Synthese
https://doi.org/10.1007/s11229-018-1809-5

The rational impermissibility of accepting (some) racial generalizations

Renée Jorgensen Bolinger[1]

Received: 17 March 2018 / Accepted: 8 May 2018

Abstract I argue that inferences from highly probabilifying racial generalizations (e.g. believing that Jones is a janitor, on the grounds that most Salvadoreans at the school are janitors) are not solely objectionable because *acting* on such inferences would be problematic, or they violate a moral norm, but because they violate a distinctively epistemic norm. They involve accepting a proposition when, given the costs of a mistake, one is not adequately justified in doing so. First I sketch an account of the nature of adequate justification—practical adequacy with respect to eliminating the $\neg p$ possibilities from one's epistemic statespace. Second, I argue that inferences based on demographic generalizations tend to disproportionately expose group members to the risks associated with mistakenly assuming stereotypical propositions, and so magnify the wrong involved in relying on such inferences without adequate justification.

Keywords Epistemology · Acceptance · Generalizations · Statistical evidence · Moral encroachment · Epistemic risk

1 Introduction

There are cases—some familiar from debates over racial profiling, others from debates over naked statistical evidence—in which even though an agent's evidence makes a proposition p very probable, it seems she still shouldn't believe that p. This holds

Work on this paper was supported by ARC Grant D170101394.

✉ Renée Jorgensen Bolinger
renee.bolinger@anu.edu.au

[1] School of Philosophy, Australian National University, HC Coombs Building, Canberra 0200, Australia

Published online: 23 May 2018

even when the evidence seems to justify a credence in p that would ordinarily suffice for rational belief. Gendler (2011) offers one such case:

> **Cosmos Club**: The night before he is to be presented with the Presidential Medal of Freedom, John Hope Franklin hosts a celebratory dinner party at the Cosmos Club. All the other black men in the club are uniformed attendants. While walking through the club, a woman sees him, calls him over, presents her coat check ticket and orders him to bring her coat.

It seems the woman shouldn't have assumed that John Hope Franklin was an attendant, but as Gendler notes, it isn't obvious what the nature of her error is:

> Franklin had been the Cosmos Club's first black member, and was still one of very few. By contrast, nearly all of the club's numerous attendants were men of African descent. So when the woman was presented with the visual experience of a black man in the club's lobby, she endorsed an empirically well-supported hypothesis—one that took full account of the prior probabilities. The likelihood that a black man present in the cosmos club was a member of the staff rather than a member of the club was very high–high enough, perhaps, to make it rational to assume that even though he was wearing a suit rather than a uniform, he was nonetheless an employee rather than a host.[1]

There are several ways to gloss the case. We could say the problem is purely practical: that while she (for epistemic reasons) ought to have made the assumption, she (for reasons of morality or politeness) ought not to have acted on it. Or purely moral: that while epistemically permitted, the woman morally ought not have even believed that Franklin was staff. Gendler's discussion of the case suggests one of these two readings; on either, the woman's assumption violates no epistemic norms. Recently, a number of theorists have argued in favor of a third option: the assumption is in fact not rationally permissible. Some locate the fault in the type of evidence provided by statistical generalizations (or specifically demographic generalizations), while others argue that the moral reasons against belief undermine the epistemic permission otherwise afforded by the woman's evidence.[2]

Each of these approaches is insightful, but leaves out important parts of the explanatory story, and so either over or under-generates epistemic permissions. This paper aims to provide an analysis of the problem in COSMOS and similar cases that delivers an intuitively appealing explanation of the fault, and generates a plausible pattern of epistemic permissions and obligations. I will argue that assumptions like the one in COSMOS are rationally impermissible because they involve accepting a proposition without adequate justification, given the wrongs risked in the event of error. My analysis has two separable but complimentary parts: (i) an account of the requirements

[1] Gendler (2011), p. 3. The case is taken from Franklin (2005)'s memoir, *Mirror to America* (p. 340). Notice that we can make the evidential likelihood that Franklin is staff arbitrarily high by increasing the number of black staff members. If your first reaction to this case is to think that the woman doesn't have good evidence, it cannot be because she lacks information on which it is highly probable that Franklin is staff.

[2] For the first, see Armour (1994), Enoch et al. (2012), Buchak (2014) and Moss (2018) for a sampling; for the latter, see Fritz (2017), Basu and Schroeder (2018) and Basu (2018).

for rational acceptance, and (ii) an argument that inferences like the one in COSMOS impose disproportionate risk of certain types of error, which compound into wrongs significant enough to make the inference fail the requirements for rational acceptance.

The first step in this project is to extricate ourselves from an overly simple view of rational norms. On the simple picture, a belief-like attitude has an epistemic fault only if it violates one of the norms on purely theoretic inquiry. Any other flaws aren't with the attitude, but are only problems with *actions* the agent would subsequently be inclined to perform. This caricature neglects the interface between agents' epistemic and practical lives: the epistemic attitudes that frame practical deliberation, setting which possibilities they will treat as 'live' for the purposes of inquiry, and which propositions they will take for granted, or *accept*. When an agent accepts p, she uses p to frame her deliberative statespace: she dismisses the possibility that $\neg p$ from consideration, and is disposed to rely on p in her practical reasoning.

Acceptance is neither a purely theoretic nor purely practical attitude. It is governed by epistemic norms and sensitive to practical costs, aiming to be efficiently accurate: true enough, given what's at stake and the costs of increased accuracy. Nonetheless it remains distinctively epistemic: acceptance is an evidence-responsive cognitive attitude toward propositions, which interacts systematically with paradigmatically epistemic activities like evidence gathering, conditionalization, and inference. An agent cannot rationally accept something she takes to be false, because she can never rationally disregard what she takes to characterize the actual world.[3] It is also practically-oriented: insofar as an agent's outright beliefs (as opposed to her credences) influence her practical deliberation, they do so through acceptance. I am not particularly concerned with whether *outright believing that* p and *accepting that* p are ultimately the same or different attitudes.[4] I aim simply to demonstrate that there are problems with adopting this attitude in cases like COSMOS, and to reject the suggestion that these problems are merely practical, arising only when agents subsequently act on the attitude.

[3] 'Acceptance' as I am using it is thus broadly like the notions defended by Cohen (1989) and Bratman (1992), namely an attitude of 'taking for granted' to frame deliberation that is incompatible with thinking p to be false. It is importantly unlike Stalnaker (1984)'s notion of acceptance, which resembles supposing or pretending in allowing that an agent may accept p in one domain while simultaneously rejecting it in another; it is also unlike the notion in Engel (1998) which explicitly aims only at utility and permits agents to accept propositions they take to be false. My thanks to Julianne Chung for encouraging me to explicitly characterize the attitude.

[4] 'Belief' names two different attitudes: degrees of belief, or *credences*, are the epistemic probability that p for S, while all-out or *full belief* is an all or nothing attitude of confidence in the proposition. Since acceptance is a binary attitude, it should be distinguished sharply from the degreed notion; in what follows, I'll use 'credences' to refer to degrees and reserve 'belief' for full belief. Whether acceptance in my sense amounts to belief depends on the theoretical role assigned to full belief. If we characterize the attitude as something like a disposition to 'treat as true', to 'premise', or to 'take for granted' [as Ross and Schroeder (2014) do], then acceptance and belief appear to be the same attitude. Alternatively if we insist on distinguishing premising from 'feeling to be true' or credence 1, and take only the latter to be belief [as Cohen (1992) does], then acceptance cannot be the same attitude as belief. But insofar as the two differ, it is acceptance, rather than belief, that implies willingness to act as if p, and so it is acceptance that, together with her credences, guides an agent's action in a context.

2 The generality of the puzzle

Focusing on a single case can be hazardous for theorizing. To help fill out the profile of the puzzle, it will be useful to consider two additional cases, analogous to COSMOS in inappropriateness, but occurring in a slightly different context[5]:

> *Gatecrashers* Only 25 fares to an event are sold, but 500 people are observed in the stands. Based just on this evidence, it's .95 likely that a given event attendee, Alfred, failed to pay his fare. Tickets were not distributed and fares paid in cash, so no further evidence is available. On the standards of proof for tort liability, it need only be likely "on the balance of probabilities" that the agent incurred the liability, but it does not seem permissible to rule based on the statistical evidence alone that Alfred is liable to be fined for failure to pay his fare, despite the fact that it easily clears the probability threshold.
>
> *Buses* The Blue Bus co. (which operates all and only blue buses) owns 90% of the buses in town; the Green Bus Co. operates the other 10%. Jones is injured by a bus. It seems impermissible to rule that the Blue Bus co. is liable based just on the statistical evidence, despite the fact that it's .9 likely that the bus that injured Jones was owned by the Blue Bus co. By contrast, if the two companies operated equal proportions of buses, but an eyewitness whom we know to be 80% reliable testifies that the bus was blue, it seems that it would be permissible to rule that Blue Bus Co. is liable on the basis of the testimony, even though conditional on *that* evidence it's only .8 likely that Blue Bus Co. was at fault.

Both of these cases are like COSMOS in that the deliberating agent has evidence (a statistical generalization about a relevant group) on which it is highly probable that p, but the evidence seems inadequate to permit her to accept that p. Of course since legal rulings are official, public findings of fact, it is likely that whether it is permissible to conclude that p in legal cases is subject not only to rational and moral constraints on acceptance, but also to the stronger norms governing authoritative assertion.[6] Still, ruling that the plaintiff is liable would, in these two cases, be an institutional parallel to an individual's accepting p: it involves taking one's current evidence to establish p adequately enough to take p for granted in framing the subsequent p-dependent decisions one expects to face (in these cases, decisions about compensation). Insofar as it is an institutional correlate, the deliberative norms governing the institutional decision space are plausibly similar, and helpfully illuminate the rational constraints on individual deliberations.

In GATECRASHERS, the reason we shouldn't base a liability judgment on the highly probabilifying statistical evidence seems to be that doing so runs an unacceptable risk of wronging Alfred, if in fact he paid his fare. Something similar can be said of BUSES, though in that case what makes the risk unacceptable is clearly not—or not

[5] GATECRASHERS, also known as the 'paradox of the rodeo' is based on hypothetical given by Cohen (1977), *The Probable and the Provable*, pp. 70–83. The BUSES case is based on *Smith v. Rapid Transit, Inc.* (317 Mass. 469, 58 N.E.2d 754 (1945).

[6] In her analysis of the problem of statistical evidence, Thomson (1986) says something along these lines, suggesting that legal findings are subject to a norm like 'Assert p only if the expected costs to the hearer of relying on p are low enough.'

merely—that a mistake is too *probable*, otherwise the eyewitness testimony would also be inadequate to justify holding the Blue Bus Co. liable. The core problem doesn't seem to be with the type of property grounding the generalization, either. Though in COSMOS it's a deeply personal, stable sortal property (membership in a racial group), in GATECRASHERS it's a highly accidental, merely situational feature (presence in a crowd).

You might think that these cases show that agents should simply never accept p unless their evidence completely rules out $\neg p$. This thought is encouraged by common models of deliberation for ideal agents, but we should reject it as a standard for agents like us. Certainty is hard to get, and our cognitive resources are limited. A cognitively bounded agent cannot attend to a complete partition of logical space; to give adequate attention to the most relevant possibilities, she must allow herself to simplify the decision space, making do with rough probability estimates, tolerating rounding errors, and completely ignoring quite a lot of low-probability possibilities. In deciding whether to stop for groceries on her way home, such an agent should simply take for granted a wide array of propositions that frame her deliberation. She shouldn't consider the possibility that the store has suddenly stopped selling food, or burned down, or even that it might have changed its hours—unless a lot depends on getting this particular fact right. In other words, rationality requires cognitively bounded agents to simply accept many propositions that fall short of certainty.

What sort of justification does an agent need in order to rationally accept a proposition, then? Accepting p involves deciding to move from an epistemic partition with some $\neg p$ regions of >0 probability to one without, giving $\neg p$ no weight in future deliberation.[7] S can rationally do this only if her current evidence is sufficient to justify wholly ignoring the remaining chance that $\neg p$. I suggest that there are two ways her evidence could manage this: it could either (1) yield a probability for $\neg p$ that is low enough, given the costs of a mistake, that it makes no difference to how it is rational for S to act (rendering S *practically adequate* in assuming p), or (2) directly license her to discard $\neg p$, presuming that she is in a p-world absent special undermining evidence.[8] I'll devote the next section to filling in the details of this proposal, and the

[7] We can model this either as contracting the epistemic space Ω by removing all possibilities incompatible with p, or as setting the probability of those possibilities to 0. The differences between the two will likely have no effect for our purposes, but the former is more realistic as a model of how actual agents allocate cognitive space. The models diverge where there are uncountably infinite alternatives or in a system that doesn't validate countable additivity. One cannot have positive credence in a disjunction all disjuncts of which have been excluded from the epistemic space (the disjunction will be undefined), but one can when all the disjuncts are just set to 0 [see Arntzenius et al. (2004)].

[8] A prominent view in contemporary epistemology holds that in general, pragmatic concerns about the potential costs of error and the benefits of success *encroach* on knowledge: it is harder to know or have adequate evidence to justify believing p in 'high stakes' contexts. Regardless of its merits as a view about knowledge, something like this picture is attractive for acceptance. Beyond concerning acceptance rather than knowledge, the account I offer departs from standard articulations of pragmatic encroachment in two key ways. First, standard pragmatic encroachment proposals remain thoroughly probabilistic: the effect of high stakes is just to raise the minimum threshold of how probable the evidence must make p in order to justify accepting p. I advocate modeling the demand not as a difference in probability threshold, but strength: you need evidence that can function to *dismiss* a salient error possibility, and even extremely high probability generalizations often can't play this role. Second, I replace the somewhat mysterious talk of 'stakes' with the comparatively tractable notion 'practical adequacy.' Worsnip (2015) and Anderson

following one to demonstrating that in COMOS and many cases where the evidence is a racial generalization, it will fail to meet either standard.

3 Practical adequacy as a constraint on rational acceptance

To make things more precise, we can put an agent's available actions in a context into a partial ordering, ranked by their probability-weighted value on the agent's current evidence e.[9] Call this S's e-based ranking. If we then update e by changing the probability of p to 1 (and adjusting accordingly) and generate a new ranking, we'll get a model of what it'd be most rational for S to do if she were certain of p. If the top-ranked actions in the e-based ranking do not differ from the top-ranked action in the updated ranking, then S is *practically adequate* with respect to p: the difference between her current evidential situation and certainty in p makes no difference to how it is rationally best for her to act.[10]

There are three important caveats in how this calculus works. First, we have to be a little careful when characterizing the value of S's available actions. Since we're modeling the factors relevant to the *epistemic* rationality of accepting p, the relevant costs and payoffs are only those associated with the characteristic role of *full confidence that* p. If you know that given that p it is best to ϕ, and you are confident that p, then if your confidence plays its characteristic role you will ϕ. Consequently, costs associated with ϕing when p is false count as reasons against accepting p. Costs or payoffs *not* associated with the characteristic role—such as threats or bribes from an external agent conditional on your coming to accept that p—do not count as reasons to accept p. Second, the relevant costs are not limited to S's own prudential concerns. While the value of some elements of the ranking are up to the agent, many are not. The moral disvalue of violating another's stringent rights, for instance, counts against an action

Footnote 8 continued
and Hawthorne (forthcoming) show that it is non-trivial to define a decision-theoretic notion of stakes that will play the role needed to underwrite the classic judgments of pragmatic encroachment. Russell (forthcoming) suggests that while we *can* define a consistent notion, it will be sensitive only to ratios, not to raw magnitudes of outcomes and thus will diverge at crucial points from the rough intuitive notion. Practical adequacy escapes these difficulties.

[9] On this sort of ranking, the value of an action is determined by the sum of the values of each possible outcome weighted by its probability on S's evidence. Actions that are top-most in the ranking count as 'best', though because a partial ordering allows for ties and permits some alternatives to be incommensurable, there can be multiple 'best' actions. The rankings may be made sensitive to rights and other classically deontological concerns by invoking a dual-ranking evaluative scale [like the one proposed by Portmore (2008)], and can allow agents to discount costs to themselves, or allow personal costs to weigh against supererogatory moral values, if desired.

[10] Roughly this notion of practical adequacy is articulated in Anderson and Hawthorne (forthcoming), who offer the following definition: "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." Locke (2013)'s notion of 'practical certainty' is also very similar, but he offers it as a necessary and sufficient condition for rational acceptance (in his terminology, 'premising'), while I am proposing it only as a necessary condition. The standard can be made more stringent, if desired, by requiring minimal divergence in the n top-most actions.

even if S doesn't care about the violation. The moral disvalue of outcomes is relevant to their position in the ranking, and so too to whether S is practically adequate.

Finally, there is an important temporal element to rational acceptance that is obscured by common models of rational action or belief. When we ask what it is rational for an agent to *do*, we often focus on a particular, immediate decision she must make (like getting on a train or stopping at a bank). This misleadingly suggests that the costs and payoffs relevant to the rationality of acceptance are just those attached to the agent's immediate next decision. But accepting p has a variety of downstream effects on S's behavior as an epistemic agent: she'll consider the question whether p closed, stop being attentive for further evidence whether p, and be disposed to act as if certain in p. Accepting p will affect how S acts in more than the immediate future, so the consequences of future p-dependent decisions are relevant to whether it is rational to accept. At the other extreme, on many standard models of belief revision, becoming certain of some proposition is irreversible. Were we to use such a model for rational acceptance, we'd consider the consequences of assuming p for *all* future decisions the agent will ever face relevant to whether she is practically adequate in accepting. This, though, is too broad; acceptance is not irreversible for actual agents. It's still possible, given sufficiently striking evidence for $\neg p$, for an agent S to re-introduce non-zero probabilities for $\neg p$, but a much smaller pool of possible evidence is capable of making S attend to these possibilities after accepting p than before.

The strength of the practical adequacy requirement lies somewhere between these extremes. Accepting p will have effects on her action beyond the immediate context, and so the practical adequacy requirement must extend beyond the immediate context. But it need not extend into the whole future: she need only be adequate until she expects to receive compelling $\neg p$ evidence, if $\neg p$. Whether S may accept that p thus partially depends on how confident she can be that if she's mistaken, she'll receive compelling evidence for $\neg p$ before incurring costs due to having assumed that p.

While the practical adequacy constraint does the lion's share of the explanatory work, to complete the account we'll need to invoke a notion of evidential strength to determine when evidence is strong enough to license S to directly dismiss $\neg p$ as a salient possibility. Handily, there are many available, and my account imposes only the constraint that the chosen notion of strength not reduce to probabilistic support. It pairs naturally with a normic support model, or a modal conception like safety or sensitivity, on which e is stronger evidence for p the greater the sphere of worlds in which if S has e, p is true (for safety), or in which if p is false, S doesn't have e (for sensitivity).[11] Evidence that is sufficiently strong (in the chosen sense) licenses S to presume that the actual world is a p-world, and directly discard $\neg p$ possibilities, because given that S has this evidence, she cannot easily be mistaken whether p.

[11] Enoch et al. (2012) favor sensitivity; Pritchard (2017) advocates safety. Smith (2010) advocates Normic Support, on which e is strong evidence for p if p is more normal, given e, and we require an extra explanation to make e compatible with $\neg p$. [Gardiner (2018) also favors a variant of normic support, while Peet and Pitcovski (2018) defend something similar.] Haack (2012) offers a conception grounded in explanatory power and coherence. On any of these, evidence like eyewitness testimony will be stronger than even high-probability generalizations, since the latter are easily consistent with $\neg p$ even at the actual world.

Holding the costs fixed, S can be justified in accepting p on a body of evidence $e1$, and not on other evidence $e2$, if $e1$ but not $e2$ is strong enough to license S in dismissing $\neg p$ from salience, even if the probability of p conditional on $e2$ is higher than conditional on $e1$. Though highly probabilifying generalizations do, when properly defined, provide S with reason to assign a very low probability to $\neg p$, they don't directly license her to discard $\neg p$. When S's evidence consists only in such a generalization, she's justified in accepting p only if the probabilities yielded are asymmetric enough to render her practically adequate, given the costs. When the costs of leaving it open whether p are low, and the costs of p-based mistakes are high, even a very low remaining probability that $\neg p$ can keep S from being practically adequate.[12] This explains the evidential inadequacy in GATECRASHERS, even though frequently it is permissible to accept propositions that are less than .95 likely on one's evidence.

When the costs of error are high largely because a p-based mistake would harm or morally wrong another agent A, accepting p without adequate justification is both epistemically and morally impermissible. Epistemically, it involves dismissing relevant possibilities without adequate reason. Morally, in doing so S fails to give adequate weight to A's moral claims, and recklessly imposes a risk of the error costs on A. Even if she is lucky—p turns out to be true, so the harm to A of a p-based mistake fails to eventuate—she has still done moral wrong in recklessly imposing the risk. GATECRASHERS and BUSES have this structure.

On my analysis, while the moral stakes are relevant to explaining why probabilifying evidence is inadequate for acceptance in the motivating cases, it isn't because moral prohibitions against accepting p trump epistemic permissions. And, while I've articulated an account on which we can expect that generalizations will license accepting 'morally risky' propositions less often than stronger types of evidence, this isn't because generalizations are intrinsically incapable of grounding moral conclusions. This yields a straightforward and compelling gloss of the obligations violated in the puzzle cases. Since we are cognitively limited agents, a crucial element of epistemic responsibility is to accurately judge when the evidence we have is sufficient for acceptance, given what we risk in the event of a mistake, and when we must keep inquiry open and continue to seek better evidence. As moral agents, we also have an obligation to accord appropriate weight to the concerns of others in our deliberations. These are the obligations violated when S does but *shouldn't* accept p on e.

4 Harms, risks, and wrongs

Given what I've said so far, one might still reasonably expect that the woman in COSMOS should be permitted to assume that Franklin is staff. After all, while certainly embarrassing and a breach of propriety, it isn't obvious that the costs of her error are especially high. To explain the rational impermissibility of accepting p in this case, and the frequent inadequacy of racial generalizations more generally, we'll need to spend some time looking more closely at how patterns of inference affect risk distributions.

[12] If the costs of error are high enough, it may be that no merely probabilistic evidence (short of certainty) can justify S in accepting p, in which case to accept she'll need strong evidence that p.

I'll work up to COSMOS slowly, starting with the relevance of risk for more straightforward cases. Some theorists argue that we can morally wrong others just by forming certain beliefs about them.[13] I think this is probably true, but I don't want to rely on it as a premise. I'll start instead at the more minimal assumption that there are some propositions p such that, if you intentionally treat A as if p is true of them, and it isn't, you would wrongfully harm A. For example if you believe that Alejandro has entered a boxing match against you, you may throw the first punch. If you're right this action is permissible, but if you are mistaken in your assumption, you will have wrongfully assaulted him. Let p be a proposition that is morally relevant in this way; if S ϕs and p is false, S will wrong A.

For simplicity, let's assume for now that the wrong involves a direct, concrete harm to A's interests. Now we need to connect it to *acceptance*. Recall that to accept something is to take it for granted, adding it to the stock of propositions you're willing to act on without further consideration. If your acceptance plays its characteristic role, then you'll be disposed to act as if certain in p: you'll be less attentive to evidence whether p, and if you face a decision in which *if* p *it is best to* ϕ, then you will ϕ. You will be so disposed until your evidential context changes, either by encountering compelling evidence that $\neg p$, or by a significant change in the costs of p-based errors. If S accepts that p on evidence that leaves some chance that $\neg p$, then there's some risk she will ϕ when p is false. It's a matter of luck whether S actually faces such a decision. If p is false and she ϕs, S will impose a harm H on A. So, in accepting that p on such evidence, S imposes some risk of H on A. The risk is greater the higher the uneliminated chance of $\neg p$, and more weighty the greater the magnitude of H.

Risks can be justified either evidentially or morally, and of course these two sources of justification can interact. A risk may be evidentially justified if we have enough reason to think that the harm won't eventuate. It can be morally justified if the other error possibilities are significant enough to outweigh A's grounds for complaint against being exposed to a risk of H. If mistakenly accepting p and ϕing would seriously wrong A, S can justify acceptance only if either she risks something of similar weight if she mistakenly *fails* to accept p, or her evidence that p is especially strong. If S accepts p without adequate epistemic care, given the costs, then S imposes an unjustified risk of H on A. The magnitude of the wrong involved in this risk imposition is plausibly a function of (i) the seriousness of H, including the stringency of A's right against suffering H; (ii) the probability of $\neg p$ on S's evidence; and (iii) the severity of S's failure to do due diligence in her inquiry whether p.

Having accepted that p, S might directly *harm* A by acting as if p. But even if she luckily avoids imposing these harms—either because p is in fact true, or because S never has the opportunity to act on her assumption that p—she *wrongs* A if she accepts p without adequate justification. She's closed inquiry too early, recklessly exposing A to unjustified risk of suffering the harms involved in p-based mistakes. When H is severe or is a violation of a stringent right, unjustifiedly imposing the risk of H seriously wrongs A. While we began by assuming that H is a direct, concrete harm to

[13] See especially Basu (2018) 'What We Epistemically Owe to Each Other' and 'The Wrongs of Racist Beliefs', and Basu and Schroeder (2018).

A's interests, it should be clear now that it need not be. So long as A has a moral claim against S imposing H, S wrongs A if she imposes an unjustified risk of H on A.

In some cases (including GATECRASHERS), the severity of the harm of a single, isolated mistake suffices to explain the wrong involved in unjustified acceptance. In others, however, the wrong arises not from the one-off costs, but from the pattern of repeated risk imposition over time. The evidence used as grounds for accepting p can partially shape the pattern, disproportionately exposing an individual or group to repeated risks of the harms of that type of error. It is this that explains the failure in BUSES,[14] and why even highly probabilifying generalizations based in race, gender, or other visible social identities seem intuitively to be peculiarly inadequate to justify acceptance of stereotypical propositions about group members.[15]

In general, risk exposure is disvaluable because it tends to undermine agents' autonomy. Merely being exposed to risk reduces the modal space of 'safe alternatives' that make up an agent's deliberative options, narrowing the space for him to autonomously shape his life.[16] If the agent is aware of the risk and wishes to avoid the harm eventuating, he may change his behavior, taking costly protective or risk-reducing measures that he would not otherwise choose. If the risk imposition was wrongful, these follow-on costs can rightly be considered part of the wrong done in imposing the risk. Even despite these measures, he may live in fear of the harm eventuating, and thus suffer a very real reduction in his quality of life. All of these effects are magnified when the agent is subject to repeated exposure of the same risk. So, one-time risk exposure is a less serious harm, all else equal, than a repeated imposition of unjustified risk. Unjustified risks that are imposed in a patterned way, such that one group of individuals faces disproportionate repeated risk, are thus worse than randomly distributed risk of the same harm.

If some of A's socially visible identity-tracking features (race, sex, etc.,) are a popular heuristic for inferring that p, then simply in virtue of his visible identity, A is likely to be exposed to risk of p-based mistakes in a wide variety of contexts, by a variety of agents. When these features are also stable properties, which A will exhibit for a significant portion of his life, the risk compounds. This patterned risk exposure transforms risks of harms that might have been insignificant, taken on their own, to forces that shape and restrict A's opportunities and alternatives. The net effect of the pattern resulting from using visible group membership as a heuristic is often that it deprives A (and other members of the group) of access to social goods and opportunities that others enjoy. The fact that a given risk imposition forms part of this pattern magnifies the wrong A suffers when an identity-tracking property is used as a basis for the risky action.

[14] If we allow a policy of relying on the market-share evidence to determine fault, we would hold the Blue Bus co. liable for every crash in this region for which we lack more specific evidence, whereas the eyewitness' risk of false-positive error is not similarly concentrated on one party.

[15] Where a 'stereotypical proposition' is roughly that A has some property stereotypically had by his group.

[16] Oberdiek (2008) argues that this modal restriction is how pure risk imposition can harm agents. In that it reduces the space of valuable deliberative alternatives, risk undermines the necessary conditions for the agent's exercise of autonomy in deciding how to shape her life. It therefore can be a setback to the agent's strong interest, and thus a violation of his right.

The pattern also magnifies the wrong S commits in using A's membership in an identity-tracking group as a basis for accepting p. While a reasoner who *idiosyncratically* accepts p based on statistics about the group might impose a relatively minor risk, an agent who relies on a stereotypical heuristic contributes to the collective harm imposed by the patterned risk. The agent in the latter case is similar to Parfit's harmless torturers: though the harm she does directly, on her own, is minimal, the harm expectably caused by her action *together with* the many others who she reasonably can expect to act is significant. Consequently, the wrong she does is magnified by the fact that her action forms part of the larger pattern.[17]

This is not to say that agents must abide by a universalisability maxim, accepting p only if, were many others to do likewise, the resultant risks of error would still be acceptable. Idiosyncratic inferences would fail such a test as badly as stereotypical ones. Rather, in estimating the expected costs of acceptance, the agent must take the world as she finds it. The fact that she knows many others are *already* disposed to make inferences that, collectively, wrongfully harm a group, gives S strong moral reason to avoid contributing to the harm. The moral disvalue of acting against this reason is an additional cost of accepting p on a stereotypical heuristic e. We need not say that the *harm* attributable to S is greater, given that her action forms part of a larger collective harm; only that when S's action would contribute to a significant collective harm, the moral reason to refrain is stronger than it would otherwise be, and hence the moral wrong S does if she acts anyway is greater.

To see this more concretely, let's walk through the COSMOS case. The proposition in question is 'Franklin is a staff member.' The woman's evidence is the fact that Franklin is black, together with the generalization that all the other black men are staff. If p is true and she hands him her coat check, she gets her coat more conveniently than if she waits for someone in an attendant uniform. Suppose her evidence justifies a high credence in p. Since she would be slightly inconvenienced if he *is* staff and she failed to hand him her ticket, if nothing much hangs on $\neg p$ she'll be practically adequate with respect to p. But since the costs of failing to correctly accept p are slight, they can be easily outweighed if the costs of mistakenly accepting p are significant.

The one-off costs are relatively insignificant: it's an affront to Franklin's status, a failure to recognize his achievements, perhaps it's offensive. But part of what feels wrong about the woman's inference is that her evidence—Franklin's membership in the African American racial group—is a stereotypical heuristic for inferring lower social status (in this case, being staff rather than a member of the club).[18] The woman in the Cosmos Club is just one of many agents who will take Franklin's race as license to assume he occupies a socially subordinate role. The mistake made once, idiosyncratically, may be trivial; but made frequently it deprives the members of that race of opportunities to signal authority and high social status, which in turn limits

[17] (Parfit 1984, p. 80). In Parfit's case, "Each of the thousand torturers presses a button, thereby turning the switch once on each of the thousand instruments. The victims suffer the same severe pain. But none of the torturers makes any victim's pain perceptibly worse."

[18] As many authors have noted [see, e.g., Armour (1994); Anderson (2010)], Black Americans of high social status are routinely subject to mistakes of this kind. Examples abound, ranging from hotel guests being mistaken for vagrants, restaurant patrons for valets, etc.

their opportunities for advancement and constrains their options in a way incompatible with respecting their moral equality and autonomy. If she mistakenly accepts p based on this generalization, she participates in the collective harm constituted by the pattern of repeated impositions, which she has strong moral reason not to do.

Under these conditions, even a small risk of mistake suffices to outweigh the expected benefit of getting her coat more conveniently. So given her current evidence, it's best to play it safe and wait for a uniformed attendant, though conditional on p it'd be best to hand Franklin her coat check. So, her current evidence fails to render her practically adequate. And while the generalization makes p very probable, it isn't strong enough evidence for p (on any of the candidate notions of evidential strength) to license the woman to directly discard the possibility that $\neg p$. So accepting p on her current evidence is epistemically impermissible; to do so anyway imposes unjustified risks on Franklin, which is morally impermissible.

5 Explanatory power

One might wonder whether we should do without practical adequacy, and simply hold that strong (in the preferred non-probabilistic sense) evidence is always necessary for rational acceptance, in *any* context. There are two reasons to resist this. The first is that the standards for rational acceptance seem to actually depend on the costs, giving us independent motivation for the practical adequacy requirement.

When it doesn't much matter whether S mistakenly accepts p, or when the costs of withholding are much higher than the costs of mistaken acceptance, merely probabilifying generalizations seem sufficient to justify acceptance. Intuitions may differ, of course, but if it seems that acceptance is rationally permitted in the two cases below, then there is good reason to think that agents with the same evidence may have different epistemic permissions, and to avoid claiming that strong evidence is always necessary:

> **Hobbyist**: To unwind at night you read old court cases. You read the Blue Bus case, and based on the statistics, form the opinion that Blue Bus Co. was responsible for Jones' injuries. You then go to bed, and have no further dealings with the case or claimants.
> **Bored**: You are bored at the Cosmos club, and decide to ask the nearest staff member for a drink. You see someone (Franklin) who looks to you like an attendant. You know that all attendants wear nameplates, but you can't from this distance tell whether he has a nameplate. Knowing that if you're mistaken you'll discover your error before addressing him, you walk over.

It seems to me that you are permitted to accept p in both of these cases. In HOBBYIST you may accept that Blue Bus is responsible, while the judge in BUSES may not, because *given what they would do, were they to accept* p, the judge is in a more evidentially demanding context. Unlike you, the judge faces high-risk p-based decisions, and would likely act in a way that wrongs someone if they mistakenly accept p. In BORED, you know that all staff wear nameplates; if Franklin isn't staff, you can expect to get compelling evidence to that effect before acting in a way that wrongs him. Given

that, the only costs of mistakenly accepting p are temporarily having a false belief, and taking a fruitless walk. The fact that *after accepting* p, you can re-introduce the possibility that $\neg p$ when confronted with compelling evidence is what insulates you from risking wronging anyone. These verdicts would not follow from an unbounded commitment to the necessity of strong evidence, but are predicted by the account I have outlined: whether an agent's evidence renders her practically adequate depends on the costs, and the costs are partially determined by what p-dependent decisions she is likely to face, and whether she can expect to be saved from error by encountering compelling evidence that $\neg p$ in a timely way.

The second reason to prefer a cost-sensitive practical adequacy framework over a simpler strength-only alternative is that it grounds a better explanation of the effects of what I'll call *social signals*. Many behaviors have public social meanings: asserting that p means p is true; wearing a staff uniform means you are a staff member; extending your hand in a certain way means you invite the other party to shake hands, etc. Each of these behaviors can be performed misleadingly. But if these behaviors are adequately avoidable—their meanings are public knowledge, and agents can avoid the behaviors without undue cost—then in some sense A accepts the associated risks if he performs them misleadingly. To illustrate: if A were to assert to S "it's fine with me if you borrow my car", then even if it isn't fine, the fact that he has voluntarily made this speech undermines A's complaint against S's borrowing the car.[19]

Observing A perform a social signal with the meaning p changes S's position in two important ways. It gives S evidence on which it is likely that p, and A's voluntary performance of such a signal undermines his complaint against suffering the risks associated with S's accepting p, effectively lowering the costs of acceptance. These effects can be seen in another variant on COSMOS:

> **Alibi:** It's Halloween at the Alibi Club, and many guests are in costume. For a lark, John Hope Franklin attends dressed in a staff uniform. All the staff are white, but all the other people in uniforms are staff. Seeing him, a woman assumes he is staff, and handing him her coat check, demands her coat.

The epistemic and moral wrongs of the original case are absent here, though the proposition accepted in ALIBI is the same as in COSMOS, and is still false. The woman's evidence is still a generalization from a property Franklin shares with all and only staff in the club (race in COSMOS, attire here).[20] In both cases it should be salient to her that Franklin could have the relevant property without being staff. The key difference

[19] Obviously not all *operable* social signals are adequately avoidable to have this moral effect, and so not all have these epistemic 'licensing' effects. To fully address the moral significance of voluntary risk, or thoroughly characterize the conditions under which social signals are morally active in this way would require a longer discussion than I have space for here. I have, however, begun to address this question in other work; see 'The Case for Conventional Defensive Permissions', (ms).

[20] In both cases, the woman also disregards a property he has that is lacked by all the staff (wearing a non-uniform in COSMOS, being non-white in ALIBI.) Not all accounts of non-probabilistic evidential strength will consider the woman's evidence in ALIBI strong: since p is false, the woman's evidence is not sensitive, and since the fact that some people are wearing costumes is already part of her total evidence, she needs no additional explanation to make Franklin's attire consistent with his being a guest. So the fact that he's wearing a uniform doesn't normically support p. It's less clear whether the evidence satisfies a *safety* standard.

in ALIBI is that her inference is based on a social signal that undermines Franklin's complaint against her assumption. Not all operable social signals are avoidable enough to have this moral effect, but the ones that are appear to help justify acceptance by lowering the costs of error.

If the analysis I've offered is correct, we should expect puzzle cases to arise for acceptance when p-based mistakes risk a significant wrong, S's primary evidence is a generalization, and there is no permissible signal. But we shouldn't expect statistical direct inference to always fail: when the risks of mistakenly accepting p are lopsided and fall mostly on S, or her evidence e is a permissible signal, it is likely that S is practically adequate in accepting p on e.

6 Wrapping up

I have argued that the puzzle cases share a structural flaw: accepting p while lacking the evidence sufficient to justify ignoring the possibility that $\neg p$, given the costs of mistake. The consequence of mistakenly accepting p in each case is that we treat the individual who is the target of the belief in a way we ought not. This account is flexible, not forcing any particular conception of the wrongs involved, but the model can be made more or less demanding by adopting a more or less expansive picture of these wrongs.

When coupled with an understanding of repetition as magnifying the wrongs of unjustified risk imposition, we can explain what has gone wrong in COSMOS as an epistemic error, parallel to the more obvious problems in GATECRASHERS and BUSES. Given the wrongs risked by accepting the relevant proposition, the agent's evidence—though it may make p highly probable—is not strong enough to justify *accepting* p. The additional or distinctive feeling of wrong in COSMOS can be traced to the way the woman's use of a stereotypical heuristic relates to the harms risked. Widespread use of such heuristics not only subjects group members to increased risk of p-based mistakes, but also deprives them of social goods for which being able to signal that $\neg p$ is a necessary condition.

This explanation is fully general, and accurately predicts the conditions under which we should expect puzzle cases to arise, while generating a plausible pattern of epistemic permissions and prohibitions. To be epistemically permitted to accept p, an agent's evidence must either be strong enough to license her to directly presume that p, or render her practically adequate, given the costs. Though most highly probabilifying evidence is also strong, statistical generalizations are an anomaly in this regard. Together with the fact that generalizations over social identities like race tend to raise the costs of error, this explains why the puzzle appears at first to be unique to inferences from identity-tracking generalizations like the one in COSMOS.

A nice upshot of this account is that in explaining why identity-tracking inferences tend to raise the stakes, it provides a way to reconstruct the legal notion of 'protected categories', and gives a transparent rationale for presumptively excluding evidence consisting in generalizations over them from legal proceedings. Better still, it does this without having to give a positive statement of the nature or membership conditions for such categories, a project which can be both metaphysically and ethically fraught.

Acknowledgements This work grew out of conversations with Rima Basu, John Hawthorne, and Mark Schroeder, who each also generously gave me comments on several drafts, and to whom I owe particular gratitude. Thanks also to Michael Ashfield, Julianne Chung, Justin D'Ambrosio, Kenny Easwaran, Georgi Gardiner, Alan Hájek, Elizabeth Jackson, Ethan Landes, Dustin Locke, Jonathan Quong, and James Willoughby for helpful comments, and to Matthew Babb, Maegan Fairchild, Sahar Joakim, Colin Klein, Jeremy Strasser, Jake Ross, and the audiences at the 2016 Arché Epistemology Workshop at St. Andrews, the Talbot Philosophical Society at Biola University, and the 2017 St. Louis Annual Conference on Reasons and Rationality, for fruitful discussion of earlier versions of this material.

7 Appendix: Credences and reference classes

Arguments that a generalization e justifies accepting a proposition p can be divided into three claims:

1. e yields a high conditional probability for p,
2. So e justifies high credence in p,
3. So e justifies accepting p.

In the main paper I objected only to the move from (2) to (3), but it is worth briefly noting that the move from (1) to (2) is not unproblematic. The following discussion draws heavily on Moss (2018) and Hájek (2007).

The setup of cases like COSMOS invites us to assume that one's credence that a given member of F has a further property G should be roughly equal to the ratio of Fs that are G. The cases mention only a single salient group (race, or presence in the stands, or buses in a city), so the inference seems straightforward. But objects and individuals can be sorted into indefinitely many groups, each of which may have a different ratio of members that are G. To rationally match one's credence to the ratio of Fs, we need some justification for taking F to be uniquely relevant to whether a thing is G. To illustrate, consider a slight variant on GATECRASHERS case:

> **Gate Crashers 2** — Of the 1,000 people in F, only 10 paid the fare. Of the ten former Boyscouts in F, 9 paid the fare; of the 3 Canadians in F, 2 paid. Alfred is a Canadian former boyscout in F.

Let Ga be the proposition 'Alfred failed to pay'. Conditional on being in F, P(Ga) = .99; conditional on Alfred's being a former boyscout, P(Ga) = .1; conditional on Alfred's being Canadian, P(Ga) = .33. With this as evidence, what credence should a rational agent have in Ga? Presumably in asking this question, we think that credences should be substantially constrained by one's evidence; the problem is that the evidence is not univocal. It supports multiple, competing probability assignments, depending on which reference class we attend to. To identify what credence is rational, we'll *first* need to determine whether being a member of F, or former Boyscout, or Canadian, is most relevant to determining whether Alfred failed to pay.[21]

The justificatory hurdle can be skirted only if we lack any information about alternative classes, so that F is the relevant class simply by being the only class we have any evidence about. While the stipulatively sterile environments of puzzle cases some-

[21] This 'Problem of the Reference Class' most famously affects frequentist interpretations of probability, but as Hájek (2007) demonstrates, it is not unique to it.

times give information only about one group, the actual world is positively buzzing with detail, giving agents a wealth of classificatory information about the people who are the subjects of their inferences. So in the kinds of inferences the puzzle cases are supposed to be about, the justificatory challenge must be faced head-on: we must have a reason for taking F to be the relevant class before we can assume it is rational to set one's credences in G*a* to match the ratio of Fs that are G.

In other contexts, we justify relying on a particular class by appeal either to the naturalness of the properties on which it is based, or to the causal relationship between those properties and the target property. For instance, though we could divide precious stones by color into any number of classes, including <grue, bleen>, division into <green, blue> is *natural* in a way that the other alternatives are not. Similarly, though a biopsy reveals many properties about a growth, our knowledge of relevant causal patterns allows us to select some features (size and shape) as relevant and dismiss others (e.g. being discovered on a Tuesday).

Each of these appeals is more dubious when the class we want to rely on is based on a social category like race. Appeals to the naturalness of the property feel out of place, and prospects for a causal connection depend very much on the nature of the target property. This may explain why there seems to be *something* epistemically objectionable even when racial generalizations are only used to ground credences. Doing so presupposes that race is explanatorily relevant to the target property, which, depending on the property, may be both false and insulting. Some (including Moss) argue that when anticipating the properties of autonomous agents, we must always be sensitive to the fact that A may choose to act differently from others in the reference class. If so, it will take more to justify the relevance of a class for claims about agents than about objects, and demographic generalizations will rarely ground either high credence or acceptance of p.

So it is worth distinguishing three possible cases in which S has a generalization e that yields a high conditional probability for p. There are the two that were the focus of the main paper,

1. e justifies high credence in p, *and* justifies accepting p,
2. e justifies high credence in p, *but not* accepting p,

but there are likely also many cases in which

3. e neither justifies high credence in p nor accepting p, because S lacks justification for taking the reference class for e to be uniquely relevant.

I am inclined to think that though statistics about demographic groups will very often lack relevance and hence fail to justify high credence, this need not always be so. Plausibly there are some properties that tend to arise from causal forces (social or natural) that converge on members of these groups, such that group membership is strongly predictive of having the further property. If so, the demographic statistics may justify high credence in these cases, while *still* failing to justify acceptance.

On the view I have offered, the downstream effects on evidence-gathering and action explain this asymmetry. While accepting involves treating the question *whether* p as closed, simply having a high credence in p is compatible with continuing to seek stronger evidence whether p and considering the possibility that $\neg p$ when choosing

one's action. It seems morally and epistemically permissible to increase your credence in proportion with your (properly justified) statistical evidence; this credence will still be sensitive to reform in the face of additional evidence. Furthermore, in being attentive to the possibility that $\neg p$, agents' reasoning takes into account the cost of being mistaken whether p; when these costs are high, even agents with high credence will be cautious when making p-dependent decisions.

References

Anderson, C., & Hawthorne, J. (forthcoming). Knowledge, practical adequacy, and stakes. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 6). Oxford University Press.

Anderson, E. (2010). *The imperative of integration*. Princeton, NJ: Princeton University Press.

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

Arntzenius, F., Elga, A., & Hawthorne, J. (2004). Bayesianism, infinite decisions, and binding. *Mind*, *113*(450), 251–283.

Basu, R. (2018). Beliefs that wrong. Ph.D. thesis, University of Southern California.

Basu, R., & Schroeder, M. (2018). Doxastic wrongings. In B. Kim & M. McGrath (Eds.), *Pragmatic encroachment in epistemology*. Routledge Studies in Contemporary Philosophy (forthcoming).

Bratman, M. (1992). Practical reasoning and acceptance in a context. *Mind*, *101*(401), 1–15.

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

Cohen, J. (1977). *The probable and the provable*. Oxford: Clarendon Press.

Cohen, J. (1989). Belief and acceptance. *Mind*, *98*, 367–89.

Cohen, J. (1992). *An essay on belief and acceptance*. Oxford: Oxford University Press.

Engel, P. (1998). Believing, holding true and accepting. *Philosophical Explorations*, *1*(2), 140–151.

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

Franklin, J. H. (2005). *Mirror to America: The autobiography of John Hope Franklin*. New York, NY: Farrar, Straus and Giroux.

Fritz, J. (2017). Pragmatic encroachment and moral encroachment. *Pacific Philosophical Quarterly*, *98*, 643–661.

Gardiner, G. (2018). The burden of proof and statistical evidence. In D. Coady & J. Chase (Eds.), *The Routledge handbook of applied epistemology*. Routledge Handbooks in Philosophy (forthcoming).

Gendler, T. (2011). On the epistemic costs of implicit bias. *Philosophical Studies*, *156*(1), 33–63.

Haack, S. (2012). The embedded epistemologist: Dispatches from the legal front. *Ratio Juris*, *25*(2), 206–35.

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

Locke, D. (2013). Practical certainty. *Philosophy and Phenomenological Research*, *90*(1), 72–95.

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

Oberdiek, J. (2008). Towards a right against risking. *Law and Philosophy*, *28*(4), 367–392.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.

Peet, A., & Pitcovski, E. (2018). Normal Knowledge: Toward an explanation-based theory of knowledge. *The Journal of Philosophy*, *115*(3), 141–157. https://doi.org/10.5840/jphil201811539.

Portmore, D. (2008). Dual-ranking act-consequentialism. *Philosophical Studies*, *138*(3), 409–427.

Pritchard, D. (2017). Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence*, *9*(1), 108–119. https://doi.org/10.1080/20403313.2017.1352323.

Ross, J., & Schroeder, M. (2014). Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, *2*, 259–288.

Russell, J. (forthcoming). How much is at stake for the pragmatic encroacher. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 6). Oxford University Press.

Smith, M. (2010). What else justification could be. *Noûs*, *44*(1), 10–31.

Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.

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

Worsnip, A. (2015). Two kinds of stakes. *Pacific Philosophical Quarterly*, *96*, 307–324.