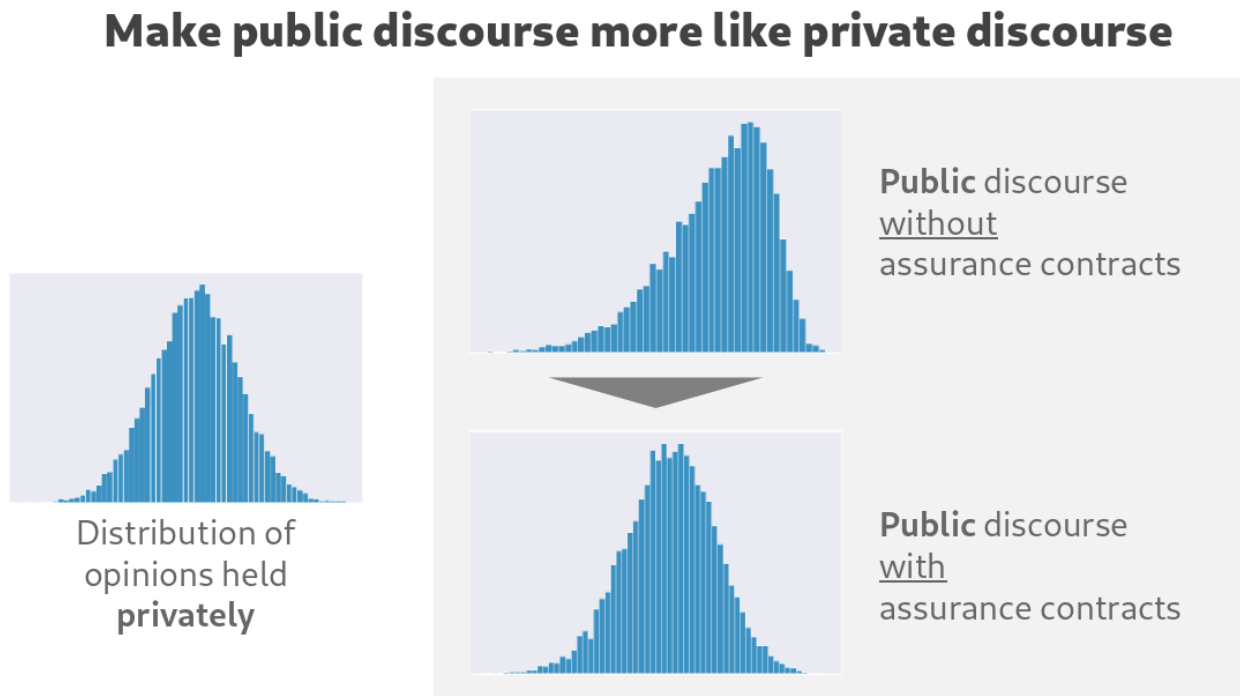


**Motivation:**

Political and ideological polarization is a core problem facing our civil society today, and one of the drivers of the increase in polarization in recent years is the uneven reflection of privately-held beliefs in the public sphere. If we were to try to infer what the average American is like on the basis of news coverage and social media, we would end up with a caricature: this person is Blue or Red, Left or Right, R or D. He carries a gun, drives a truck, and goes to church—or he eats vegetarian, drives a hybrid, and goes to protests.

There is a tyranny of the minority in public discourse, where we see the vast majority of content generated by a small minority of posters (Hughes & Asheer 2019). Instead of reflecting the diversity and relative moderation of actual beliefs and behavior, news media and social media content is as inflammatory as is profitable. Expressing beliefs that are rare in the public sphere is often perceived to be too dangerous to be worthwhile, so the mere threat of punishment has chilling effects—dissenters opt for silence, and the diversity of opinion characteristic of honest, open exchange of ideas eludes us all. The resulting distorted public discourse results in the average person wrongly inferring that talked-about opinions are commonly held, rather than sparsely held and merely profitable—a phenomenon referred to as pluralistic ignorance (Prentice & Miller 1993).

I am exploring the use of a piece of social technology designed to make the distribution of beliefs in public discourse more accurately reflect the underlying distribution actually held by the public: the assurance contract. An assurance contract takes effect only when the provision point, a pre-specified threshold, is reached.



*Figure 1: Assurance contracts will make the distribution of ideas discussed and viewpoints reflected in public discourse more similar to that held in private*

This mechanism already exists in the economic domain and is well-used if not particularly well-studied. Crowdfunding platforms like Kickstarter make use of assurance contracts (Agarwal et al. 2014). For them, the assurance contract allows for private provision of public goods: once a

group has \$500,000 in pledges, they can manufacture the new widget, say pet-carrier backpacks, those pledging want. Here, the provision point is \$500,000 and the good produced (in this case a “club good”) is the backpacks. Tabarrok (1998) refines the concept and develops “dominant” assurance contracts, which provide an incentive for investing even in the event of the contract’s failure.

When we import this mechanism into the social domain, it is an agreement something like an open letter—with the exception that the signatures on the letter become public only *after* some safety conditions (analogous to the provision point) are fulfilled. At a university it might look like this:

We, the undersigned believe [controversial idea] and think the University should take the following steps [1, 2, 3...]. Signatures on this letter will become public *only when there are at least [200] signatures from faculty and [800] signatures from students*. Until then, no one except the keeper of this letter can see who has signed.

This serves to make expressing a controversial (or controversial-feeling) idea much less costly: while a single hand raised in dissent might get cut off, a thousand can be safely raised together.

### **Proposed research:**

#### *Theoretical work: Modeling*

A first piece of the proposed research will develop a model of the assurance contract, and a second will explore possible attacks and mitigations. The modeling piece will allow some understanding of what individual agents will have to believe about the possible costs of speaking up without an assurance contract, the reliability of the contract writer, and, most importantly, the safety conditions. One possible attack is analogous to ballot-stuffing: opponents flood the contract with fake signatures, triggering the publication of the contract. Since the opponents know which signatures are fake, they also know who has signed the contract—and all before the safety conditions have been fulfilled by genuine signatures. Another is an untrustworthy contract writer: the entity collecting the signatures could, either deliberately or accidentally, leak signatories’ names before the safety threshold is reached. Both of these attacks might be mitigated using a combination of modern identity providers present at universities (for instance, MIT uses cryptographic certificates), byzantine fault tolerance, and smart contracts—but the details remain to be discovered.

#### *Empirical work: An experiment in MBA classrooms*

We propose an experiment to take place in MBA classrooms. Subjects (MBA students) will be asked if they would like to stay after class for about 30 minutes to participate in an experiment about political participation in exchange for some compensation. The subjects use their laptops to go to a website where they are given a list of 10 political issues current in public discussion and are asked to indicate which they support and which they oppose, indicating a minimum of 5 positions. Pre-testing indicates 7 of these issues are contentious within the MBA milieu, and three are not. They then make their stance known by filling out “badges”, folded paper cards which will sit on their desktops and which are visible to their peers during the exercise. Once badges are visible on all desktops, the subjects complete the distractor task, in which they predict whether a given randomly assigned peer has voted in the last local election on the basis of the badges he or she has chosen to display.

We will manipulate between groups whether an assurance contract is available while they are making their badges. Some groups have the opportunity to signal holding a given opinion publicly only if a certain number of other people do the same, while others have no such “safe” option. In the assurance contract condition, subjects will be told: “For each of these issues your stance will only be disclosed if at least 35% of the people in the classroom share the same opinion.”

We will then look for differences in which badges are displayed as well as the relative support and opposition for each badge, with and without an assurance contract. We hypothesize that assurance contracts will increase the diversity of opinions expressed, as well as increasing the support for the minority counter-narrative stance for a given opinion. This provides good experimental control, allowing us to accurately measure the effect assurance contracts have on students' willingness to express controversial opinions in front of their peers.

The general setup can be expanded to many other venues where the expression of controversial ideas can be costly. Another likely setting is social media: Twitter field experiments could measure underlying sentiment on a wide variety of ideas as well as public discussion of the same, looking for ideas which are well-represented in private data but rare in public discussion. These would be good targets for a field experiment that makes assurance contracts available to participants in certain online communities within Twitter, such as academics.

*Practical work: Implementing assurance contracts for good*

Building tools that make it easy to implement assurance contracts in the real world would be a natural next step. I envision a small nonprofit running a website that makes it very easy to deploy credible assurance contracts.

**Works Cited:**

Agrawal, A., Catalini, C., & Goldfarb, A. (2014). Some Simple Economics of Crowdfunding. *Innovation Policy and the Economy*, 14, 63–97. <https://doi.org/10.1086/674021>

Hughes, A., & Asheer, N. (2019). Those who strongly disapprove of Trump produce majority of tweets from U.S. adults that mention national politics. *Pew Research Center*, 25.

Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *Journal of personality and social psychology*, 64(2), 243.

Tabarrok, A. (1998). The private provision of public goods via dominant assurance contracts. *Public Choice*, 96, 345–362.