munton's framework is primarily offered as an elucidation of certain forms of prejudice and not, specifically, as an account

3 This aspect of the view fits nicely with the so-called structural view of implicit attitudes illustrated by, e.g., Haslanger (2006), according to which the analysis and the strategies to understand and eradicate such attitudes cannot afford to ignore structural properties of our social environment.

of implicit bias. Yet, it appears aptly tailored to shed light on this type of bias, providing a compelling answer to the central question of what it is that drives our attention. Attention is driven by salient structures of information.

If we put the biased competition view of attention and Munton's account of prejudice together, we can begin to see how to construe an account of our (in)sensitivity to moral properties that builds on the perceptual, attentional mechanisms posited by AMP. On the one hand, the biased competition view of attention allows attention to selectively enhance the representation of some inputs over others and, arguably, it could provide a useful theoretical toolbox for AMP. We just need to highlight the fact that the biased competition posited by this hypothesis ranges over both perceptual and high-level cognitive representations. Think about the empirical evidence offered in favor of AMP discussed above. If the moral pop-out effect is indeed best explained as a case of semantic priming, this effect could still be explained by the biased competition view of attention, because the presentation of a prime word biases the processing of related words by activating a network of *semantic* representations that are associated with the prime. This activation increases the likelihood that the related word will be selected for further processing, leading to a faster and more accurate recognition of the related word. The case of binocular rivalry is even clearer, if understood, as we suggested, as an effect of negativity *bias* in the retrieval of information from semantic memory. Attention is at play, and perceptual awareness is of the essence, but the key explanatory mechanisms involved are primarily cognitive.

On the other hand, if we explain which representations win the competition for attention, not in terms of their content, but in terms of their accessibility, i.e., in terms of the ease (or difficulty) with which they are called into play by either perceptual or cognitive mechanisms, we will be able to offer an account of our (in)sensitivity to morally charged situations that does not entail the perceptual representation of moral properties. In the same way in which at least some important forms of prejudicial attitudes are constituted by inappropriate attentional patterns to social categories such as race, gender, sexual orientation, or age, the sensitivity of the virtuous agent is constituted by appropriate attentional patterns to moral categories, such as right or wrong. The salience structures constituted by the different accessibility of certain representations would be the central explanatory tool. Of course, that means that we need to appeal to an independent story to account for the content of the morally relevant representations themselves, but this is just part of the general aim of any theory of content and something that goes well beyond the scope of this paper.

Acknowledgements

For helpful comments on previous drafts of this paper, I would like to thank the audience at the University of Antwerp, Centre for Philosophical Psychology, and two reviewers for this Journal.

Funding

Research for this article was supported by MICINN/AEI/10.13039/501100011033 and by ERDF A way of making Europe, under grant agreement PID2021-124100NB-I00, by AGAUR, under grant agreement 2021-SGR-00276, and by Grant CEX2021-001169-M funded by MCIN/AEI/10.13039/501100011033.

Biographical Note

Josefa Toribio is an ICREA Research Professor at the University of Barcelona. She holds a PhD from Complutense University, Madrid.

References

- Anderson, E., Siegel, E. H., Bliss-Moreau, E., and Barrett, L. F. (2011). "The visual impact of gossip." *Science*, 332(6036): 1446–1448.
- Audi, R. (2013). *Moral perception*. Oxford: Oxford University Press.
- Awh, E., Belopolsky, A. V., and Theeuwes J. (2012). "Top-down versus bottom-up attentional control: A failed theoretical dichotomy." *Trends in Cognitive Sciences*, 16: 437–444.
- Carrasco, M. (2011). "Visual Attention: The Past 25 Years." *Vision Research*, 51: 1484–1525.
- Clifton, W. S. (2013). "Murdochian Moral Perception." *Journal of Value Inquiry*, 47(3): 207–220.
- Collins, J., and Olson, I. (2014). "Knowledge is power: How conceptual knowledge transforms visual cognition." *Psychonomic Bulletin and Review*, 21, 843–860.
- Cowan, R. (2015). "Perceptual intuitionism." *Philosophy and Phenomenological Research*, 90(1), 164–193.
- Crow, D. (2014). "The mystery of moral perception." *Journal of Moral Philosophy*, 13, 187–210.
- Cullison, A. (2010). "Moral perception." *European Journal of Philosophy*, 18(2), 159–175.
- Decety, J., Michalska, K. J., and Kinzler, K. D. (2012). "The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study." *Cerebral cortex* 22(1): 209–220.

- DesAutels, P. (2012). "Moral perception and responsiveness." *Journal of Social Philosophy* 43(3): 334–346.
- Desimone, R., and Duncan, J. (1995). "Neural mechanisms of selective visual attention." *Annual Review of Neuroscience*, 18: 193–222.
- Duncan, J. (2006). "Brain mechanisms of attention." *The Quarterly Journal of Experimental Psychology*, 59: 2–27.
- Faraci, D. (2015). "A hard look at moral perception." *Philosophical Studies*, 172(8), 2055–2072.
- Firestone, C., and Scholl, B. J. (2015). "Enhanced visual awareness for morality and pajamas? Perception vs. memory in 'top-down' effects." *Cognition*, 136: 409–416.
- Gantman, A. P., and Van Bavel, J. J. (2014). "The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli." *Cognition*, 132(1): 22–29.
- Goldie, P. (2007). "Seeing what is the kind thing to do: Perception and emotion in morality." *Dialectica*, 1–15.
- Griffin, J. (1996). *Value Judgment: Improving Our Ethical Beliefs*. Oxford: Clarendon Press).
- Harman, G. (1977). *The Nature of Morality*. Oxford: Oxford University Press.
- Hochman, G., Glöckner, A., Fiedler, S., and Ayal, S. (2016). "I can see it in your eyes', Biased Processing and Increased Arousal in Dishonest Responses." *Journal of Behavioral Decision Making*, 29: 322–335.
- Huang, P., Liu, C., and Chen, H. (2017). "The influence of Outcome Status on the attending and memory of positive and negative feedback." *Bulletin of Educational Psychology*, 48: 469–486.
- Huemer, M. (2005). *Ethical Intuitionism*. Palgrave MacMillan.
- Ito, T. A., Larsen, J. T., Smith, N. K., and Cacioppo, J. T. (1998). "Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations." *Journal of Personality and Social Psychology*, 75(4): 887–900.
- James, W. (1890). *Principles of Psychology*. New York: Henry Holt.
- Kastner, Rebecca M. (2010). "Moral judgments and visual attention: an eye-tracking investigation." *Chrestomathy: Annual Review of Undergraduate Research, School of Humanities and Social Sciences, School of Languages, Cultures, and World Affairs* 9: 114–28.
- Katsuki, F., and Constantinidis, C. (2014). "Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems." *The Neuroscientist*, 20(5), 509–521.
- Lupfer, M. B., Weeks, M., and Dupuis, S. (2000). "How pervasive is the negativity bias in judgments based on character appraisal?" *Personality and Social Psychology Bulletin*, 26(11): 1353–1366.
- Machery, E. (2015). "Cognitive penetrability: a no-progress report." In J. Zeimbekis and A. Raftopoulos (eds.) *The cognitive penetrability of perception. New philosophical perspectives*. New York: OUP.

- Macpherson, F. (2012). "Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism." *Philosophy and Phenomenological Research*, 84(1): 24–62.
- McBrayer, J. (2010). "Moral perception and the causal objection." *Ratio*, 23(3), 291–307.
- McGrath, S. (2004). "Moral knowledge by perception." *Philosophical Perspectives*, 18(1): 209–228.
- McGrath, S. (2011). "Moral knowledge and experience." In R. Shafer-Landau (Ed.), *Oxford studies in metaethics, volume 6*. Oxford University Press.
- Mickley, K. R., and Kensinger, E. A. (2008). "Emotional valence influences the neural correlates associated with remembering and knowing." *Cognitive, Affective, and Behavioral Neuroscience*, 8(2): 143–152.
- Munton, J. (2021). "Prejudice as the misattribution of salience." *Analytic Philosophy*. DOI: <https://doi.org/10.1111/phib.12250>.
- Nussbaum, M. (1990). *Love's Knowledge*. Oxford: Oxford University Press.
- Reiland, I. (2021). "On experiencing moral properties." *Synthese*, 198: 315–325.
- Siegel, S. (2010) *The content of visual experience*, New York: OUP.
- Siegel, S. (2012). "Cognitive penetrability and perceptual justification." *Noûs*, 46(2): 201–222.
- Skulmowski, A., Bunge, A., Kaspar, K., and Pipa, G. (2014). "Forced-choice decision-making in modified trolley dilemma situations: a virtual reality and eye tracking study." *Frontiers in Behavioral Neuroscience*, 8: 426.
- Speaks, J. (2010). "Attention and intentionalism." *The Philosophical Quarterly*, 60(239): 325–342.
- Stokes, D. (2013). "Cognitive penetrability of perception." *Philosophy Compass*, 8(7): 646–663.
- Stokes, D. (2018). "Rich perceptual content and aesthetic properties." In A. Bergqvist and R. Cowan (Eds.), *Evaluative Perception*. Oxford: Oxford University Press.
- Teoh, Y. Y., Yao, Z., Cunningham, W. A., and Hutcherson, C. A. (2020). "Attentional priorities drive effects of time pressure on altruistic choice." *Nature Communications* 11(1): 1–13.
- Vaish, A., Grossmann, T., and Woodward, A. (2008). "Not all emotions are created equal: the negativity bias in social-emotional development." *Psychological Bulletin*, 134(3): 383–403.
- Vance, J., and Werner, P. (2022). "Attentional moral perception." *Journal of Moral Philosophy*. DOI: 10.1163/17455243-20220001.
- Väyrynen, P. (2018). "Doubts about moral perception." In A. Bergqvist and R. Cowan (Eds.), *Evaluative perception* (pp. 109–128). Oxford: Oxford University Press.
- Watzl, S. (2017). *Structuring Mind: The Nature of Attention and How It Shapes Consciousness*. Oxford/New York: OUP.

- Werner, P. J. (2016). "Moral perception and the contents of experience." *Journal of Moral Philosophy*, 13(3): 294–317.
- Werner, P. J. (2018). "Moral perception without (prior) moral knowledge." *Journal of Moral Philosophy*, 15(2): 164–181.
- Whiteley, E. (2022). "Harmful salience perspectives." In S. Archer (Ed.), *Salience: A Philosophical Inquiry*. Chapter 11. London: Routledge.