Demographic Statistics in Defensive Decisions*

Renée Jorgensen Bolinger

Abstract

A popular informal argument suggests that statistics about the preponderance of criminal involvement among particular demographic groups partially justify others in making defensive mistakes against members of the group. One could worry that evidence-relative accounts of moral rights vindicate this argument. After constructing the strongest form of this objection, I offer several replies: (i) most demographic statistics face an unmet challenge from reference class problems, (ii) even those that meet it fail to ground non-negligible conditional probabilities, (iii) even if they did, they introduce new costs likely to cancel out any justificatory contribution of the statistic, but (iv) even if they didn't, demographic facts are the wrong sort to make a moral difference to agents' negative rights. I conclude that the popular argument should be rejected, and evidence-relative theories do not have the worrisome implication.

1 Setting the stage

When predicting how a stranger will act, we often rely on information about how *other people like them* have acted in similar circumstances. By 'like them', we mean something quite coarse-grained: people from their city, or of their gender, or racial or ethnic group. In most cases we probably actually rely on generics or stereotypes we have internalized about the group, like *white people can't dance*, or *Asians are good at math*, and simply assume that they will accurately characterize the stranger.[1] We are usually willing to concede that generics and stereotypes lack justificatory force to support these kinds of inferences to new cases. But if we were a bit more careful — or had assistance from computer programs designed to predict behavior — we might instead base our expectations on demographic statistics, matching our credence that the particular stranger will have a feature f to the proportion of their social group that has f.

There are a lot of questions we should ask about the appropriateness of this kind of statistical inference, even when nothing much hangs on it. But these questions are more pressing when more depends on our predictions, as when making hiring, threat assessment, or sentencing decisions. They are most urgent in self-defense contexts, when we are attempting to judge whether someone

*Author's draft. Please refer to the published version in *Synthese* for citation purposes.

[1] See Leslie (2017) for discussion.

is going to unjustly attack us. If they are, then (on most accounts) it is morally permissible for us to impose proportionate harm necessary to defend ourselves. If they would not attack, then obviously we should not harm them. A mistake either way involves serious harm. And since we'll have to decide how to act before we can be certain about their intentions—and often without knowing much about them—we will have to guess, and it is tempting to rely on generalizations and statistics about demographic groups to guide our decision.

This paper focuses on the use of demographic statistics in decisions about personal self-defense, rather than in institutional settings. Primarily this is for simplicity: it can be difficult to articulate precisely what rights individuals have concerning the decision-processes used in institutional contexts, and murky to balance risks to a profiled individual against social interests in security. The personal self-defense case is comparatively tractable: focusing on one-to-one decisions allows us to bracket most questions about how to make difficult tradeoffs, and the content of the rights at stake is relatively well-theorized. There are two additional reasons to prioritize cases of this sort, leaving extension to other contexts as a project for another paper. First, a popular argument in informal discussions presents specific demographic statistics as grounds for the claim that defensive mistakes made against members of certain groups are reasonable, and agents who make them should not be punished, indicted, or even blamed. Second, one might think these claims are actually vindicated by moral theories according to which agents' objective moral duties to each other—and thus their corresponding objective moral rights—are evidence-relative: e.g. whether B has a right against A's imposing harm on him depends on A's evidence.

These theories come in a few important sub-varieties. The *expected-value variant* maintains that rights and duties depend on facts about the moral value of the outcomes of an agent's actions, weighted by their probability on the agent's evidence. On Zimmerman (2008)'s view, for instance, objective moral duties are *prospective*: they require agents to do what, given their rational credences, is their morally best prospect ("what is best in light of one's evidence about what matters morally"[2]). This understands A's right against being harmed by B as actually a right that B avoid actions that, on her available evidence, are likely to unjustifiedly harm A. Such a right is not violated if B harms A when her evidence suggests that the harm is justified (e.g. because it suggests A is an unjust aggressor), or that though her action has some risk of harming an innocent A, this is sufficiently unlikely that given the comparative values of the more likely outcomes, running the risk is her morally best option. The *justified-belief variant*, meanwhile, construes an agent's moral duties as a function of her justified beliefs. van der Vossen (2016), for example, holds that agents are permitted to impose defensive harm on someone Y when the agent "…is justified in believing either that Y is about to perform an objectively rights-violating act, or Y has culpably acted in a

[2]Summarized in Zimmerman (2018); developed in detail in Zimmerman (2008, 2014). He construes agents' moral duties as requiring them to perform actions with sufficiently good *prospective* moral value; this is not a straightforward expected-value maximization standard. Zimmerman clarifies that "I take an agent's prospectively best option to be that which it would be most reasonable for him to choose in light of his evidence, were he morally conscientious." (2018:454)

way that would, if successful, be rights-violating act, or both."[3] Along similar lines, Frowe (2010) argues that self-defense is permissible when the agent "reasonably believes that (a) she is innocent, and (b) if she does not kill this person, then they will kill her."[4]

One might think that evidence-relative views vindicate the popular argument by implying something like the *statistics claim*:

> STATISTICS CLAIM: if A is a member of a group G, and I match my credence that A is a violent aggressor to the probability of a person's being involved in violent crime, conditional on being in G, and that credence is sufficiently high, then even if in fact (unbeknownst to me) A is not an aggressor, I would not wrong A in preemptively defensively harming them.

Whether this implication (if it is one) is presented as an objection to evidence-relative views or as a salutary consequence depends on how comfortable one is with its conclusion. I take it that when an agent could exercise responsible control neither over his membership in a group, nor the overall statistical odds of criminal involvement, conditional on that membership, these are the wrong sorts of facts to undermine his moral claim against being intentionally harmed. If so, and if evidence-relative views imply the STATISTICS CLAIM, that would provide strong reason to prefer a different account of rights.[5]

This paper aims to show that the STATISTICS CLAIM is not a genuine implication of evidence-relative theories, and moreover, it is false. Despite the apparent intuitive appeal of the popular argument, demographic statistics cannot do the justificatory work that it assumes they do. I'll start by articulating a common, but flat-footed form of the concern about demographic statistics and outline an obvious response before presenting and replying to a more nuanced version. The replies I offer have a cascading structure: (i) most demographic statistics face an unmet challenge from reference class problems, (ii) of those that can meet that challenge, most ground only minimal conditional probabilities, (iii) those that do more introduce new costs likely to outweigh any justificatory contribution of the statistic, but (iv) even if they are not outweighed, demographic facts are the wrong sort to make the moral difference that needs justifying. To simplify the discussion, I'll focus exclusively on cases where a single agent faces a single possible aggressor, and her de-

[3] van der Vossen (2016, p. 143)

[4] Frowe (2010, 269). Frowe advocates evidence-relativity only for defensive permissions, not moral rights generally, and is emphatic, however, that this sort of permission can come apart from the liability of the other party: if the defender's beliefs are false, then her permissible defense will constitute a justified infringement of the other party's rights against being harmed. So unlike the other views discussed, on Frowe's analysis it does not follow from it being permissible for B to harm A that B would not thereby violate A's rights.

[5] The standard alternative is a kind of fact-relative view on which what agents owe each other is independent of what evidence they have, and (roughly) extensionally equivalent to what an omniscient, morally-motivated third party would advise them to do. On such views, if A is *in fact* not an unjust aggressor, imposing preemptive defensive harm would wrong them no matter how strongly misleading my evidence was. For a defense of fact-relative views of this kind, see McMahan (2005); Otsuka (1994).

fensive options (and perceived threatened harm) are extremely limited: she must either suffer or defensively inflict serious bodily harm or death.

2 Articulating the Objection

The flat-footed objection

The most obvious form of the objection targets accounts that uses a decision-theoretic structure to determine whether an agent is under a duty. Such accounts release an agent from the duty to refrain from an action when, given the comparative moral costs and gains weighted by their probability on the agent's evidence, it is better to do it than to refrain. The high cost of a false-negative defensive error —*failing* to defend oneself against a genuine threat— has the effect of significantly increasing the comparative expected value of imposing defensive harm. If this cost is close to equal to the moral cost of a false-positive error —mistakenly *imposing* unnecessary defensive harm— then even if it's only slightly more likely than not that our agent faces an aggressive threat, DEFEND will have highest expected moral value. So then it might be that, according to evidence-relative theories of rights, knowledge of certain demographic statistics suffices to justify pre-emptive defensive harm against members of those demographic groups.

Justified-belief variants face a similar objection if they allow that what counts as a justified belief depends on decision-theoretic factors. *Pragmatic encroachment* pictures hold that the degree of evidential support necessary for justified belief in, or knowledge of, a proposition P is greater when the stakes are higher. Many advocates of this position clarify that an agent's evidence clears the relevant threshold if she is *practically adequate*: if the action with highest expected value on her current evidence is not different from the action that would have highest expected value were she certain of P.[6]

In cases of self-defense, the at-issue proposition P — whether a stranger (let's call him 'Alan') is an unjust aggressor — is not of merely academic interest to Defender. There is, in some intuitive sense, quite a lot at stake. But the stakes are roughly symmetric, whether she errs on the side of believing P or not. If the contemplated defensive harm is roughly proportionate to the apparently threatened aggressive harm, then the two error possibilities are roughly proportionate in severity. Similarly, if Defender *correctly* assumes P and imposes defensive harm, she avoids suffering the threatened harm, while if she correctly assumes ¬P, she avoids wrongfully imposing harm. These two outcomes also seem roughly symmetric. So the stakes seem balanced: each option has one good and one disastrous outcome, and each is roughly proportionate to the corresponding outcome

[6] Anderson and Hawthorne (2018) interpret pragmatic encroachment as the thesis that S knows that P only if she is 'practically adequate' with respect to P, which they define as follows: "S is practically adequate with respect to P iff the top-ranked element(s) in S's actual preference ranking do not differ from the top-ranked element(s) in her ranking conditional on P." Fantl and McGrath (2002) and Hawthorne and Stanley (2008) can be read as offering variants of this condition for justified belief.

of the other option. But as Schroeder (2012) and Russell (2018) have both noted, when we model stakes decision-theoretically, what matters is the *ratio* between the outcome values, not the absolute values. There's no difference between a case where the cost of errors are low and one where either type of error would be a disaster: the costs cancel each other out. So if the stakes are symmetrical, the evidential requirements for DEFEND to be Defender's morally best bet, or for her to justifiedly believe that it is, will be low, and the relatively weak probabilistic evidence yielded by demographic statistics may suffice.

Now that we've set the stage, here's how the flatfooted version of the *Objection from Statistical Evidence* runs. Assume that our agent starts out neutral about whether a given stranger is a violent aggressor: she initially assigns this a credence of 0.5 (or slightly lower). Now add just two facts to her evidence:

(1) members of demographic group G are *x*% more likely than average to be aggressors in violent crimes.

(2) this particular individual, Alan, is a member of G.

Updating her credence in P on this evidence, our agent should be *x*% more confident of P than she was.[7] Suppose that *x* is a moderately high percentage (perhaps 10-20). Given that the costs of error are roughly equal, this slight difference in probabilities may make assuming that Alan is an aggressor Defender's best prospect, or justify Defender in believing that Alan is an aggressor. If it follows that Alan now lacks a right against suffering defensive harm, the derivation is complete: Defender's knowledge of demographic statistics is sufficient to permit her to preemptively impose defensive harm on Alan.

Answering the flat-footed objection

This version of the worry is relatively easy to rebuff. First, to get the objection off the ground, we've assumed that agents start out roughly as confident that Alan *is* an aggressor as that he is not; but this is radically implausible as a rational prior. Assuming that every new person is as likely to be an unjust aggressor as not fails to take account of base-rates in a way that strongly favors the possibility that they are an aggressor. With the possible exception of active war zones, the vast majority of strangers one encounters are *not* violent aggressors; so rational agents should start far more confident that a given stranger is not an aggressor, and thus with quite a low credence in P. But the lower Defender's prior in P, the more her new evidence must shift the probabilities to justify accepting that P, and the less likely it is that any true demographic statistic will be probabilifying enough to do that.

Second, while the *harms* associated with either type of error in defensive scenarios are roughly equal, to conclude that the errors are therefore *morally* balanced we'd have to deny a principle that

[7] There are at least two problems with this step, one involving the mismatch between the property in the statistical claim and the property of interest to our agent, and one involving an unjustified assumption of probabilistic independence of evidence, but I'll set those aside for now to give the objection as good a run as I can.

many rights-theorists either accept, or do not want to rule out. Namely, that there is an asymmetry between doing and allowing harm, such that, all else equal, it is more difficult to justify imposing unjust harm on someone than merely allowing someone to suffer the same unjust harm. Nor is this a merely marginal difference. On classic articulations of the distinction, the fact that it would save five innocent lives is not sufficient to justify killing one innocent person, so the moral disvalue of killing appears at least several times as difficult to justify as the disvalue of the harm of allowing to die.[8] Since Defender's case is one-to-one, if we accept something like the doing-allowing asymmetry, the risk of unjustly harming Alan (if he's innocent) outweighs the risk of Defender allowing herself to suffer unjust aggression, at least until it is several times more likely that Alan is an aggressor than that he is not.[9]

There is a possible rejoinder here. The asymmetry between doing and allowing is classically understood as showing that negative duties (e.g. duties not to kill) are more stringent than positive duties (e.g. duties to save). But this does not entail that the defender's duty to not kill Alan is stronger than her *permission* to save herself. In non-risky cases, theorists (including Draper, 2005; Montague, 1981; Quong, 2009) have argued that agents enjoy special permissions to preserve their own lives, even if that requires acting in a way that causes the death of a non-threatening innocent. Quong puts it this way: "Morality cannot require you to sacrifice your life for another single person when you rightfully possess the means to save yourself."[10] Could something similar be said for the subjectively risky defensive decisions that are our focus?

It's not obvious that this thought translates. There is a significant difference between requiring Defender to allow herself to be killed and merely requiring Defender to bear some risk of death. It certainly doesn't seem plausible to say that Defender is permitted to preserve herself against a moderate risk of death when that involves running a high risk of *intentionally killing* an innocent non-threat. Suppose that a villain has captured Defender and put her in a room with Alan, announcing that if both are still alive in five minutes, Villain will flip a coin and will kill Defender if the coin lands heads; otherwise he will release them both. Even if it would be permissible for Defender to kill Villain in an attempt to escape, it is surely not permissible for Defender to preemptively kill Alan in order to escape just the 50% risk of death. Nor does this judgment change substantially even if we stipulate that Defender has some evidence—is .6 confident, say—that Alan culpably arranged the whole thing. Perhaps Defender is permitted to assign some extra moral weight to the fact that it's *her* death that is risked if she does not kill Alan. Still, it is difficult to imagine that an agent-relative prerogative would allow Defender to weight her own life as several times more valuable than Alan's, and thus as restoring the moral balance of the error possibilities.

[8] See discussions in Foot (1967), Quinn (1989), and Scanlon (2008, p. 91).

[9] This asymmetry is even more pronounced if Defender chooses to waive her rights against suffering the aggressive harm; she can thereby make the harm she suffers non-wrongful. But she need not be so magnanimous, and, given that we are investigating the conditions of permissible defense against unjust threatened harms, I will assume that she does not waive her rights. (Thanks to Seth Lazar for discussion on this point.)

[10] Quong (2009, p. 518)

So, in cases where the probabilities are close, even if the harms risked are balanced, the expected moral disvalues of the error possibilities are not, and so the evidential threshold for justifying defensive harm remains high. Consequently, the objection from demographic evidence can only get traction in cases where the odds are stacked heavily in favor of the supposition that Alan is an aggressor. And, given that in most contexts the agent's rational priors should be quite low, we can safely say that in realistic cases, generalizations based on demographic groups like age, race, or sex *alone* will not ground probabilities asymmetric enough to permit Defender to impose defensive harm.

The Revised Objection

There are two ways to modify the objection to escape these difficulties. Rather than invoking generalizations over paradigmatic demographic categories like age, race, class, etc., we could instead invoke narrower groups (like an age group *in* a geographic area and social network), or more artificially defined sets (like the group consisting of 99 violent murderers and you), which, given how they are defined, will yield high conditional probabilities for violent aggression.

Alternatively, we could develop the objection as a concern that demographic statistics could simply make a worrisome *contribution* to the justification for imposing preemptive defense. Let's stipulate that, conditionalizing on all her other evidence, Defender is only *just* shy of being justified in believing P (or of preemptively harming Alan being her best prospect). Given this stipulation, small changes in the probability that P could make all the difference to whether she is under a duty to refrain. If her knowledge of the demographic facts (1) and (2) above count as evidence, then even the small additional probabilistic support they furnish may push her past the relevant evidential threshold, making the decisive difference to whether Defender is permitted to assume that Alan is an aggressor. If this happens, then while the statistical probabilities don't by themselves permit others to preemptively harm Alan, they do disproportionately expose Alan—and the other members of G—to risks of preemptive harm, since in contexts that would otherwise be borderline, these statistical probabilities will make it permissible for others to harm a G-member.

These more nuanced versions of the objection threaten both the expected-value and justified-belief variants, regardless of whether they embrace a stakes-sensitive account of justified belief. Answering it adequately will require a serious inquiry into the evidential value of demographic statistics.

3 The Limitations of Demographic Evidence

As an initial reply, advocates of the *justified-belief* variant might take advantage of recent work in epistemology contending that statistical evidence is simply inappropriate grounds for all-out attitudes like belief or knowledge. There are a wide array of such arguments to choose from. To summarize just a few: Jackson (2018) argues that evidence which raises error possibilities to

salience—including evidence that explicitly provides a probability distribution, like statistics—is inappropriate as a basis for belief, though it can be a good basis for credences. Thomson (1986) contends that evidence must be causally connected in the right ways in order to justify belief, and statistical evidence rarely is. Some invoke modal conditions that statistical evidence usually fails: Enoch, Spectre, and Fisher (2012) argue that evidence must be *sensitive* to justify belief (if P weren't the case, we wouldn't have *e*); Pritchard (2017) argues that it must be *safe* (for *e* to justify belief in P, it must not be easily consistent with ¬P). Some others appeal to explanatory power: Smith (2017) argues that to justify belief in P, a body of evidence *e* must 'normically support' P; we would need some additional explanation to render ¬P consistent with *e*. Risinger (2018) similarly argues that *e* justifies belief in P only if we would be surprised, given *e*, to learn that ¬P. Still others (including Armour, 1994; Basu, 2019; Bolinger, 2018; Buchak, 2014; Moss, 2018a) hold that the moral reasons against basing a belief about an individual on statistical data can defeat or undermine the epistemic justification such evidence provides. Along similar lines, Lippert-Rasmussen (2011) contends that individuals have a right to be appraised on the basis of all the relevant, reasonably available evidence, which disallows basing decisions on statistics alone.

Adopting any one of these accounts would permit an advocate of the *justified-belief variant* to say that demographic statistics are never the right sort of evidence to ground justified belief that Alan is an aggressor, and consequently can never render it permissible to impose preemptive defensive harm on Alan. However, if demographic information is permitted to influence one's credences, then we could still worry that it will, *together with the agent's other evidence*, render her justified in believing that Alan is an aggressor. Most of the accounts mentioned above are aimed at explaining the evidential shortfalls of *purely* statistical evidence; not all will equally disqualify demographic evidence from *contributing* to the justification provided by other evidence. To fully secure the justified-belief variant, or to defend the *expected-value* variant of the evidence-relative rights view from the objection on these grounds, we'd need to say that statistical evidence is not appropriate even as a basis for raising one's credence that Alan is an aggressor; that it should simply be disregarded entirely.

One might meet this suggestion with incredulity; surely, if the statistics are skewed enough, then it is not only rational to allow them to inform one's credences, it is *irrational not to!* To give this reply some heft, let's use a concrete illustration, drawn from the demographic group most disproportionately likely to commit violent single-victim crimes. To keep things simple, I'll focus only on crime rates in the United States. The most recent complete statistics released by the FBI Uniform Crime Reporting program (2015) report that close to 89% of single-victim single-perpetrator homicides, and 79% of violent crimes overall, were perpetrated by males, with the overwhelming majority being between the ages of 15 and 34. Suppose a solitary pedestrian knows these statistics, and notices someone following behind her at night. On seeing that he is male, shouldn't she increase her credence his being a violent aggressor accordingly?

There are a lot of things to untangle here. As a start, the case is misdescribed: it's not the fact that the person following her is *male* that should make our agent concerned. After all, if she

noticed a young boy, or a very old man, we wouldn't similarly feel that she's irrational in taking the fact that Alan is male to be irrelevant to whether he is an aggressor. Restating the intuition more carefully, we should say that rational agents should consider the fact that *Alan is an able-bodied adult male* relevant to the probability that he is a threat. Even if we do not reject the use of statistical evidence wholesale, it is easy to over-estimate the support this kind of information can provide. The evidential value of demographic statistics, even for the purposes of setting one's credences rather than full-beliefs, is extremely limited in a few key ways.

Reference Class Challenges

The first limitation is a version of the reference class problem.[11] It isn't epistemically permissible to update one's credences about whether Alan is an aggressor on just any facts of the form '*x*% of members of G are aggressors'. Even if she knows that Alan is a member of a group of which fully half are aggressors, these facts alone do not license her to raise her credence in *Alan* being an aggressor *at all*. Group membership is cheap; it can be stipulated, or based on any property (or set thereof) you like. There are arbitrarily many groups with Alan as a member, with widely varying proportions of aggressors. We could make G the group consisting of 99 murderous assassins and you. Given how we've defined G, it is 99% likely that an arbitrarily selected member of this group is murderous, but we would go wrong to conclude that *you* are 99% likely to be murderous — or even that your membership in G has *any* predictive bearing on whether you're murderous. To be justified in updating our credences based on statistics about one of Alan's group memberships, we need some reason to treat that group as uniquely relevant to the question of whether Alan is an aggressor.

In simple cases, we can justify treating a given class as the relevant one by appealing to our knowledge of background causal forces. For example, as Moss (2018b) explains, we know that if we're trying to set our credence in whether a given growth is cancerous, we should take the class of similarly shaped and colored growths as relevant, rather than the class of other growths discovered on a Tuesday. We can justify this by noting that cancer cells tend to cause certain growth and coloration patterns, and therefore the shape and color are more predictively relevant to the property of interest (whether it is cancerous) than alternative reference classes.

While easy to come by for most banal questions about our environment, this sort of justification is more elusive when the property of interest is something that depends on the exercise of agency.

[11] Reichenbach (1949, p. 374): "If we are asked to find the probability holding for an individual future event, we must first incorporate the case into a suitable reference class. An individual thing or event may be incorporated into many reference classes, from which different probabilities will result. This ambiguity has been called the problem of the reference class." As Hájek (2007) demonstrates, we must face this challenge regardless of which underlying theory of probability we assume. On most theories of probability, the probability of F*a simpliciter* is undefined; to get a determinate value we must *first* select *the single relevant reference class*. So long as we remain ignorant of how G relates to the relevant reference class, we are not justified in assuming that the objective probability of F*a* is anything like the probability of F*a* conditional on *a*'s membership in G.

To justify taking a group G as predictively relevant for an agency-involving property, we'd need reasons to think that being a member of G is intimately connected with choosing to exercise one's agency in that particular way. I think it's too quick to say that *no* group membership can be predictive in this way;[12] in fact, there are at least three reasons for which we could justifiably take membership in G to be predictively relevant to whether a given member will exercise their agency in some particular way f:

(i) G is a biological group, and the feature that is the membership condition for G biologically makes it difficult to resist choosing to f;

(ii) G is a self-selecting group, and the same disposition that leads a person to join G also inclines them to choose to f; or

(iii) G is an identity group, and exerts group-conforming pressure on its members to choose to f.

In the cases that interest us here, the target property—whether Alan is an aggressor—is agency-involving. Some groups that intuitively *are* good reference classes for this property include certain species of aggressive reptiles (biological groups), or, among humans, violent gangs (which are often both self-selecting and an identity group). But the sortals we feel most uncomfortable about using as a basis for predicting properties like violence—e.g. race or ethnicity—are also bad candidates for being relevant groups.

Consider race. As plenty of theorists have pointed out, the folk concept of race and the properties taken to indicate a person's race do not correspond to an actual biological kind, so (i) isn't eligible as a reason to take race as relevant.[13] Racial groups are also not self-selecting, so it's doubly implausible that agents with a predisposition towards violent aggression also tend to choose to be members of some particular racial or ethnic group, so (ii) is out as well. (iii) stands a better chance: to the extent that race is a real sortal, it is principally a social category, exerting pressure on individuals to conform to the practices and values of the group with which they are identified.[14]

[12] Some theorists disagree, holding that it is in fact never permissible to treat agency-involving properties as predictable, because doing so involves taking a 'predictive stance' toward an agent (see especially Basu, 2019). Basu only explicitly discusses the wrongfulness of forming *beliefs* based on statistical evidence, but since using demographic statistics to set one's credences involves taking group membership as predictively relevant, it seems this too violates Basu's prohibition on treating other agents as "something whose behavior is to be predicted", and so would also be impermissible on her account.

[13] On the multiple variances between biological kinds and the folk notion of race, see for example Appiah (1996), Glasgow (2003). Of course, even if race did correspond to some biological kind, we would still lack license to treat it as predictively relevant to violent behavior unless (implausibly) we had evidence that the heritable traits *cause* higher rates of violent aggression.

[14] Whether we should take 'race' to refer to some real property is a matter of significant debate, but many even in the eliminativist camp will allow that there is a socially constructed property of 'racial identity', that can exert pressures on individuals to conform to the cultural expressions of the identity they are taken to have (see, e.g., Appiah, 1996). 'Identifying' with a racial group is plausibly a two-dimensional process: in part it is a matter of the agent herself

This means that race *may* be predictively relevant for agency-involving properties that are an important part of the racial identity. However, unless we are prepared to say that violent aggression is itself part of that identity, we will still lack a reason to take race as a relevant class for predicting whether Alan is a violent aggressor. It is more plausible that gender can be justified as relevant for predicting violent aggression, insofar as certain forms of aggression are socialized as masculine traits.

Generally, unless we can justify taking a group G to be the relevant reference class, we cannot legitimately move from observations about the relative frequency of the property f among Gs to a similar credence that this particular member of G will exhibit f. Without that inference, the demographic probabilities do not justify changing our credences about a particular individual.

Minimal Predictive Value

Second, even when we *can* justify taking a group as a relevant reference class, the base-rates for violent aggression are quite low—most people, most of the time, are not aggressors—so even members of the most disproportionately violent demographic groups are highly unlikely to be violent aggressors. Given the statistics mentioned earlier, conditional on an event being a violent crime in the US, it's highly probable ($\sim$80%) that the perpetrator was a male. But what can we learn about whether someone is likely to be violent, when we learn that they are male? US Census Bureau (2015) population estimates for the same year put the total number of men in the United States at around 158,138,060. If we (improbably) assume that each violent crime in the UCR database had a unique perpetrator, around 954,570 of those men were violent aggressors in the US in 2015. That's roughly equivalent to 0.6%—hardly a high conditional probability. Even though nearly all violent assaults are perpetrated by men, the *overwhelming* majority of men are not violent aggressors. For an agent to increase her credence that Alan is an aggressor by more than about 0.006 on learning that he is male, she would have to irrationally neglect the base-rate of aggression in the group.

So rather than characterize the pedestrian as gaining evidence that Alan is an aggressor, I think the better way to understand the effect of seeing that Alan is an adult male is that the pedestrian hasn't received information that helps *rule out* aggression. When we compare this to her relief at seeing a woman or child instead, the felt difference between dismissing the possibility that Alan is an aggressor and having to keep it open leads us to mischaracterize her as having received evidence that makes it likely that Alan *is* an aggressor. But this sort of information does not actually contribute *positively* to her justification for preemptive defensive harm; it just highlights the absence of information that would undercut such justification. If this is true even for a demographic responsible for close to 80% of violent crimes, it is sure to hold for generalizations across less probabilifying demographic groups as well.

taking the racial identity to be part of her identity, but in part it is a matter of being taken by others to have that racial identity. For thorough discussion, see Gooding-Williams (1998), Glasgow (2006).

There is a third problem. The thought that, absent other evidence, an agent should match her credence in Alan's exhibiting f to the proportion of G-members who exhibit f is reasonably intuitive. But incorporating this sort of information is much more delicate if she already has some other evidence about Alan. To do so responsibly (and avoid double-counting), she needs to be sure that the evidential support given by the group statistic isn't captured by the other facts already in her evidence. To make it clear why, imagine you know that there is a 90% chance that an arbitrary vehicle selected from fleet G will be very safe. You go to the garage to select a vehicle, and see that the fleet is nine minivans and one motorcycle. You know that motorcycles are generally unsafe, but reason that, statistically, since this one is part of fleet G, it is more likely to be safe than an average motorcycle. What goes wrong in this reasoning chain is that the statistic does not provide new information; it only reflects the fact that the nine minivans are safe and the motorcycle is not. Once you've taken into account the vehicle's type, its being a member of fleet G provides no additional evidence about whether it is safe. To be justified in treating a demographic statistic as *additional* evidence, you need a reason to think that it contains relevant information that is probabilistically independent of the evidence you already have. Careful statistical analysis can meet this challenge, but agents making in-the-moment judgments are unlikely to; given that our agent has already taken into account what Alan is wearing, how he is walking, what time and where their encounter is taking place, etc., demographic statistics likely provide no further evidence about whether Alan is an aggressor.

Fourthly, there is a mismatch between the property reflected in the demographic statistics usually invoked (having *ever* been arrested for a violent assault), and the property of interest to our defensive agent (being about to perpetrate a violent assault *now*). Even if she were certain that Alan had previously been arrested, it's unclear how that fact bears on the probability that Alan poses an aggressive threat to her now. The vast majority of even the most ruthless hitman's interactions are not acts of violent aggression, and our agent is seeking to form a credence not about whether Alan has ever been an aggressor, but about whether he is one *in this interaction*. Taken together, these four limitations on evidential value of demographic statistics make it highly unlikely that they can make any significant contribution to an agent's rational credence in whether a given stranger is an aggressor.

The New Costs of Demographic Evidence

But let's imagine that we found some demographic intersection that we could justify taking to be relevant, that yielded non-negligible conditional probabilities for Alan's being a violent aggressor, and was appropriately independent of our other evidence. Could an agent update her credences in accordance with her knowledge of these statistics, and in so doing, come to be justified in imposing preemptive defensive harm? Even if she updates her credences, for most easily recognizable demographic groups, it is unlikely to make a decisive difference to whether she is permitted to defensively harm Alan. This is because generalizations of this kind tend to disproportionately increase the risk exposure of visible group members. Consequently, relying on such generalizations

introduces a new cost to the decision table.

Using identity-tracking heuristics—like generalizations based on race, sex, etc.,—is highly likely to impose disproportionate risk of mistakes in a strongly patterned way. Features like these are highly visible; people display them in most contexts. If they are also a popular heuristic for inferring that P, then simply in virtue of their visible identity, certain people are more likely to suffer a P-based mistake in a wide variety of contexts, made by a variety of agents. Less abstractly: if features of Alan's visible social identity—like his race, gender, and physical build—are a popular heuristic for threateningness, then whether in a hotel, at a grocery store, on university grounds, or while trying to hail a cab, Alan is objectively more likely to be mistakenly perceived to be a threat than those with a different social identity. Facing an increased risk of being mistaken as a violent aggressor is already a significant cost—possibly a significant wrong—because of the severity of the harms one will suffer if the risk eventuates. But *patterned* risk exposure exacerbates this, transforming even minor risks into forces that shape and restrict Alan's opportunities and alternatives in an unjustified way on the basis of his social identity. These follow-on costs magnify the wrong done in contributing to the patterned risk imposition, and are weighty reasons against adopting epistemic practices that would make such a contribution.[15]

Whether or not Alan actually is an aggressor, if Defender reaches that conclusion partially on the basis of a generalization about the threateningness of members of *G*, Defender participates in the imposition of patterned risk of mistake on members of the demographic group *qua* group members. She not only risks having unnecessarily harmed someone; she risks having harmed him *because* he was a *G*. This wrong is more severe the more the heuristic increases the group's disproportionate risk exposure. If the conditional probability of P, given that Alan is a member of G, is relatively low, then the heuristic won't expose members of G to noticeably increased risks, but it also won't noticeably raise Defender's credence that Alan is an aggressor. But as the conditional probability rises, innocent members of G are exposed to increasingly disproportionate risk of suffering mistaken harm, and so have an increasingly serious complaint against other agents taking their visible membership in G as evidence that they are likely an aggressor.[16]

In short, the more a demographic generalization increases one's credence that P, the more it

[15] In Bolinger (2018), I appeal to these kinds of reasons, and in particular the way that heuristics pattern and amplify risk exposure, to explain why inferences from visible social-group membership to stereotype-congruent conclusions are frequently rationally and morally impermissible.

[16] In more technical terms, if agents rely on demographic statistics about G in determining their epistemic risk of error, they will more readily act preemptively against members of G, all else equal, in proportion to the strength of the conditional probability. The more widespread this attitude is, and the more visible and stable membership in G is, the more being a group member increases A's objective *ex ante* risk of being mistaken for an aggressor. When these effects are especially stark (e.g. when being black doubles one's risk of being mistakenly killed by police (Swaine, Laughland, & Lartey, 2015)), A's knowing that his group membership puts him at increased risk may trigger additional harms: it may take a psychological toll, and he may feel forced to engage in costly measures to unilaterally reduce his risk exposure (e.g. avoiding contact with police or legal institutions, adopting mannerisms and dress aimed at appearing non-threatening, etc.).

increases Alan's exposure to serious risks and harms *simply on the basis of his group-membership*, and thus the greater the wrong Defender would participate in by relying on it. Appealing to demographic statistics may raise the probability of P on her evidence, but it also adds 'imposing disproportionate risk of suffering mistaken harm on Alan, on the basis of his being in G' to the costs. If membership in G is not an appropriate basis for the increased risk, this imposition is wrongful. So, if the decision table was otherwise balanced, we should expect it to at most remain balanced: whatever contribution the marginal increase in probability makes to the expected value will be matched or outweighed by the additional wrongs involved.

4 When relying on demographic probabilities can be appropriate

Of course, just because identity-tracking inferences based on race, ethnicity, or skin tone contribute to structural injustice by imposing wrongful risks, it doesn't follow that all inferences about threateningness based on demographic statistics do so. There are two kinds of demographic inference we might think to be exempt from this problem: (a) sufficiently narrowly defined intersectional demographic categories, or (b) inferences based on self-selecting violent groups like gangs.

The wrong sorts of reasons

Narrowly defined groups could yield non-negligible probabilities. However, even assuming that the membership property is defined narrowly enough to affect relatively few people (just the able-bodied young white men in a particular district, say), it doesn't seem that properties over which he has no agential control should be the difference-maker in whether Alan loses his right against harm. So, particularly if agents do not have control over whether they bear the membership-determining property—and thus count as a member of the demographic group—there is something morally objectionable about allowing demographic statistics for the group to play a decisive role in determining whether they retain rights against being harmed.

It is less clear whether the individual has any complaint against others' merely using statistics to inform their credences about whether he is a threat. *If* the conditional probabilities arising from narrowly-defined groups are high enough, we may think they contribute to making defense *rational* without equally contributing to stripping Alan of rights against suffering that harm, if in fact he is not aggressive. This position does not cohere well with the evidence-relative position that what duties B owes to A is a function of B's evidence, rather than A's interests or exercise of agency. So if A's agency seems decisively important here, that is some reason to prefer an alternative to evidence-relative views of moral rights.

The Difference Behaviors Make

Gangs are illuminatingly different from the other groups we have discussed. When a group is self-selecting, Alan can choose whether to be a member of the group or not. The more voluntary membership is, and the easier it is for Alan to exit the group, the less we need worry that conditionalizing on group membership will let Alan's rights be determined by factors outside his control. And the more directly connected the group is with violent activity, the better a claim it has to being predictively relevant to whether group-members will engage in violent aggression. So if there are any cases where Defender's knowledge of the statistics about a group contributes substantially to the permissibility of preemptive defensive harm, they will involve groups like these, where group membership is voluntary and signalled by things Alan has agential control over (e.g. behaviors and attire) rather than by identity-tracking properties like race, gender, or class.

I contend that this is in part because they clear the evidential hurdles that disqualify the others, and in part because signalling membership in a violent group is a communicative behavior that can undermine Alan's standing to complain against defensive harm. There are some familiar ways that Alan can act to cancel his complaint against a treatment that harms him: he can consent to the treatment, he can waive his right against it, he can incur a liability to it, etc. But he can also undermine his standing to complain in a way that doesn't fit neatly into any of these actions, by being the one who invites the treatment, as in the following case.

> CAR: We have a standing agreement that you may borrow my car to run errands when I don't need it. On Tuesday you ask whether I need the car, and I say that I don't. My testimony leaves you less than certain, but you trust me and borrow it. In fact, I needed it to get to a job interview.

In CAR, I can't complain that, since the error is costly for me, you shouldn't have believed on the basis of the available evidence that I didn't need the car. The reason seems to be that through my testimony, I gave up standing to complain against your accepting this proposition. To generalize, we can say that when Alan intentionally performs a communicative act with the conventional meaning that P, Alan cannot complain against his addressee accepting that P. So, in defensive contexts, when Alan intentionally signals that he is a member of a group whose membership principle is violent aggression, he undermines his complaint against being taken to be a violent threat.

If this is true, then observing signalled evidence of violent gang-membership can have a double-effect on the decision table: *qua evidence*, it increases the probability that Alan is a violent aggressor; *qua communicative behavior*, it lessens Alan's complaint against the risk of mistake, and so reduces the costs associated with false-positive error. As a consequence, even when this kind of evidence yields the same (or less) probabilistic support for as evidence from other sources, it may have greater justificatory effect. Assuredly there are some moral limits on what sorts of behavioral choices can be treated as complaint-undermining, and we'll need to do some significant moral theorizing.[17] But, provided we can ground some criteria to distinguish which signals can do the

[17] I've outlined the beginning of such an account elsewhere (see Bolinger, 2017).

moral work, we have the resources to give a principled explanation of why probabilities based in Alan's gang membership can, but based in Alan's other demographic groups cannot, contribute to justifying preemptive defensive harm.

5 Conclusion

The initial problem I set out to discuss was whether evidence-relative theories of rights imply that demographic statistics—either on their own or together with some otherwise inadequate evidence—can justify imposing preemptive defensive harm. In §2, I noted that because imposing unjustified harm is morally worse than allowing oneself to suffer unjustified harm, it is implausible that any realistic demographic statistics would yield probabilities high enough to *suffice* to make imposing such harm the agent's moral 'best bet', or justify her in flat-out believing that she must defend herself. So the real challenge comes not from this sufficiency worry, but from the more nuanced concern that demographic information could make a decisive *contribution* to whether an agent has a duty not to impose harm on a particular group member, Alan.

In §3 I turned to examining the relevant limitations of evidence that is based in statistical generalizations. While there are many accounts of the inadequacies of *purely* statistical evidence on offer, they typically do not indict its ability to merely contribute to the justification provided by other evidence, or to provide grounds for rational credence. So evidence-relative views of rights cannot answer the contribution objection simply by pointing to general faults with statistical evidence. Still, there is plenty to be said. First, most generalizations over demographic groups face an unmet reference class challenge: before we can permissibly base our credence in whether Alan is an aggressor on the rate of aggressors in a demographic group to which he belongs, we must justify taking that group to be predictively relevant, and this is no small task.

Second, even if we can justify the relevance of a group, most demographic groups have an incredibly low baserate for violent aggression, and so ground only negligible probabilities that their members will be violent aggressors. So they will not make a non-negligible difference to whether an agent has a duty to not impose harm on Alan. And in the event that a relevant group successfully grounds a non-negligible probability, there is a third problem. Relying on visible identity-tracking demographic characteristics in order to make determinations about whether someone is a threat introduces new costs to the decision table, which are likely to at least offset the evidential contribution of the statistic. The fourth problem, raised in §4, is that when Alan can exercise control neither over his membership in a group, nor over the statistical distribution of violent aggression within the group, these are the wrong sorts of facts to undermine his right against being intentionally harmed, even when they are relevant to whether it is *rational* for an agent to impose defensive harm.

These arguments threatened to prove too much: they seemed to suggest that information about the rates of violent aggression among groups to which Alan belongs can never help justify the belief that Alan is an aggressor. But surely this is wrong; if I notice that Alan sports the armband or

jacket-patch of a violent gang, it seems that this information should contribute to the justification of just such a belief. A closer look at this sort of case revealed that when displaying the property indicative of group-membership is sufficiently within Alan's agential control, the display undermines his complaint against others coming to believe that the group is predictively relevant. When the group's organizing principle is violent aggression, such a display can be powerful justification for believing that Alan is likely to engage in violent aggression. But the justificatory power of group-based evidence diminishes in proportion to its distance from Alan's responsible action. The less direct the connection to Alan's agency, the weaker the evidence is, until (at the far extreme) relying on it *introduces* new harms by exposing Alan to unjustly disproportionate risk of P-based error.

All told, evidence-relative theories face no special problem from the objection from demographic evidence. There are adequate epistemic resources to draw on to capture the appropriate verdicts, and block the troublesome implications. Consequently, the popular argument, claiming that demographic statistics justify or fully excuse agents who make defensive mistakes against members of particular social groups, should be rejected.

Acknowledgements

Thanks to Rima Basu, Hannah Bourandt, Maegan Fairchild, Greg Keating, Seth Lazar, Kirsten Mann, Jonathan Quong, Mark Schroeder, Brian Weatherson, and James Willoughby, as well as the Moral Philosophy and Social Theory working group at Australian National University for helpful discussion on many of the questions covered in this paper.

References

Anderson, C., & Hawthorne, J. (2018). Knowledge, practical adequacy, and stakes. In *Oxford studies in epistemology* (Vol. 6). Oxford: Oxford University Press.

Appiah, K. A. (1996). Race, culture, identity: Misunderstood connections. In *The tanner lectures on human values* (Vol. 17, p. 53-136). University of Utah Press.

Armour, J. (1994). Race ipsa loquitur: of reasonable racists, intelligent bayesians, and involuntary negrophobes. *Stanford Law Review*, *46*(4), 781-816.

Basu, R. (2019). What we epistemically owe to each other. *Philosophical Studies*, *176*(4), 915-931.

Bolinger, R. J. (2017). Reasonable mistakes and regulative norms: Racial bias in defensive harm. *Journal of Political Philosophy*, *25*(2), 196-217.

Bolinger, R. J. (2018). The rational impermissibility of accepting (some) racial generalizations. *Synthese*.

Buchak, L. (2014). Belief, credence and norms. *Philosophical Studies*, *169*, 285-311.

Draper, K. (2005). Rights and the doctrine of doing and allowing. *Philosophy and Public Affairs*, *33*(3), 253-80.

Enoch, D., Spectre, L., & Fisher, T. (2012). Statistical evidence, sensitivity, and the legal value of knowledge. *Philosophy and Public Affairs*, *40*(3), 197-224.

Fantl, J., & McGrath, M. (2002). Evidence, pragmatics, and justification. *Philosophical Review*, *111*(1), 67-94.

FBI Uniform Crime Reporting program. (2015). *Crime in the United States.* Retrieved June 2017, from https://ucr.fbi.gov/crime-in-the-u.s/2015/crime-in-the-u.s.-2015

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5-15.

Frowe, H. (2010). A practical acount of self-defence. *Law and Philosophy*, *29*(3), 245-272.

Glasgow, J. (2003). On the new biology of race. *Journal of Philosophy*, *100*(9), 456-74.

Glasgow, J. (2006). A third way in the race debate. *Journal of Political Philosophy*, *14*(4), 163-185.

Gooding-Williams, R. (1998). Race, multiculturalism and democracy. *Constellations*, *5*(1), 18–41.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, *156*, 563-585.

Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy*, *105*(10), 571-90.

Jackson, E. (2018). Belief, credence, and evidence. *Synthese*. Retrieved from https://doi.org/10.1007/s11229-018-01965-1

Leslie, S.-J. (2017). The original sin of cognition: Fear, prejudice and generalization. *Journal of Philosophy*, *114*(8), 393-421.

Lippert-Rasmussen, K. (2011). 'we are all different': Statistical discrimination and the right to be treated as an individual. *Journal of Ethics*, *15*, 47-59.

McMahan, J. (2005). The basis of moral liability to defensive killing. *Philosophical Issues*, *15*(1), 386-405.

Montague, P. (1981). Self-defense and choosing among lives. *Philosophical Studies*, *40*, 207-219.

Moss, S. (2018a). Moral encroachment. *Proceedings of the Aristotelian Society*, *118*(2), 177-205.

Moss, S. (2018b). *Probabilistic knowledge*. Oxford University Press.

Otsuka, M. (1994). Killing the innocent in self-defense. *Philosophy and Public Affairs*, *23*(1), 74-94.

Pritchard, D. (2017). Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence*, *doi: 10.1080/20403313.2017.1352323*, 1-12.

Quinn, W. (1989). Actions, intentions, and consequences: The doctrine of doing and allowing. *The Philosophical Review*, *98*(3), 287-312.

Quong, J. (2009). Killing in self-defense. *Ethics*, *119*, 507-537.

Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California University Press.

Risinger, D. M. (2018). Leveraging surprise: What standards of proof imply that we want from jurors, and what we should say to them to get it. *Seton Hall Law Review*.

Russell, J. (2018). How much is at stake for the pragmatic encroacher. In *Oxford studies in epistemology*. Oxford: Oxford University Press.

Scanlon, T. (2008). *Moral dimensions*. Cambridge University Press.

Schroeder, M. (2012). Stakes, withholding, and pragmatic encroachment. *Philosophical Studies*, *160*, 265-285.

Smith, M. (2017). *Between probability and certainty: What justifies belief*. Oxford University Press.

Swaine, J., Laughland, O., & Lartey, J. (2015). The counted: Black americans killed by police twice as likely to be unarmed as white people. *The Guardian*.

Thomson, J. J. (1986). Liability and individualized evidence. *Law and Contemporary Problems*, *49*(3), 199-219.

US Census Bureau. (2015). Population estimates for 2015. Retrieved June 2017, from https://www.census.gov/quickfacts/table/PST045216/00

van der Vossen, B. (2016). Uncertain rights against defense. *Social Philosophy and Policy*, *32*(2), 129-145.

Zimmerman, M. (2008). *Living with uncertainty*. Cambridge University Press.

Zimmerman, M. (2014). *Ignorance and moral obligation*. Oxford University Press.

Zimmerman, M. (2018). In defense of prospectivism about moral obligation: A reply to my meticulous critics. *Journal of Moral Philosophy*, 444-461.