

# Social Censorship: The First Offender Model

*Posted on April 2, 2019 by Scott Alexander*



RJ Zigerell (h/t Marginal Revolution) studies [public support for eugenics](#). He finds that about 40% of Americans support some form of eugenics. The policies discussed were very vague, like “encouraging poor criminals to have fewer children” or “encouraging intelligent people to have more children”; they did not specify what form the encouragement would take. Of note, much lack of support for eugenics was a belief that it would not work; people who believed the qualities involved were heritable were much more likely to support programs to select for them. For example, of people who thought criminality was completely genetic, a full 65% supported encouraging criminals to have fewer children.

I was surprised to hear this, because I thought of moral opposition to eugenics was basically universal. If a prominent politician tentatively supported eugenics, it would provoke a media firestorm and they would get shouted down. This would be true even if they supported the sort of generally mild, noncoercive policies the paper seems to be talking about. How do we square that with a 40% support rate?

I think back to a metaphor for norm enforcement I used in an [argument](#) against Bryan Caplan:

Imagine a town with ten police officers, who can each solve one crime per day. Left to their own devices, the town's criminals would commit thirty muggings and thirty burglaries per day (for the purposes of this hypothetical, both crimes are equally bad). They also require different skills; burglars can't become muggers or vice versa without a lot of retraining. Criminals will commit their crime only if the odds are against them getting caught – but since there are 60 crimes a day and the police can only solve ten, the odds are in their favor.

Now imagine that the police get extra resources for a month, and they use them to crack down on mugging. For a month, every mugging in town gets solved instantly. Muggers realize this is going to happen and give up.

At the end of the month, the police lose their extra resources. But the police chief publicly commits that from now on, he's going to prioritize solving muggings over solving burglaries, even if the burglaries are equally bad or worse. He'll put an absurd amount of effort into solving even the smallest mugging; this is the hill he's going to die on.

Suppose you're a mugger, deciding whether or not to commit the first new mugging in town. If you're the first guy to violate the no-mugging taboo, every police officer in town is going to be on your case; you're nearly certain to get caught. You give

up and do honest work. Every other mugger in town faces the same choice and makes the same decision. In theory a well-coordinated group of muggers could all start mugging on the same day and break the system, but muggers aren't really that well-coordinated.

The police chief's public commitment solves mugging without devoting a single officer's time to the problem, allowing all officers to concentrate on burglaries. A worst-crime-first enforcement regime has 60 crimes per day and solves 10; a mugging-first regime has 30 crimes per day and solves 10.

But this only works if the police chief keeps his commitment. If someone tests the limits and commits a mugging, the police need to crack down with what looks like a disproportionate amount of effort – the more disproportionate, the better. Fail, and muggers realize the commitment was fake, and then you're back to having 60 crimes a day.

I think eugenics opponents are doing the same thing as the police here: they're trying to ensure certainty of punishment for the first offender. They've established a norm of massive retaliation against the first person to openly speak out in favor of eugenics, so nobody wants to be the first person. If every one of the 40% of people who support eugenics speak out at once, probably they'll all be fine. But they don't, so they aren't.

Why aren't we in the opposite world, where the people who support eugenics are able to threaten the people who oppose it and pre-

vent them from speaking out? I think just because the opponents coordinated first. In theory one day we could switch to the opposite equilibrium.

I think something like this happened with gay rights. In c. 1969, people were reluctant to speak out in favor of gay rights; in 2019, people are reluctant to speak out against them. Some of that is genuinely changed minds; I don't at all want to trivialize that aspect. But some of it seems to have just been that in 1969, it was common knowledge that the anti-gay side was well-coordinated and could do the massive-retaliation thing, and now it's common knowledge that the pro-gay side is well-coordinated and can do the massive retaliation thing. The switch involved a big battle and lots of people massively retaliating against each other, but it worked.

Maybe everyone else already realized something like this. But it changes the way I think about censorship. I'm still against it. But I used to have an extra argument against it, which was something like "If eugenics is taboo, that means there must be near-universal opposition to eugenics, which means there's no point in keeping it taboo, because even if it wasn't taboo eugenicists wouldn't have any power." I no longer think that argument holds water. "Taboo" might mean nothing more than "one of two equally-sized sides has a tenuous coordination advantage".

(in retrospect I was pretty dumb for not figuring this out, since it's pretty the same argument I make in [Can Things Be Both Popular And Silenced?](#) The answer is obviously yes – if Zigerell's paper is

right, eugenics is both popular and silenced – but the police metaphor explains how.)

The strongest argument against censorship is still that beliefs should be allowed to compete in a marketplace of ideas. But if I were pro-censorship, I might retort that one reason to try to maintain my own side's tenuous coordination advantage is that if I relax even for a second, the other side might be able to claw together its own coordination advantage and censor *me*. This isn't possible in the "one side must be overwhelmingly more powerful" model of censorship, but it's something that the "tenuous coordination advantage" model has to worry about. The solution would be some sort of stable structural opposition to censorship in general – but the gay rights example shows that real-world censors can't always expect that to work out for them.

In order to make moderation easier, please restrict yourself to comments about censorship and coordination, not about eugenics or gay rights.