

## Unbiased Awarding of Art Prizes? It's Hard to Judge

Ema Sullivan-Bissett and Michael Rush

We have higher-order evidence that aesthetic judgements in the context of awarding art prizes may be affected by implicit bias, to the detriment of artists from marginalized groups. Epistemologists have suggested how to respond to higher-order evidence by appeal to bracketing or suspending judgement. We explain why these approaches do not help in this context. We turn to three ways of addressing the operation of implicit bias: (i) anonymization, (ii) the production of objective criteria, (iii) direct implicit bias mitigation techniques. We show that, in the art prize case, strategy (i) is sometimes counterproductive and any benefits are partial, and strategy (ii) is difficult or impossible to implement. This means that the need for (iii) direct implicit bias mitigation techniques is more pressing here than elsewhere. The art prize context is one where mitigation of *a particular kind* is all we are left with. However, domain-specific problems arise for this strategy too, which call for further empirical work on the operation of implicit bias in the artworld. We conclude that the problem of implicit bias as it arises in the specific context of awarding prizes in the artworld is especially challenging, and given the unavailability of alternative mitigations in this context, the need for direct bias mitigation is even more pressing here than in society in general.

### 1. Aesthetic evaluation and bias

In this paper we will use the term *aesthetic evaluation* interchangeably with *aesthetic judgement*. We use *evaluation* and *judgement* to pick out some inner mental event which informs an aesthetic act (in the Lopesian sense to be explicated shortly). Differences in how these terms are used elsewhere, or ways in which they might usefully be distinguished, can be put aside for our purposes.

Let us situate our discussion alongside some recent work on irrelevant influences on aesthetic judgement. Matthew Kieran points to how aesthetic judgement is influenced by

some irrelevant factors identifiable by empirical science (Kieran, 2010; 2011). We will see shortly how the challenges presented by the factors he identifies are different from the challenge presented by implicit bias. He argues that we are bad at tracking the reasons why we like the art that we do due in part to a range of subconscious factors, for example, status cues and subliminal familiarity to social signals (Kieran, 2011, p. 32). Kieran also examines snobbery (something he takes to be endemic in the artworld (2010, p. 246)), where certain motivational features such as the desire to feel or appear superior to other groups may influence one's aesthetic judgement (2011, p. 36).<sup>1</sup> For Kieran, appreciation and judgement in cases of snobbery are driven by reasons external to appreciation proper. In particular, reasons relating to the elevation of one's social status (Kieran, 2010, p. 244). The epistemic role afforded to pleasure is one reason that aesthetic judgement is especially vulnerable to these irrelevant influences. That role is granted in the aesthetic context since 'it is partly constitutive of an object's being valuable that appreciation is pleasurable, or at least gives rise to pleasure' (Kieran, 2010, p. 249). When one takes pleasure from a work, one is provided with an (albeit defeasible) reason to judge that the work is good. That pleasure might be mistakenly taken to be based on aesthetic goodness, rather than as resulting from various biases.

In addressing the foregoing Kieran proposes engaging in critical self-awareness of one's motives as well as cultivating aesthetic virtues (Kieran, 2011, p. 43). He takes it that doing so would help us prevent our aesthetic judgements being impugned by the irrelevant, and develop us into better aesthetic appreciators.

It is important to note that these irrelevant influences often have their tracks covered, since we do not recognize that our judgements have been guided by them. As Dominic Lopes puts it, 'a well-established body of research in social psychology indicates that we routinely

---

<sup>1</sup> See also Sherri Irvin (2017) for discussion of how various injustices arise due to the convergence of aesthetic judgement on certain bodies and faces as attractive or unattractive.

confabulate the reasons we give for our aesthetic response' (Lopes, 2018, p. 43). Confabulatory reports of reasons for choices have been demonstrated by Petter Johansson and colleagues' (2005) choice blindness studies, where participants are asked for reasons for a choice which they (unbeknown to them) did not in fact make (see also Hall and Johansson, 2009; Hall et al., 2010; 2012; 2013; Johansson et al., 2006; 2008, Strandberg et al., 2018). A natural interpretation of such experimental results is that our choices are not guided by reasons, but our reasons are guided by (what we take to be) our choices.<sup>2</sup>

Let us turn now to a particular kind of irrelevant influence on our aesthetic judgements: implicit biases. These are understood as mental items<sup>3</sup> which are inaccessible to consciousness, automatically activated, and prevalent among even those who identify as egalitarian. They are posited as items which cause common micro-behaviours that cannot be tracked, predicted, or explained by explicit attitudes. In general they are biases which affect our judgments about, and behaviour towards, members of particular, often marginalized, social groups.

The challenge posed by implicit bias in the artworld has received little attention (although Kieran mentions it in passing (2010, p. 249)). It also represents, as we noted earlier, a different challenge from the biases Kieran discusses, which are ones peculiar to aesthetic evaluation due to features of the aesthetic context (i.e. the epistemic role afforded to pleasure, the vulnerability to snobbery). These biases might thus be amenable to what Kieran calls

---

<sup>2</sup> Lopes (2014) offers an alternative interpretation of such studies. The standard interpretation is that participants are wrong about the choice they made (due to experimental manipulation), but nevertheless give reasons for that choice. Lopes suggests instead that the agent might be right about what her choice is, and gives reasons for her choice, but simply does not realise that (due to experimental manipulation) her choice has changed. (For more on these two interpretations and some reasons to prefer the latter, see Bortolotti and Sullivan-Bissett, 2021).

<sup>3</sup> We use the theory neutral 'mental items' so as not to enter the debate over the ontology of implicit bias, since it does not affect the problem we identify (although what we say in §4.3 may need to be sensitive to this). The canonical view developed in psychology is that they are *associations* (see Sullivan-Bissett, *forthcoming* for an overview of the options within this approach). Recently however, some philosophers have argued for propositional understandings of implicit bias, positing that there is a specific relation between their constituents, a relation which is absent if they are associations (Levy, 2015, p. 804). Propositionalists have argued that implicit biases are *patchy endorsements* (Levy, 2015) or *unconscious beliefs* (Mandelbaum, 2016). A third kind of view has it that biases come in both associative and propositional flavours, and are constituted by unconscious imaginings (Sullivan-Bissett, 2019). Other views still have it that implicit biases are *character traits* (Machery, 2016), *aliefs* (Gendler, 2008), or mental imagery (Nanay, 2021). More radical perhaps is the mosaic view (Chehayeb, 2020), or the non-representational account (Johnson, 2021).

*aesthetic education*. Implicit biases though are not peculiar to the aesthetic context, but their operation here is especially difficult to overcome, and behaviours guided by them are liable to confabulatory explanations (see Sullivan-Bissett, 2015). The problem we want to identify then is structurally different, arising from the operation of a more general kind of cognitive bias in the specific context of aesthetic evaluation. It will turn out that the mitigation strategies used elsewhere will suffer under the strain of the particular features of this context.

Let us precisify how we will understand the relationship between an aesthetic evaluation, plausibly an inner mental event, and an action, specifically the awarding of a prize. These need not always walk in step. Obvious unworrying cases include when the work judged best aesthetically does not meet the eligibility criteria for the prize after all (though we might prefer to think of the artwork never, strictly speaking, having been a candidate for the prize in the first place). Or in cases of disagreement among the awarding panel, the prize may legitimately be awarded to an entry that is every judge's second choice, but no one's first. In these cases there might be no bias in either evaluation or awarding, and so no injustice to correct.

More pertinently for our purposes, bias can infect either the aesthetic evaluation or the act of awarding the prize. I could non-biasedly judge that artwork A is the best, but my dislike of the artist may explain why I do not avow my judgement and disallow it from guiding my awarding of the prize. Equally, I could biasedly judge that artwork B is the best, but be guided by other considerations in awarding the prize to a different piece.

Implicit biases and the behaviours they prompt can also dissociate; changes to the former do not always produce corresponding behavioural changes (see e.g. Forscher et al., 2019), and we can change some behaviours without changing the underlying biases which might have disposed us to other behaviours. Focusing on the possible ways in which implicit bias comes apart from biased behaviour prompts a slightly different view of the issue. If we thought that changing implicit bias did not lead to better outcomes for marginalized artists,

we might pay little heed to direct implicit bias mitigation, and instead focus just on the behaviour, perhaps by introducing quotas. If we know that expert evaluators are going to favour awarding prizes to men or White artists, we might specify that a proportion of the prizes go to women or BAME artists.

Such strategies likely have their place, and may take their seat at the table of various interventions. Our focus though is on those cases where biased judgements lead to corresponding biased awarding of prizes, such that if we want to de-bias the awarding of prizes, we should work to de-bias the judgements informing the awarding. Instructive here is Lopes's characterization of *aesthetic acts*, understood as acts operated on an object,  $x$ , counterfactually depending on the agent's aesthetic evaluation of  $x$  (where aesthetic evaluation consists of a mental representation of some object having some aesthetic value) (Lopes, 2018, 34). That is not to say that aesthetic evaluations can always be read off from particular actions (cases where the action did not depend on the aesthetic evaluation would not be Lopesian aesthetic acts). Nor are particular aesthetic acts always predictable from the presence of particular biases (after all, biases are presumably just one part of an overall aesthetic evaluation). We restrict our attention to cases where we have higher order evidence (in a sense to be specified, §3) that our aesthetic acts (the awarding of a prize) might be biased (due to a biased aesthetic evaluation).

## 2. Implicit bias in the artworld

We have good reason to think that implicit biases are rife in our everyday lives (and not only alive and well in the laboratory).<sup>4</sup> Since we have no grounds for positing their absence from the artworld in particular, we should take seriously their operation there too. When

---

<sup>4</sup> One meta-analysis of 2.5 million Implicit Association Test results across seventeen topics found that '[i]mplicit and explicit comparative preferences and stereotypes were widespread across gender, ethnicity, age, political orientation, and region' (Nosek et al., 2007, p. 40). As Jules Holroyd, Robin Scaife, and Tom Stafford have put it following their survey of empirical work on implicit bias in a range of contexts: 'It is very likely that any one individual harbours and is influenced by some sort of implicit bias; [...] we cannot rule out that, for each of us, we are influenced in our judgements or behaviours in discriminatory ways' (2017, p. 2).

considering the participation and recognition of marginalized groups in this context, we can think about two broad kinds of operation of implicit bias. The first is sociological: the participation of marginalized groups in the artworld may be hindered by the biased society of which that world is a part. This creates a problem of *opportunity*, culminating in an *underrepresentation*<sup>5</sup> of marginalized groups, fewer artworks produced by them, and fewer candidates for prizes.<sup>6</sup>

The second kind of operation of implicit bias is more individualistic (although is in part downstream of structural injustice). These are cases in which the *work itself*, as known to be produced by a member of a marginalized group, is judged in a biased way. Though related, these two operations are independent. There could be underrepresentation of members of a certain group, without judgements of their work systematically affected by this underrepresentation (people named ‘Sandra’ might be underrepresented without arty Sandras in particular having judgements of their work influenced by implicit bias). Equally, the artworld could have *overrepresentation* of members of marginalized groups, and yet there still be systematic discrimination exhibited in judgements concerning the value of their work.

It is an empirical question whether marginalized artists are subject to negative effects of implicit bias when their work is evaluated. To speak to this, we reflect briefly on reasons that the art context might make our judgements especially vulnerable to bias, before providing some statistics showing that women and BAME artists<sup>7</sup> are underrepresented on

---

<sup>5</sup> It would be difficult to determine precisely how many artists belonging to marginalized groups are working at any one time in any given medium. We will assume that the underrepresentation of shortlisted and prize-winning artists from certain groups is not always a straightforward artifact of those groups participating in smaller numbers (that is, the second kind of operation of implicit bias we identify is operative) (see fn. 10 for more on this).

<sup>6</sup> Fewer opportunities may impact on the kind of work that gets produced (because of lack of resources, training, etc.), so in some cases artworks by members of certain groups may be of lower quality. We presume that the explanatory work done by this possibility falls at least some way short of explaining the low levels of shortlisting and prize winning for artists in certain groups.

<sup>7</sup> Discrimination on the grounds of other social categories (and indeed intersectional discrimination) likely operate in this context. We focus on women and BAME people because they often come with more identifiable markers (than e.g. sexuality or class), and these are groups on which the empirical work on implicit bias has most often concentrated.

shortlists for prestigious prizes, and even more underrepresented as recipients of those prizes.<sup>8</sup>

The art prize case offers opportunities for bias to operate that are not present in more mundane evaluative cases (e.g. recruitment). To illustrate, let us make some quick remarks on the relationship between aesthetic properties on the one hand, and aesthetic experience and judgements of aesthetic value on the other. Philosophers generally agree that the presence of aesthetic properties need not result in sameness of aesthetic experience or judgements of aesthetic value for all consumers. That is, the presence of particular aesthetic properties does not guarantee that all observers will *thus* have the same aesthetic experience or that all observers will *thus* make the same judgements regarding a work's aesthetic value. Philosophers will disagree on whether such cases reveal that (1) aesthetic value is instrumental, (2) aesthetic experience is not exhausted by acquaintance with aesthetic properties, or, (3) the experience not being intrinsically valued tells us that the relevant aesthetic properties have not, in fact, been experienced (see Stecker, 2006, for discussion).

Whatever explanation one prefers of the presence of aesthetic properties not guaranteeing sameness of aesthetic experience or sameness of judgement regarding a work's value, there is room for the operation of implicit bias to carry at least some of the explanatory burden. Return to (1): if aesthetic value is instrumental, then the operation of implicit biases could play a role in fixing whether a given aesthetic experience is judged to be intrinsically valuable. On (2), if aesthetic experience is not exhausted by acquaintance with certain aesthetic properties, the operation of implicit bias could be part of the story of the further constituents of aesthetic experience. On (3), acquaintance with aesthetic properties does entail some aesthetic experience, but acquaintance with those properties is not guaranteed

---

<sup>8</sup> Implicit bias likely also affects the pricing of artworks. Renée Adams and colleagues document a 'gender discount' of 47.6% in the auction prices for paintings by women, a discount which is higher in countries with greater gender inequality. They hypothesize that artists' gender (and not the quality of work) drives this effect, since in experiments 'participants are unable to guess the gender of an artist simply by looking at a painting and they vary in their preferences for paintings associated with female artists' (2017, p. 1). Adams and colleagues conclude that '[w]omen's art appears to sell for less because it is made by women' (2017, p. 1).

merely by the presence of them. The operation of implicit bias might explain the failure of acquaintance in some cases of presence. We see then that the aesthetic judgement necessary for the awarding of art prizes brings with it special vulnerability to implicit bias.

Let us turn now to some statistics. At time of writing (April 2022), the Turner Prize for outstanding exhibition of art by an artist born or working in Britain has been awarded 36 times, 12 times to women and 6 times to BAME artists.<sup>9</sup> The Booker Prize for the best literary work written in English and published in the UK has been awarded 56 times, 17 times to women, and 10 times to BAME authors. The Nobel Prize for Literature has had 118 winners, including 16 women and 11 BAME authors. The most striking underrepresentation statistics come from prizes for directing. The Academy Award for Best Director has had 97 recipients, 3 of those have been women, and only 9 out of 468 places on the shortlists have been taken by women. More striking still, there have only ever been 4 BAME winners (with one of them winning twice). Until 2020/21, no BAME woman had ever been shortlisted. The story is similar for the British Academy Film and Television Award for Best Director. Of 53 awards, 3 have been made to women, and only 16 women have ever been shortlisted. There have only been 3 BAME winners. Since we know from the empirical literature on implicit bias that women and BAME people *do less well* when their membership of those groups is known by assessors, we suggest that in the artworld, a similar phenomenon is in play.<sup>10</sup>

### 3. Higher-order evidence

We will not join the discussion on how best to characterize higher-order evidence, but will focus on what epistemologists have said about how we ought to respond to it. For our

---

<sup>9</sup> The Turner Prize shortlist for 2021 includes no individual artists, consisting instead of five artist collectives. This makes an assessment of diversity and representation more complicated and we set it aside here.

<sup>10</sup> In the case of the Turner Prize and the Booker Prize, one might think that the levels of BAME representation are roughly in line with the demographic make-up of the population. However, neither prize does well on the representation of women, and the Nobel Prize for Literature, the BAFTA and Academy Awards for Best Director significantly underrepresent both BAME artists and women. Our claim here is not that every art prize radically underrepresents every marginalised group, but as we noted earlier, even without underrepresentation members of marginalized groups could still be subject to systematic discrimination when it comes to judgements concerning the value of their work. Such discrimination need not be a straightforward artifact of underrepresentation.



purposes, following Andy Egan and Adam Elga, we can understand higher-order evidence as *evidence of defective cognitive or perceptual faculties*, that is, evidence which indicates that our mechanisms of belief formation are defective<sup>11</sup> (Egan and Elga, 2005, p. 77).

Egan and Elga distinguish between higher-order evidence of a cognitive faculty's *unreliability* (it does not track truth well), and of its *anti-reliability* (it tends to go positively wrong). That distinction maps different responses called for by higher-order evidence.

An example of *unreliability* is learning that a cognitive faculty with a clearly delineated domain is liable to go awry. Suppose one author of this paper is an excellent baker, he has a good sense of ingredients ratios, can adapt recipes when particular ingredients are unavailable, and reliably produces excellent baked goods. The other author has none of these skills. When she attempts to bake and has to form judgements about how best to respond to issues in the process, she often goes wrong. For example, she might overcompensate with masses of baking powder in response to the unavailability of self-raising flour. When the final product is baked, we might be skeptical that things have gone well.

An example of *anti-reliability* is learning that a cognitive faculty tends to deliver outputs that are mistaken. Suppose one author of this paper can complete a Rubik's cube. She knows the relevant algorithmic movements to apply in a range of recognizable conditions, and reliably completes the cube from any starting state in good time. The other author has none of these skills. When he picks up the cube, he moves it around in a way that is not merely unreliable in improving the starting condition, but is systematic in its making things worse. Knowing this, for any given adjustment of the cube he makes in the service of moving it closer to a state of completion, we might predict that he has in fact moved it further away.

---

<sup>11</sup> A referee notes that characterizing higher-order evidence in terms of evidence of *defect* will rule out cases where we have higher-order evidence that our mechanisms of belief formation are performing especially well. We take the point. However, since we are interested in higher-order evidence of implicit bias infecting our judgements, characterizing it in terms of defect is consistent with our purposes. We also take ourselves to be following standard characterizations of the phenomenon in epistemology.

### 3.1 Responding to higher-order evidence

Once we have higher-order evidence, what ought we to do? David Christensen describes a case of being a subject in an experiment, in which participants are given a drug and then asked to complete some logic puzzles. Although the drug leaves people feeling normal, 80% of those who take it are susceptible to impaired logic-puzzle reasoning, and usually come to incorrect conclusions. Christensen is given a practice question but is then told that the coffee he was given contained the drug. He has evidence that the drug has likely caused epistemic malfunction (Christensen, 2010, p. 187). This is a case of higher-order evidence of *unreliability*.

Christensen claims that the rational response is to become less confident in his answer to the puzzle, even if it turned out that he was in the 20% of folk who are immune to the drug's effects (2010, p. 194). He appeals to *bracketing*:

In accounting for the [higher-order evidence] about the drug, I must in some sense, and to at least some extent, *put aside* or *bracket* my original reasons for my answer. In a sense, I am barred from giving a certain part of my evidence its due.

(Christensen, 2010, p. 195)

Bracketing the evidence is 'to react reasonably to the evidence that [he's] been drugged' (Christensen, 2010, p. 196). Overall, Christensen understands higher-order evidence as 'requiring agents to bracket some of their reasons' (2010, p. 202). Similarly, Egan and Elga suggest, in cases of unreliability we should 'bracket off' that faculty's outputs (treating them with a caution akin to how one might treat the outputs of a faulty watch) (2005, pp. 80-1).

What should we do in cases of anti-reliability? Suppose that instead of having taken a drug, Christensen just knows that he's really bad at logic puzzles, and systematically produces incorrect solutions. According to Egan and Elga, higher-order evidence of an anti-reliable cognitive faculty calls for the suspension of judgement in that faculty's domain (2005,

p. 83). Christensen ought not merely bracket off original reasons for his answer, but suspend judgement on the matter of logic puzzles altogether.

### 3.2 Responding to higher-order evidence in aesthetic evaluation

The claim that knowing about implicit bias can be a source of higher-order evidence regarding the status of evaluative beliefs is not new. Suppose a subject gets higher-order evidence that she is likely to overrate the CVs of men relative to those of women. Christensen notes that whatever else we say about this case, it is clear that the subject, upon receiving such evidence, ought not reason that the first-order evidence (CVs) strongly supports her hypothesis (that a given male is the best candidate), despite her bias (2019, p. 14). In the art prize case, we have higher-order evidence that the judgements regarding the quality of an artwork may not be grounded in only appropriate considerations, due to implicit bias.

This is usually a case of *unreliability* structurally similar to Christensen's drugs case. Implicit biases may impair aesthetic judgement, but this is something which does not affect everyone. Some folk, in some contexts, lack judgement-distorting biases, but one ought not presume that one is such a person, and not give the higher-order evidence of one's likely susceptibility its due.

So what of the advice to *bracket off* either the outputs of the cognitive faculty responsible for aesthetic judgement (Egan and Elga), or one's reasons for the judgement arrived at (Christensen)? Such responses might be suited to reading a faulty watch or answering a logic puzzle but are unhelpful in awarding art prizes. The cognitive faculties involved in aesthetic evaluation affected by bias, and the reasons leading to an evaluative judgement affected by bias, are not easily delineable. Thus we can't *bracket off* some bias-infected sub-set of our cognitive faculties or reasons and make an aesthetic judgement anew in a way properly responsive to higher-order evidence.

Could we do what we epistemically ought if our case were better understood as one of *anti-reliability*, where we ought to suspend judgement in the domain of the faculty for which we have evidence of anti-reliability (Egan and Elga, 2005, p. 83, see also Feldman, 2005, p. 117)? This isn't feasible in the context of awarding art prizes where judgements need to be made one way or another. A judge could respond to higher-order evidence of anti-reliability by recusing herself from the panel, but another judge (subject to the same impugning evidence) would need to be found.

More generally, in cases of unreliability and anti-reliability, where we might bracket off cognitive outputs or reasons, or suspend judgement in a particular domain, other routes to knowledge might be available. These routes might not be vulnerable to the higher-order evidence that prompted us to seek them out. In both the baking (unreliability) and cubing (anti-reliability) cases, the subjects could gain knowledge by consulting an authoritative source or expert testimony. The higher-order evidence in these cases does not generalize to the subjects' abilities to pursue these alternative epistemic routes, nor does it generalize to the likely reliability of those routes.

The art prize case is different. When shortlisting for the Booker prize, upon learning that one's judgements are vulnerable to implicit bias, the judges have no text to consult to guide them (see §4.2). What about (other) expert testimony? Let us quickly note two fairly obvious points: (1) we probably don't want our art prize panelists basing their judgements on the testimony of others, and (2) that testimony is likely to be subject to the same impugning higher-order evidence.

In addition, the possibility of rational belief or knowledge from aesthetic testimony is a contested issue. Many philosophers have argued for some version of aesthetic pessimism, the view that denies 'the legitimacy of forming aesthetic beliefs on the basis of testimony' (Robson, 2013, p. 750). This could be pessimism of the 'unavailability' type, which has it that the aesthetic case includes particular features which rule out knowledge transmission via

testimony (Hopkins, 2011, p. 140). Or it could be pessimism of the 'unusability' type which has it that aesthetic knowledge can be had via testimony, but that forming beliefs on its basis is unacceptable (Hopkins, 2011, p. 140). If either type of aesthetic pessimism is true, we could not endorse judges for art prizes responding to higher-order evidence of their bias by seeking out testimony from others, even other experts.

Of course, another view is aesthetic optimism, which has it that aesthetic knowledge or rational belief *is* available via testimony. However, even if that were right, we could still not advocate testimony as a way to mitigate against bias in the judging of art prizes. Even aesthetic optimists would recognize the inappropriateness here. For example, optimist Jon Robson points out that:

[I]t is not difficult to think of pre-theoretically plausible requirements for being a successful aesthetic appreciator that could not be met by someone whose epistemic expertise in aesthetics is based exclusively on testimony, such as affective engagement with the work, employing their own aesthetic sensibilities, undergoing personal and social development of a peculiar kind, and so on.

(Robson, 2013, p. 759)

Robson goes on to note that we ought not confuse the absence of factors which contribute to being a successful aesthetic appreciator with the absence of aesthetic knowledge (see also Meskin, 2004, p. 76). So even if optimism were true, and even if there were no professional obligations on judges ruling out forming their judgements on the basis of testimony, and even if such testimony could be known to not be impugned by the same higher-order evidence, we take it that we would nevertheless want our judges to be *aesthetic appreciators*, not merely vehicles for aesthetic knowledge.

The idea that knowledge falls short of proper assertion also earns its keep in other domains. Jennifer Lackey has argued that in a range of cases involving testimony with

successful knowledge transfer, there is nonetheless an epistemic shortcoming based on the fact that assertions in these cases are accompanied by expectations about the kinds of grounds on which they are based (Lackey, 2011, p. 271). Of four categories of case where she thinks this occurs, one is in judgements made by experts, and a thought complementary to Robson's optimism is that in our art prize case, knowledge via testimony is possible, but there is still a shortcoming rendering assertions of that knowledge improper.

In sum, the suggestions from the literature on responding to higher-order evidence are less suited to the case of art prizes and implicit bias. In order to work towards making these judgements fairer, we need to respect the higher-order evidence differently, by identifying mitigation strategies.

#### **4. Mitigation strategies**

In what follows we discuss anonymization, objective judging criteria, and direct implicit bias mitigation strategies for panel members, drawing on recruitment as an instructive contrast case. We argue that in the art prize case anonymization is sometimes counter-productive, that the production of criteria raises special problems, and that direct bias mitigation is the best route, but calls for context-sensitive research.

First though, let us speak to the debate over the kind of change we ought to seek if we want to mitigate the effects of implicit bias, or more ambitiously, work to reduce its existence. Broadly speaking, there are two options: approaches which prioritize *individual change* and those which prioritize *structural change*. These kinds of changes roughly map the two kinds of operation of implicit bias that we noted earlier (§2). The first was sociological, where the participation of certain groups in the artworld may be hindered by the biased society of which the artworld is a part. The second was more individualistic, where *biased individuals* were forming problematic evaluative judgements about particular artworks.

Sally Haslanger (2015) defends prioritizing structural change, arguing that if we want to explain how implicit bias functions, we must 'situate it within a broader theory of social structures and structural injustice' (2015, p. 1). This is, in part at least, because a precondition for changing individuals' patterns of thought and action is changing the structures in which they operate. If implicit bias is an internalized ideology which comes about from one's presence within certain social structures, then so long as such structures are maintained, it is a waste of time for individuals within those structures to correct for implicit bias (Haslanger, 2015, p. 8):

If we attempt to change how we perceive and think without changing the social reality that is responsible for the schemas we employ, our efforts are unlikely to be sustainable. (Haslanger, 2015, p. 12).

This has been borne out empirically where it is found that even techniques which do reduce bias do so only in the short-term, since biases 'will reflect whatever local environments [people] are chronically immersed in' (Dasgupta, 2013, p. 271). Indeed, 'their very presence hints at their being not only generated but also maintained by culture' (FitzGerald et al., 2019, p. 9).

So why do we approach the problem of implicit bias in the artworld from the perspective of individual change? The mitigations we consider all focus on correcting or mitigating the biases of individual judges, rather than changing the structures which give rise to such biases. Space constraints prevent a full discussion of the merits of a programme of measures which prioritizes the individual, versus a programme which focuses on the structural. Instead, we draw on work arguing that structural change is crucial, but a focus on individuals can and should be part of that change.

Robin Zheng argues that focusing on issues of individual responsibility is necessary to start and maintain the process of change, and that we cannot understand and adjust the structural side without beginning with the individual:

A normative focus on individuals remains practically necessary insofar as an important part of the work of social change consists precisely in efforts to make normative claims on others that can elicit transformative agency.

(Zheng, 2018, p. 18)

So even for those who think that the action is in looking at structural change, in pursuit of that aim we must focus on individuals.<sup>12</sup> For Zheng, individuals are ‘appropriate bearers of responsibility for structural transformation’ (2018, p. 5), and although we might not be able to attribute biased actions to agents, we can nonetheless hold them accountable for them. And, by extension, agents can be held accountable for structural injustice even though we cannot properly attribute fault to them (2018, p. 8).

Let us turn to our particular context to see how this might work in practice. Consider the artworld as an institution, with its own bias-facilitating structures, that is, as something which isn’t just a small mirror held up to a larger society in general. As we discuss later, the stereotypes peculiar to the artworld are relevant in assessing the likely success of particular interventions. The specific culture of the artworld is explanatorily relevant to the kinds of biases which exist within it. Those who construct panels of judges for artworks, and those who sit on such panels, are the obvious starting points for those wishing to change the local structures of the artworld. Holding such gatekeepers and award-givers accountable for the biased operation of the world in which they play a substantial role should be part of the

---

<sup>12</sup> Zheng also develops a framework of implicit biases according to which they are a *particular type of social structure* (2018, p. 19). If that’s right, then implicit biases are not mere parts of individual psychology to be de-prioritized by those in favour of structural approaches, but are themselves precisely the kind of thing that such approaches are interested in.



project of debiasing the process of prize giving, and the more general operation of the artworld.

Alex Madva takes a similar approach insofar as he takes it that the success of structural change depends in part on the success of individual change. He argues that structural interventions are not ‘two-bird stones’ (interventions which will both address broader conditions of injustice *and* appropriately change the minds of individuals). Rather, we should think of the problems to be addressed (discrimination and inequality) as *two-stone birds* (as requiring both individual and structural interventions) (Madva, 2016, pp. 715–16).

One upshot of Madva’s discussion is that although it can make sense to argue in favour of a particular structural reform over an individual one (e.g. integration over diversity training), it is more instructive to ask which structural reforms we should prioritize, and which individual reforms we should prioritize (2016, pp. 706–7). We agree, and see the work in this paper as asking about which individual reforms we should prioritize *within a particular context*.

We make one final point. Our context is a very narrow one; our conclusions concern not how best to mitigate and prevent social injustice *simpliciter*, but on how to deal with implicit bias in particular in the highly specific context of awarding prizes in the arts. Broad questions about how we approach implicit bias and other mechanisms of injustice *in general* may well not translate straightforwardly to a particular case. We argue that the context of awarding art prizes is a special one, where mitigations used elsewhere (e.g. recruitment) are not appropriate. That intellectual project is of course downstream of debate concerning whether we ought to be prioritizing such methods of mitigation at the expense of looking more broadly at the social structures in which such methods might be embedded. We turn now to three methods of implicit bias mitigation in the context of awarding art prizes.

#### 4.1 Anonymization

In recruitment, anonymization may involve removing identifying information from applications so that biases (implicit or otherwise) are not triggered (studies suggest that identifying information leads to less favourable evaluation of certain candidates, see e.g. Steinpreis et al. 1999). The success of this strategy may still be partial, since often at interview a candidate's membership of certain groups will be harder to obscure.

A similar strategy might be appropriate in some art contexts. Claudia Goldin and Cecilia Rouse found that when auditioning orchestral musicians, using a screen to hide the candidate's identity increased the probability of a woman being advanced out of preliminary rounds by up to 50%, and enhances 'by severalfold' the chances of a woman being the winner of the final round (2000, p. 738). Depending on how different the norms for auditioning and qualifying for a prize are in orchestral music, it may be a case where anonymization is appropriate.

However, anonymization will not work in all art prize cases, for at least two reasons. Firstly, criticism, shortlisting, and consumption of artworks, takes place in the public arena. An artist's name is attached to the work from the beginning, and is often essential to the commercial life and purpose of the work; without an overhaul of the production and consumption of art, we could not mitigate implicit bias using anonymization. Secondly, sometimes anonymization will interfere with aesthetic evaluation. For instance, it may be central to judging an artwork depicting struggle that we know the artist struggled in the relevant way. A mural depicting the experience of female oppression, or a novel about systemic racism in the police, may lose some of their impact if they turn out to have been produced by an influential White male police chief constable. The distinctively perspectival message of some artworks is part of their power, and may contribute to the aesthetic assessment of the work.<sup>13</sup>

---

<sup>13</sup> A similar idea is suggested by Fowler when he says '[t]he persuasiveness of the artist's perspective, the power of his communication to move the recipient toward his position, is decisive in judging the artist's work' (1968, p. 93). See also Irvin (2017, p. 6) for an argument for why parallels of anonymization are inappropriate for attractiveness judgements.

#### 4.2 Objective judging criteria

In recruitment we commonly use person specifications and role criteria. These might mitigate the implicit biases of recruiters to some extent since the candidates likely to be favoured by biased panels must meet some specified and testable criteria. The panel cannot simply hire its friends with no regard for the quality of every applicant; the right kind of criteria will rule this out. In addition, good practice involves the prior weighting of each criterion, to prevent the panel hiring a particular candidate on the basis of his having some feature  $x$  and not hiring another on the basis of her merely having some feature  $y$ , without it being the case that  $x$  is—ahead of time—valued more than  $y$ . The tendency to hire the male candidate on grounds insufficient to hire the female candidate, and so to redefine merit in favour of the male candidate, has been demonstrated empirically (Uhlmann and Cohen, 2005).

In this section we consider whether similar criteria for art prizes might be introduced, either in practice or in principle, as a way of mitigating the potential bias in the members of the judging panel? We argue that the answer is ‘no’.

It is worth first noting that objective judging criteria, analogous to the person specification, are not currently used in the context of awarding art prizes. Indeed, Kieran suggests that the problem of irrelevant influences on aesthetic judgement is ‘compounded by the lack of straightforward publicly available regulative norms’ (Kieran, 2011, p. 36). Gaby Wood, Literary Director of the Booker Prize Foundation, says ‘Man Booker judges have to choose the best. But what does that mean? There is no more specific remit – each set of judges has to make up its own model’ (West, 2018, p. 5). The Hyundai Mercury Prize explains on its website that ‘[t]he judges’ decisions are based solely on the quality of the music on the albums’ and talks of the judges ‘expressing their opinions’ ([mercuryprize.com/about-the-prize](http://mercuryprize.com/about-the-prize)). No further details are given beyond terms and conditions for candidacy. We argued

earlier (§2) that there are several sites at which bias may be operating when formulating aesthetic judgements, and this is the case regardless of judges' credentials.

The dispute over whether there are general principles underpinning aesthetic judgements is the dispute between particularists and generalists. The particularist thinks that general principles do not play a central role, the generalist that they do.

Claire Kirwin (2011, pp. 203–4) distinguishes three claims:

- (1) good aesthetic judges do not need to (or ought not to) appeal to general principles;
- (2) there are no general principles; and
- (3) judges don't in fact appeal to any such principles.

Kirwin takes (1) and (2) to be particularist stances, and (3) as '[a] claim which might be involved in the development of a particularist position' (Kirwin, 2011, p. 204).

We have already said that (3) seems to be true, but that of course does not settle the matter of how things ought to be done, and neither does it settle the question of whether there are actual or possible general principles of the kind under consideration.<sup>14</sup> If (2) is correct then it straightforwardly follows that we can't, even in principle, mitigate bias by appeal to objective criteria. Kirwin's two versions of (1) (*need not* and *ought not*), and of course (3), leave room for thinking that general principles are possible in principle, even if not desirable or possible in practice.

One reason for thinking we *ought not* to appeal to general principles, whether or not there are any to be had, has been articulated by Frank Sibley<sup>15</sup>:

---

<sup>14</sup> A referee observes that the lack of *published* criteria does not mean that judges are not appealing implicitly to general principles. That may in the end be true but part of what is needed to address the problem of bias in which we're interested is documented consistency between judges in the principles appealed to, and confidence that no biases that the judges have are infecting their application of the principles. Implicit appeal to principles would deliver neither of these.

<sup>15</sup> There is an interesting debate between Kirwin (2011) and Anna Bergqvist (2010) over whether Sibley was really a particularist or a generalist. Bergqvist says particularist, Kirwin says generalist (probably).

If someone did merely follow a rule we should not say he was exercising taste, and we should hesitate to admit that he had any real notion of [the aesthetic term in question] until he satisfied us that he could discern it in other instances where no rule was available.

(Sibley, 1959, p. 433)

This chimes with Robson's point above (§2.2) regarding testimony and aesthetic appreciation: we want our art prizes to be decided on the basis of a judgement of the right, distinctively aesthetic, kind.

If general principles were available *in principle*, and offered a viable bias mitigation strategy, but appeal to them would undermine distinctively aesthetic judgement *in practice*, it may be hard to decide whether we would rather our art prizes be infected with bias, or not based on aesthetic judgements. Different oughts would pull in opposing directions and neither prospect is appealing.

For those particularists who think we *need not* appeal to general principles, what we learn about aesthetic judgement from the empirical sciences might be enough to convince this particularist to try for those principles after all. There is also a wealth of more general empirical evidence that appealing to even well-trained expert judgement is usually less reliable than appealing to a general rule. The evidence in other contexts, from psychiatric diagnosis and the likelihood of criminal recidivism, to prediction of academic performance and career satisfaction, is extensive (see e.g. Grove and Meehl, 1996, Grove et al., 2000) and, as Jennifer Zamzow (2015) has argued, the aesthetic particularist would need either to deny the value of rules and principles in general in improving our judgements, in the face of a mass of evidence, or to argue that the aesthetic case is one very isolated exception. We cannot engage further in this debate here, suffice to say that there is some evidence that there are sites for the operation of bias in aesthetic judgement which might be addressed with the

adoption of general principles, giving us good reason to hope for the availability of such principles. Let us consider the prospects of our hopes being met.

In a foundational paper, Monroe Beardsley (1962) outlines the sort of thing we are looking for to be general criteria of aesthetic appreciation and value. He illustrates the point by way of an analogy: ‘if a certain degree of sharpness is a merit in knives [...] then to say that a knife has that degree of sharpness must *always* be a reason to support the conclusion that it is good, and it must apply to all knives of the relevant sort’ (Beardsley, 1962, p. 479). However, there are grounds for thinking that general claims of this kind cannot be articulated for judging some art prizes.

Consider the following problem that arises from the significant heterogeneity in the genres or media of artworks contending for the same prize. Candidates for Best Director may produce dramas or comedies, and the directors have to be judged against each other based on whether they have achieved the quite different things appropriate to those genres. Our basic point generalizes across art forms. Turner Prize entrants may produce sculptures or videos, and candidates for the Booker Prize may include a semi-autobiographical epistolary novel set in 1930s Devon, and a stream-of-consciousness sci-fi story written from the perspective of a canine astronaut. Judges must evaluate the relative merits of different things, done differently, by different people, with different aims. This speaks against the possibility of formulating a list containing features like Beardsley’s knife’s sharpness, since even in the case of a single-medium prize like the Booker individual candidates might be so different from each other as to potentially have very few, and perhaps no particularly interesting, features in common beyond those that make them candidates for the prize in the first place. This amounts to saying that attributive judgements, that an artwork is a good example of its kind – a beautiful *sculpture*, for example – and the production of clearly applicable principles by which to decide such questions, are likely always to be extremely difficult in practice or,

worse, impossible in principle.<sup>16</sup> Mixed-medium prizes like the Turner Prize only make this worry about heterogeneity more obvious.

Perhaps there is open to the particularist an appeal to *objectively assessable features* of artworks that might offer viable mitigation strategy even if general principles cannot, in principle, be formulated. Given that the kind of case we're looking at (the awarding of an art prize) is known in advance, we could specify ahead of time the ways in which different sorts of objectively assessable features of the various candidates should be taken into account. If this works as planned, we will be in a position to make objective judgements despite the lack of any general principles to which the judges can appeal, since for the particularist these features will derive their value or disvalue from interaction with other features of the particular case, rather than from conformity with a general principle.

There are two issues with this. The first is that the exemplification of the objectively assessable features of a work which we might specify in advance will presumably not exhaust the features relevant to which work is most deserving of a prize. Features relevant to mere candidacy may be objective in this way, but features relating to a work's beauty, execution, profundity, and so on will require aesthetic judgement liable to bias. Moreover, even if the objectively assessable features were sufficient to judge the aesthetic value of a work in a given medium, we would still encounter the problem of cross-genre prizes, since no matter how objective the judgement about a sound installation and the judgement about a concrete cast of a shed might separately be, there is no objective way to rank sound installations against concrete casts of sheds that would allow such collections of objective particularist judgements to inform a panel decision in the awarding of a cross-genre prize.

---

<sup>16</sup> Relevant here is the idea that aesthetic evaluation is category-relative. Kendall Walton argues that the features of a given artwork being standard, variable, or contra-standard for its category makes a difference to the aesthetic properties a work seems to have and the properties we might be inclined to attribute to it (1970, p. 354). For Walton, *correctly* perceiving a given work as belonging to a particular category is necessary for perceiving the work correctly. So, an additional complication for formulating objective criteria even for single-medium prizes is identifying what would count as standard, variable, and contra-standard features for a given medium. Since Walton offers some criteria for a work being correctly perceived as belonging to a particular category (1970, pp. 357–8), this additional complication may well be surmountable, and so we leave it aside.

Whether a given installation is better than a given cast will be a matter for the small number of judges on the panel, so if the members of the panel have a damaging bias against sound installations, or against the group to which the artist that produced the sound installation belongs, there is no safeguard here despite the objectivity.

In a last effort to save the possibility of using general principles as part of a bias mitigation strategy in the awarding of art prizes, one might wonder whether we need our criteria for aesthetic evaluation to apply without exception in order for them to mitigate bias. Might it instead be sufficient to formulate criteria that for the most part deliver suitable aesthetic judgements? Consider the distinction between ‘standards’ and ‘guides’ (see e.g. McKeever and Ridge, 2006, pp. 7–9).<sup>17</sup> A ‘standard’ in this sense is an exceptionless rule; perhaps cumbersome to appeal to in practice, but reliably giving clear answers in principle. By contrast a ‘guide’ is easily applied in practice, but might have exceptions.

There are three problems with settling for guides to mitigate bias in the artworld. First, if the criteria applied by the guides are not themselves ones that we can reliably identify and measure, then we leave room for the operation of the very biases we are trying to keep out. It is no use having an easily applicable guide that, e.g., just asks judges ‘did this work move you?’, if the extent to which one is moved, or the way one is moved, varies depending on the biases one harbours. Second, whether or not we can formulate standards or guides for judging particular genres, we are left with the problem of cross-genre prizes. Third, our problem in the aesthetic case is not how to make it practical to judge artworks against an agreed principle. In the moral case we might formulate a guide that is both easy to apply and delivers a morally acceptable course of action in some satisfactory portion of cases, we can be confident this has been achieved only if we have in advance agreed on a more difficult to apply, but more reliable, moral standard against which the success of the guide is to be

---

<sup>17</sup> McKeever and Ridge make their comments in the context of particularism in ethics, and acknowledge that work is needed to show that the same considerations apply in other domains. For our purposes here the distinction is useful even if some of what they go on to say about the moral case is not strictly relevant to the aesthetic case.



judged. The problem in our target case is not merely a practical problem of application of this sort: settling for guides in the face of established but cumbersome standards is one thing; settling for guides when we cannot be confident there are any standards would be quite another.

The general worry with this approach is that we were looking at the possibility of applying general criteria precisely as a way to prevent the personal preference of biased individuals having any illicit influence on the outcome. If we settle for criteria with exceptions, we don't achieve that aim. Absent exceptionless standards, which we have argued are not available, we cannot remove the influence of implicit bias in the judges' decision making by introducing unexceptionless guides.

Let us conclude this section. We began by noting that judging panels for art prizes do not currently appeal to objective judging criteria in coming to their decisions, but that this does not settle the matter of whether such criteria could in principle be formulated, or the question of whether judges ought in fact to appeal to them if they are available.

We argued that particularists endorsing either version of Kirwin's claim (1) face problems. Those taking the view that we *ought not* to appeal to general principles would, if such principles were available and could contribute to bias mitigation, face an uncomfortable choice between prize panels basing their decisions on biased judgements and basing them on judgements that are not distinctively aesthetic. Those taking the view that in practice we *need not* appeal to general principles might (and perhaps should) be moved by evidence from the empirical sciences and decide that appeal to general principles would be desirable if it proved possible.

We argued that heterogeneity in the candidates for art prizes (even within a single medium) speaks against the possibility of producing such general principles. We considered the possibility of a particularist appeal to objectively assessable features of artworks, that might facilitate bias mitigation even in the absence of general principles, but rejected it on the

grounds that either we would encounter again the choice between unbiased but non-aesthetic judgements and distinctively aesthetic but biased judgements, or we would again run up against the problem of cross-genre prizes.

The final option for general criteria we considered was the use of non-exceptionless 'guides' rather than exceptionless 'standards'. We argued that the appeal to guides is designed to solve the problem of using unwieldy but agreed-upon general principles, and cannot help us in the case of awarding art prizes where there are no such general principles to be had.

Overall then, introducing objectively trackable criteria for the awarding of prizes should not be pursued as a way to mitigate implicit bias, for a variety of in principle and in practice reasons. In the next section we turn to our third option for addressing this bias: direct mitigation techniques.

#### *4.3 Direct implicit bias mitigation*

We turn now to addressing the biases directly rather than preventing their being triggered (anonymisation) or compensating for their effects (objective criteria). We have two kinds of approach in mind, which mirror the distinction drawn earlier between the operation of implicit bias at the sociological and individual levels, and our discussion of structural versus individual interventions. To the first, we might address our efforts to society as a whole, reducing systematic structural injustice against the relevant groups. Underrepresentation itself may facilitate the operation of biases, so correcting the structural injustice at the societal level might also address the problems at the individual level (see fns. 5, 6). Reducing systematic structural injustice against the relevant groups could facilitate and encourage the entry of a representative number of members of certain groups into boardrooms and creative disciplines alike. If more people from such groups were present in these spaces, hiring committees and art prize judges would continually be exposed to counter-stereotypical

exemplars (something that has been shown to reduce bias, see e.g. Blair et al., 2001) that could have long-term effects on the relevant biases.

At the individual level, in the recruitment case we can address the problem further downstream by having panel members engage in mitigation strategies. Some of these are increasingly widespread and their efficacy is increasingly well understood.<sup>18</sup> However, these strategies have not been tested in the context of art prizes, which raises two issues for our suggestion that they be implemented.

The first issue relates to context, which some studies have revealed is relevant to the operation of bias. For example, Bernd Wittenbrink and colleagues (2001) found that race bias changed when a Black figure was placed against a church interior background as opposed to an urban setting. Other studies, on virtual reality immersion and implicit bias, have found that embodying participants in dark skinned virtual avatars can lower levels of implicit bias towards Black faces (Peck et al., 2013; Banakou et al., 2016). Similar studies, though, have found that embodying participants in dark skinned avatars or female avatars *increased* levels of implicit bias towards Black and female faces (Groom et al., 2009; Lopez et al., 2019). One explanation of these latter results appeals to context: participants were in a job interview context and a sports context in the virtual reality simulations, and these are negatively stereotyped situations for Black people and women respectively. The suggestion is that the increase in bias was due to being placed in a situation known for discrimination (Peck et al., 2013, p.785; Banakou et al., 2016, p. 8; Slater, 2017, p. 26), and that stereotype activation could have ‘overwhelmed any positive effects of perspective-taking’ (Lopez et al., 2019, p. 3).

These studies teach us that context is key in understanding the likely operation of bias, and how to mitigate it. One question we should ask then is whether various art contexts are negatively stereotyped for certain groups. For example, if photography is a negatively

---

<sup>18</sup> Such strategies have had their efficacy challenged, and there are growing doubts that changes in implicit biases as measured by implicit measures make a difference to bias-infected judgements or behaviours (see e.g. Forscher et al., 2019). We leave these broader issues aside (which apply across contexts), and focus instead on particular problems with implementing these strategies in the artworld.

stereotyped enterprise for women but sculpture is not, that might suggest that biases might be more likely to be operative in photography, but not in sculpture. When we turn to mitigation strategies, in the photography case, but not the sculpture case, we might need neutral spaces. Judges of photography might engage in general mental imagery exercises concerning strong women (as Blair et al., 2001, found to be effective), whilst judges of sculpture might engage in virtual reality environments that abound with female sculptors. Before recommending a given kind of mitigation technique ahead of aesthetic evaluation in the service of art prize awarding, work is needed on stereotypes in the artworld, and the effects of placement in a stereotyped context on successful mitigation.

The second issue relates to different kinds of bias in play in the context of awarding art prizes. Of course implicit biases differ in their contents, insofar as they are *about* different social groups (e.g. women, Black and ethnic minority people, etc.). But implicit biases concerning the *same social groups* can vary with respect to their expression. That is, a given individual may score highly on an Implicit Association Test (IAT) on associations between e.g. Black men and stereotypical traits, but they may not score highly on an IAT testing for negatively valenced associations involving their concept of Black men, and vice versa (see Amodio and Devine, 2006). These distinct biases are predictive of different behaviours (the former influencing judgements of competence, the latter influencing seating distance from a Black confederate). Returning to the art prize case, it is not as straightforward as employing strategies which result in adjusted IAT scores, since those strategies may not result in adjusted behaviours *of the relevant sort*.

So, are results on race and gender IATs of various kinds predictive of less good evaluative judgements of artworks by members of certain groups? For example, it might be that semantic associations are in play (the kind which affect judgements of *competence*) rather than valenced associations (the kind which affect *seating distance*), which might call for different mitigation strategies. Relatedly, are different implicit measures (e.g. affective

priming tasks) better predictors of biased aesthetic judgement than IATs? Empirical work on the operation of implicit bias in the context of art prizes can help us answer these questions, better equipping us to tackle bias in this context.

If we better our understanding of the operation and mitigation of bias in this context, we face the broader problem that strategies for mitigating the expression of implicit biases are limited, dependent on local interventions, and isolated from the kinds of societal inputs which shape our biases. As we saw earlier in our discussion of structural versus individual reform, even bias techniques which are deemed successful are only so in the very short term. Our social structures are powerful enough to undo the good work of intervention strategies. Judges would need *repeated* engagement with mitigation strategies, and, at least for those involving exposure to counter-stereotypical exemplars, we know that *repeated exposure* can influence our biases long-term (Huebner, 2016, p. 70). If repeated interventions are not possible, then different intervention strategies need to be devised to have more durable changes in implicit bias (FitzGerald et al., 2019, p. 9).

In the absence of alternative routes of anonymization or the introduction of objective criteria, we suggest that implicit bias mitigation techniques represent the best route for tackling implicit bias in the context of awarding art prizes. However, we have raised two issues for this strategy based on paucity of research in this area, which call for further empirical work.

## 5. Conclusions

We have argued that it is particularly difficult to respond appropriately to higher-order evidence of the operation of implicit bias in awarding art prizes. Advice from epistemologists regarding bracketing or suspending judgement is of little help in this case. Common approaches to mitigating implicit bias in other contexts (anonymization or clearly applicable objective criteria) are unpromising. We suggested that direct implicit bias mitigation

techniques represent the best route to tackling bias in this context, but identified two issues which call for further empirical investigation. The purpose of our work here was to argue that implicit bias mitigation *of a particular kind* is called for in the context of awarding art prizes, but we note that the details of the implementation of this kind of mitigation is for another day. To conclude, we suggest that further work ought to be done on the operation of implicit bias in the context of aesthetic evaluation, so that we may better understand the art of responding to higher-order evidence.

### **Acknowledgements**

Ema Sullivan-Bissett acknowledges the Arts and Humanities Research Council for funding her work on this paper (*Deluded by Experience*, grant no. AH/T013486/1). Both authors are very grateful to Daniel Whiting and two referees for this journal for extremely helpful comments which improved the paper.

### **References**

Adams, R., Kräussl, R., and Navone, M. (2017) 'Is Gender in the Eye of the Beholder? Identifying Cultural Attitudes with Art Auction Prices'. Working paper. Goethe University.

Amodio, D. M. and Devine, P. G. (2006) 'Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior'. *Journal of Personality and Social Psychology*, 91, pp. 652–661.

Banakou, D., Groten, R., and Slater, M. (2013) 'Illusory Ownership of a Virtual Child Body Causes Overestimation of Object Sizes and Implicit Attitude Changes'. *Proceedings of the National Academy of Sciences U.S.A.*, 110, pp. 12846–12851.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Beardsley, M. C. (1962) 'On the Generality of Critical Reasons'. *The Journal of Philosophy*, 59(18), pp. 477–486.

Bergqvist, A. (2010) 'Why Sibley is Not a Generalist After All'. *BJA*, 50(1), pp. 1–14.

Blair, I. V., Ma, J. E., and Lenton, A. P. (2001) 'Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery'. *Journal of Personality and Social Psychology*, 81(5), pp. 828–841.

Bortolotti, L. and Sullivan-Bissett, E. (2021) 'Is Choice Blindness a Case of Self-Ignorance?' *Synthese*, 198, pp. 5437–5454.

Chehayeb, F. (2020) *Contra Implicit Bias*. PhD Thesis. University of Birmingham.

Christensen, D. (2010) 'Higher-Order Evidence'. *Philosophy and Phenomenological Research*, 81(1), pp. 185–215.

Christensen, D. (2019) 'Formulating Independence', in Skipper, M. and Steglich-Petersen, A. (eds), *Higher-Order Evidence: New Essays*. Oxford: Oxford University Press, pp. 13–34.

Dasgupta, N. (2013) 'Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes, and the Self-Concept'. *Advances in Experimental Social Psychology*, 47, pp. 233–279.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Del Pinal, G. and Spaulding, S. (2018) 'Conceptual Centrality and Implicit Bias'. *Mind & Language*, 33, pp. 95-111.

Egan, A. and Elga, A. (2005) 'I Can't Believe I'm Stupid'. *Philosophical Perspectives*, 19, pp. 77-93.

Feldman, R. (2005) 'Respecting the Evidence'. *Philosophical Perspectives*, 19, pp. 95-119.

Fitzgerald, C., Martin, A., Berner, D., and Hurst, S. (2019) 'Interventions Designed to Reduce Implicit Prejudices and Implicit Stereotypes in Real World Contexts: A Systematic Review'. *BMC Psychology*, 7(29), pp. 1-12.

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., and Devine, P. G. (2019) 'A Meta-Analysis of Procedures to Change Implicit Measures'. *Journal of Personality and Social Psychology*, 117(3), pp. 522-559.

Fowler, C. B (1968) 'Perspectivism: An Approach to Aesthetic Evaluation'. *The Journal of Aesthetic Education*, 2(1), pp. 87-99.

Gendler, T. (2008) 'Alief in Action (and Reaction)'. *Mind & Language*, 23(5), pp. 552-585.

Goldin, C. and Rouse, C. (2000) 'Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians'. *American Economic Review*, 90(4), pp. 715-741.

Groom, V., Bailenson, J. N., and Nass, C. (2009) 'The Influence of Racial Embodiment on Racial Bias and Immersive Virtual Environments'. *Social Influence*. 4(3), pp. 231-248.



Grove, W. M. and Meehl, P. E. (1996) 'Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy'. *Psychology, Public Policy and Law*, 2(2), pp. 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., and Nelson, C. (2000) 'Clinical Versus Mechanical Prediction: A Meta-Analysis'. *Psychological Assessment*, 12(1), pp. 19–30.

Hall, L. and Johansson, P. (2009) 'Choice Blindness: You Don't Know What You Want'. *New Scientist*, 2704, pp. 26–27.

Hall, L., Johansson, P., and Strandberg, T. (2012) 'Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-Transforming Survey'. *PLoS ONE*, 7(9), p. e45457.

Hall, L., Johansson, P., Tärning, B., Sikström, S, and Deutgen, T. (2010) 'Magic at the Marketplace: Choice Blindness for the Taste of Jam and the Smell of Tea'. *Cognition*, 117(1), pp. 54–61.

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., and Johansson, P. (2013) 'How the Polls can be Both Spot on and Dead Wrong: Using Choice Blindness to Shift Political Attitudes and Voter Intentions'. *PLoS ONE*, (8), p. e60554.

Haslanger, S. (2015) 'Distinguished Lecture: Social Structure, Narrative, and Explanation'. *Canadian Journal of Philosophy*, 45(1), pp. 1–15.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Holroyd, J., Scaife, R., and Stafford, T. (2017) 'Responsibility for Implicit Bias'. *Philosophy Compass*, 12, pp. 1–13.

Hopkins, R. (2011) 'How to be a Pessimist About Aesthetic Testimony'. *The Journal of Philosophy*, 108(3), pp. 138–157.

Huebner, B. (2016) 'Implicit Bias, Reinforcement Learning, and Scaffolded Moral Cognition', in Brownstein, M. and Saul, J. (eds), *Implicit Bias and Philosophy: Metaphysics and Epistemology*. Oxford: Oxford University Press, pp. 47–79.

Johansson, P., Hall, L., and Sikström, S. (2008) 'From Change Blindness to Choice Blindness'. *Psychologia*, 51, pp. 142–155.

Johansson, P., Hall, L., Sikström, S., and Olsson, A. (2005) 'Failure to Detected Mismatches between Intention and Outcome in a Simple Decision Task'. *Science*, 310(5734), pp. 116–119.

Johansson, P., Hall, L., Sikström, S., Tärling, B., and Lind, A. (2006) 'How Something Can be Said about Telling More Than We Know'. *Consciousness and Cognition*, 15, pp. 673–692.

Johnson, G. M. (2020) 'The Structure of Bias'. *Mind*, 129(516), pp. 1193–1236.

Kieran, M. (2010) 'The Vice of Snobbery: Aesthetic Knowledge, Justification and Virtue in Art Appreciation'. *The Philosophical Quarterly*, 60(239), pp. 243–263.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Kieran, M. (2011) 'The Fragility of Aesthetic Knowledge: Aesthetic Psychology and Appreciative Virtues'. In Schellekens, E. and Goldie, P. (eds), *The Aesthetic Mind: Philosophy and Psychology*. Oxford: Oxford University Press, pp. 32–43.

Kirwin, C. (2011) 'Why Sibley is (Probably) Not a Particularist After All'. *BJA*, 51(2), pp. 201–212.

Lackey, J. (2011) 'Assertion and Isolated Second-Hand Knowledge'. In Brown, J. and Cappelen, H. (eds), *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, pp. 251–276.

Levy, N. (2015) 'Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements'. *Noûs*, 49(4), pp. 800–823.

Lopes, D. M. (2014) 'Feckless Reason'. In Currie, G., Kieran, M., and Meskin, A. (eds), *Aesthetics and the Sciences of Mind*. Oxford: Oxford University Press, pp. 21–36.

Lopes, D. M. (2018) *Being for Beauty: Aesthetic Agency and Value*. Oxford: Oxford University Press.

Lopez, S., Yang, Y., Beltran, K., Kim, S. J., Hernandez, J. C., Simran, C., Yang, B., and Yuksel, B. F. (2009): 'Investigating Implicit Gender Bias and Embodiment of White Males in Virtual Reality with Full Body Visuomotor Synchrony'. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, paper 557, pp. 1–12.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Machery, E. (2016) 'De-Freuding Implicit Attitudes'. In Brownstein, M. and Saul, J. (eds), *Implicit Bias and Philosophy. Volume one: Metaphysics and Epistemology*. Oxford: Oxford University Press, pp. 104–129.

Madva, A. (2016) 'A Plea for Anti-Anti-Individualism: How Simple Psychology Misleads Social Policy'. *Ergo*, 3(27), pp. 701–728.

Mandelbaum, E. (2016) 'Attitude, Inference, Association: On the Propositional Structure of Implicit Bias'. *Nouûs*, 50(3), pp. 629–658.

McKeever, S., and Ridge, M. (2006): *Principled Ethics: Generalism as a Regulative Ideal*. Oxford: Oxford University Press.

Meskin, A. (2004) 'Aesthetic Testimony: What Can We Learn from Others about Beauty and Art?' *Philosophy and Phenomenological Research*, 69, pp. 65–91.

Nanay, B. (2021) 'Implicit Bias as Mental Imagery'. *Journal of the American Philosophical Association*, 7(3), pp. 329–347.

Nosek, B., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., and Banaji, M., R. (2007) 'Pervasiveness and Correlates of Implicit Attitudes and Stereotypes'. *European Review of Social Psychology*, 18(1), pp. 36–88.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013) 'Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias'. *Consciousness and Cognition*, 22, pp. 779–787.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Robson, J. (2013) 'Aesthetic Testimony and the Norms of Belief Formation'. *European Journal of Philosophy*, 23(3), pp. 750–763.

Sibley, F. (1959) 'Aesthetic Concepts'. *The Philosophical Review*, 68(4), pp. 421–450.

Slater, M. (2017) 'Implicit Learning Through Embodiment in Immersive Virtual Reality'. In Liu, D., Dede, C., Huang, R., and Richards, J. (eds), *Virtual, Augmented, and Mixed Realities in Education*, Springer, pp. 19–34.

Stecker, R. (2006) 'Aesthetic Experience and Aesthetic Value'. *Philosophy Compass*, 1(1), pp. 1–10.

Steinpreis, R. E., Anders, K. A., and Ritzke, D. (1999) 'The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study'. *Sex Roles*, 41(7), pp. 509–528.

Strandberg, T, Sivén, D., Hall, L, Johansson, P, and Pärnamets, P. (2018) 'False Beliefs and Confabulation can Lead to Lasting Changes in Political Attitudes'. *Journal of Experimental Psychology*, 147(9), pp. 1382–1399.

Sullivan-Bissett, E. (2015) 'Implicit Bias, Confabulation, and Epistemic Innocence'. *Consciousness and Cognition*, 33, pp. 548–560.

Sullivan-Bissett, E. (2019) 'Biased by Our Imaginings'. *Mind and Language*, 34(5), pp. 627–647.

Forthcoming in *The British Journal of Aesthetics*. [Please cite published version.]

Sullivan-Bissett, E. (forthcoming) 'Implicit Bias and Processing'. In Thompson, Robert (ed), *The Routledge Handbook of Philosophy and Implicit Cognition*.

Uhlmann, E. L. and Cohen, G. L. (2005) 'Constructed Criteria. Redefining Merit to Justify Discrimination'. *Psychological Science*, 16(6), pp. 474–480.

Walton, K. L. (1970): 'Categories of Art'. *The Philosophical Review*, 97(3), pp. 334–367.

West, G. (2018) 'For the Taking'. Originally in *The Times Literary Supplement*.

<https://thebookerprizes.com/sites/manbosamjo/files/TLS%20piece%20as%20published.pdf> Accessed 07/01/21.

Wittenbrink, B., Judd, C. M., and Park, B. (2001) 'Spontaneous Prejudice in Context: Variability in Automatically Activated Attitudes'. *Journal of Personality and Social Psychology*, 81(5), pp. 815–827.

Zamzow, J. L. (2015) 'Rules and Principles in Moral Decision Making: An Empirical Objection to Moral Particularism'. *Ethical Theory and Moral Practice*, 18(1), pp. 123–134.

Zheng, R. (2018) 'Bias, Structure, and Injustice: A Reply to Haslanger'. *Feminist Philosophical Quarterly*, 4(1), article 4.