Growing together: assessing the peer-effects of a growth mindset intervention.

(Tech Report)

Guillaume Basse

July 3, 2020

Abstract

Do students who did not receive the mindset intervention benefit from being in contact with students who did? This question is important from a policy perspective. Indeed, the overall effectiveness of the mindset intervention is a combination of its direct effect and its peer effects, so evidence of peer-effects would bolster the case for a wider adoption of mindset interventions. In this technical report, we study the opportunities and limits of a large scale mindset experiment for estimating spillover effects. We address methodological complications arising from the design and provide a blueprint for designing future experiments focusing on the spillover effects.

1 Introduction

Are experimental units assigned to a control condition affected by an intervention if some fraction of their peers receives the treatment condition? That is, are there spillovers from the treatment condition to the control condition? This phenomenon, called interference, has received increasing attention in the causal inference literature (Aronow et al., 2017; Hudgens and Halloran, 2008; Basse and Feller, 2018) and can lead to serious bias if not accounted for.

Our motivating example is the National Study of Learning Mindsets (NSLM), a large randomized experiment for evaluating an online growth mindset intervention in american K-12 schools (Yeager et al., 2019). In the study, students were assigned at random to receive a growth mindset intervention and a variety of educational outcomes were collected. The analysis conducted by Yeager et al. (2019) implicitly assumes that there are no treatment spillovers. In this paper, we relax this assumption and propose an approach for conducting inference on the spillover effects.

We build on the framework of Aronow et al. (2017), but address a number of complications that arise in our application. The two key issues are the fact that spillovers are expected to be negligible except when a unit has a high proportion of treated neighbors, but that it is unlikely to observe units with very high (or very low) proportions of treated neighbors given the way the experiment was designed. We propose an approach that partially alleviates these issues. We also highlight the fundamental limitations of the NSLM design for studying spillovers: although our approach improves the precision of existing methods, it is likely insufficient to detect small effects. Indeed, our analysis finds no evidence of spillover effects for the mindset intervention.

Some of the limitations of the NSLM design could be addressed in future experiments by considering alternative assignment mechanisms – we discuss a class of design that would allow a better study of spillovers.

2 A general approach to estimating spillover effects

2.1 Background

We consider a setting with N units labelled $i = 1, \dots, N$. Assume for now that each unit i has a set of peers $\mathcal{N}_i$, known to the experimenter. Let $Z_i \in \{0, 1\}$ be the binary treatment indicator for unit i and $Z = (Z_1, \dots, Z_N)$ the treatment assignment vector. Denote by $Y_i(Z)$ the potential outcome of unit i under assignment i and $Y(Z)$ the potential outcome vector.

Notice that without additional assumptions, the potential outcome of a unit may depend on the entire assignment vector. In practice, we may have reasons to believe that a unit's outcome is affected only by its own assignment and that of a group of its peers. We formalize this idea in the exposure framework of Aronow et al. (2017). For each unit i , define its exposure H_i , as the fraction of treated peers,

$$H_i = h_i(Z) = |\mathcal{N}_i|^{-1} \sum_{j \in \mathcal{N}_i} Z_j.$$

For now, we assume that this exposure is well specified in the following sense:

Assumption 1. *We assume that for all $i = 1, \dots, N$:*

$$\forall Z, Z', \quad \begin{cases} h_i(Z) = h_i(Z') \\ Z_i = Z'_i \end{cases} \Rightarrow Y_i(Z) = Y_i(Z')$$

Under this assumption, the potential outcomes can be written $Y_i(Z) = Y_i(Z_i, H_i)$. A typical spillover estimand would then be:

$$\tau_{h,h'} = \frac{1}{N} \sum_{i=1}^N \{Y_i(0, h) - Y_i(0, h')\}$$

which can be estimated unbiasedly with the Horvitz-Thompson estimator:

$$\hat{\tau}_{h,h'} = \frac{1}{N} \sum_{i=1}^N \frac{\mathbb{1}(Z_i = 0, H_i = h)}{P(Z_i = 0, H_i = h)} Y_i - \frac{1}{N} \sum_{i=1}^N \frac{\mathbb{1}(Z_i = 0, H_i = h')}{P(Z_i = 0, H_i = h')} Y_i \quad (1)$$

Aronow et al. (2017) show that under some mild additional conditions, this estimator is asymptotically normal, and they derive conservative estimators of the variance.

2.2 Continuous exposures

In practice, the exposures may be close to continuous, and so the probabilities $P(Z_i = 0, H_i = h)$ for any given h may be very small. This is problematic for the estimator in Equation 1 for two reasons. First, because it implies that there may be very few (perhaps even zero) units with each exposures h or h' , and so both summations in Equation 1 may have very few terms. Second, because having small probabilities in the denominator in Equation 1 may cause the variance of the Horvitz-Thompson estimator to be unreasonably large.

One way to address this problem is to consider exposure strata. Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two sets, and suppose that the following assumption holds:

Assumption 2. *For all i , we have:*

$$\forall h, h' \in \mathcal{H}_1, \quad Y_i(0, h) = Y_i(0, h') \equiv Y_i(0, \mathcal{H}_1)$$

and

$$\forall h, h' \in \mathcal{H}_2, \quad Y_i(0, h) = Y_i(0, h') \equiv Y_i(0, \mathcal{H}_2)$$

where, with a slight abuse of notation, $Y_i(0, \mathcal{H}_1)$ and $Y_i(0, \mathcal{H}_2)$ denote the common value of the potential outcomes for exposures in $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively.

We could then focus on the following estimand:

$$\tau(\mathcal{H}_1, \mathcal{H}_2) = N^{-1} \sum_{i=1}^N \{Y_i(0, \mathcal{H}_1) - Y_i(0, \mathcal{H}_2)\} \quad (2)$$

and adapt the Horvitz-Thompson estimator as follows:

$$\hat{\tau}(\mathcal{H}_1, \mathcal{H}_2) = N^{-1} \sum_{i=1}^N \frac{\mathbb{1}(Z_i = 0, H_i \in \mathcal{H}_1)}{P(Z_i = 0, H_i \in \mathcal{H}_1)} Y_i(0, \mathcal{H}_1) - N^{-1} \sum_{i=1}^N \frac{\mathbb{1}(Z_i = 0, H_i \in \mathcal{H}_2)}{P(Z_i = 0, H_i \in \mathcal{H}_2)} Y_i(0, \mathcal{H}_2) \quad (3)$$

The key idea is that $P(Z_i, H_i \in \mathcal{H}_1)$ and $P(Z_i, H_i \in \mathcal{H}_2)$ are not as small as the probabilities of each individual exposure — thus alleviating some of the issues we discussed above. In addition, contrasting sets of exposures may actually be of more substantive interest than contrasting individual exposures. Nevertheless, the probabilities $P(Z_i = 0, H_i \in \mathcal{H}_1)$ and $P(Z_i = 0, H_i \in \mathcal{H}_2)$ may still be very low. One way to address this problem is to consider the effect on a subset of the population defined by the observed exposure. Specifically, define the length- N vector $\phi(H)$ as:

$$[\phi(H)]_i = \mathbb{1}\{H_i \in \mathcal{H}_1 \text{ or } H_i \in \mathcal{H}_2\}$$

and let $\mathcal{I}(Z) = \{i : \phi(H)_i = 1 \text{ and } Z_i = 0\}$. We also let $\mathcal{I}^{obs} = \mathcal{I}(Z^{obs})$.

Definition 1 (Average exposure effect on the exposed). *Given an observed assignment vector Z^{obs} and induced exposure vector H^{obs} , we define the average exposure effect on the exposed as:*

$$\tau^{exp}(\mathcal{H}_1, \mathcal{H}_2) = |\mathcal{I}^{obs}|^{-1} \sum_{i \in \mathcal{I}^{obs}} \{Y_i(0, \mathcal{H}_1) - Y_i(0, \mathcal{H}_2)\}$$

We can then state the following proposition:

Proposition 1. *Let $\pi_i(1) = P(H_i \in \mathcal{H}_1 \mid H_i \in \mathcal{H}_1 \cup \mathcal{H}_2)$ and $\pi_i(2) = P(H_i \in \mathcal{H}_2 \mid H_i \in \mathcal{H}_1 \cup \mathcal{H}_2)$. If Assumption 1 and Assumption 2 hold, then the estimator:*

$$\hat{\tau}^{exp} = \frac{1}{|\mathcal{I}^{obs}|} \sum_{i \in \mathcal{I}^{obs}} \frac{\mathbb{1}(H_i \in \mathcal{H}_1)}{\pi_i(1)} Y_i - \frac{1}{|\mathcal{I}^{obs}|} \sum_{i \in \mathcal{I}^{obs}} \frac{\mathbb{1}(H_i \in \mathcal{H}_2)}{\pi_i(2)} Y_i$$

is unbiased for τ^{exp} conditionally. That is $E\{\hat{\tau}^{exp} \mid \mathcal{I}(Z) = \mathcal{I}(Z^{obs})\} = \tau^{exp}$.

Inference can be conducted as for the regular Horvitz-Thompson estimator, but considering conditional instead of marginal probabilities. Here again, the key is that the probabilities $\pi_i(1)$ and $\pi_i(2)$ are generally less extreme than $P(Z_i = 0, H_i \in \mathcal{H}_1)$ and $P(Z_i = 0, H_i \in \mathcal{H}_2)$, and therefore $\hat{\tau}^{exp}$ should have a lower variance than $\hat{\tau}$. The price to pay is an estimand that is somewhat less interpretable since it depends on the observed value of the assignment vector.

2.3 A practical tradeoff

In practice, Assumption 2 will rarely hold exactly. If Assumption 1 does hold, then we might hope that Assumption 2 will be more likely to hold if the sets $\mathcal{H}_1$ and $\mathcal{H}_2$ are small. The tradeoff, of course, is that the smaller these sets are, the least useful the aggregation actually is. A proper study of the sensitivity of our approach to misspecification of assumption 2 is beyond the scope of this technical report, but it is a subject of interest for further study.

3 Challenges and suggestions for designing future experiments

3.1 Challenges with the current study

The NSLM study was designed to estimate the direct effect of the intervention — not the possible existence of spillovers. Certain design choices that guarantee maximum power for detecting direct effects render the study of spillovers very challenging in this setting. We describe the two key challenges below.

3.1.1 Identifying peer-groups

The methodology discussed in Section 2 assumes knowledge of the network of ‘peers’ – indeed, the definition of the exposure depends on $\mathcal{N}_i$, the set of peers of unit i . The data collection for the NSLM study did not include peer-information for most units in the study, since this information is not required for estimating the direct effect of the intervention. Lacking this information, one strategy is to construct a sensible notion of peers based on available covariate information and then use the methods of Section 2 with this constructed set of peers. Here again, a rigorous analysis would require studying the sensitivity of inference to this approximation – this is an open problem, and the subject of ongoing research.

3.1.2 Assignment mechanism

The biggest challenge, however, is the assignment mechanism used in the study. The intervention was assigned independently at random to each student with probability $p = 0.5$. If one is interested in the direct effect, then this is close to optimal. If, on the other hand, one is interested in assessing spillover effects, then it is a particularly challenging choice of design.

Indeed, for unit i , let $\mathcal{N}_i$ be the set of peers for unit i . It is easy to see that the exposure H_i of unit i follows the following distribution:

$$H_i \sim |\mathcal{N}_i|^{-1} B_i; \quad B_i \sim \text{Binomial}(|\mathcal{N}_i|, 0.5)$$

Using a normal approximation, one can see that as the size of peer groups increases, most of the exposures will be concentrated around the expected value $E[H_i] = 0.5$. If $|\mathcal{N}_i| = 30$, then with probability 0.95, the exposure will be between 0.32 and 0.67. If we apply the strategy of Section 2.2, it means that we must select sets $\mathcal{H}_1$ and $\mathcal{H}_2$ that contain values close to 0.5.

The key issue is that if they exist at all, we expect the spillover effects $\tau_{h,h'}$ to be noticeable only for extreme contrasts, such as $h = 0.9$ and $h' = 0.1$. That is, we might expect that having 90% percent of one’s peers treated versus 10% might make a difference; but it seems unlikely that having 55% of one’s peers treated versus 45% will make a difference.

This creates a difficult tradeoff. Natural candidates for the sets $\mathcal{H}_1$ and $\mathcal{H}_2$ would be $\mathcal{H}_1 = [h^{up}, 1]$ and $\mathcal{H}_2 = [0, h^{low}]$ where ideally h^{up} would be close to 1 and h^{low} would be close to 0. Unfortunately, as we have seen above, the probability of observing an exposure in $\mathcal{H}_1$ and $\mathcal{H}_2$ is very small for extreme values of h^{up} and h^{low} . For instance if $|\mathcal{N}_i| = 30$, then $P(H_i \geq 0.75) \leq 0.003$.

So on the one hand we should select extreme values for h^{up} and h^{low} in order to observe any effect (assuming there is one). But doing so reduces the probability of observing the associated exposure, and therefore increases the variance of our estimator, reducing our power to detect the effect.

3.1.3 Analysis for the NSLM data

With all these issues in mind, we nevertheless attempted to analyze the NSLM data. We focused on female students and constructed peer groups $\tilde{\mathcal{N}}_i$ of female students taking the same math class. We used a difference in means estimator:

$$\hat{\tau}_{DM}^{exp} = \frac{1}{N_{up}^{obs}} \sum_{i \in \mathcal{I}^{obs}} \mathbb{1}(H_i \in \mathcal{H}_1) Y_i - \frac{1}{N_{low}^{obs}} \sum_{i \in \mathcal{I}^{obs}} \mathbb{1}(H_i \in \mathcal{H}_2) Y_i, \quad (4)$$

with $h^{up} = 0.75$ and $h^{low} = 0.25$. In the special setting of the NSLM experiment, this estimator is unbiased for τ^{exp}

Proposition 2. *If Assumption 1 and Assumption 2 holds, then the estimator $\hat{\tau}_{DM}^{exp}$ satisfies:*

$$E\{\hat{\tau}_{DM}^{exp} \mid \mathcal{I}(Z) = \mathcal{I}^{obs}, N_1(Z) = N_1^{obs}, N_0(Z) = N_0^{obs}\} = \tau^{exp}$$

We then computed $\hat{\tau}_{DM}^{exp}$ for with $h_{low} = 0.75$ and h_{high} . In our setting, $\mathcal{I}^{obs} = 169$ down from an initial 4,365 female students in the 9th grade. The confidence interval constructed includes 0 so there is no evidence of peer-effects at the 95% level. We must, however, keep in mind the limitations of the design: absence of evidence, in our case, is very far from being evidence of absence.

3.2 Designing future experiments

Many of the challenges highlighted in Section 2 could be mitigated or altogether resolved with a different choice of design. The design used in the experiment assigned units to each treatment condition independently at random with probability 0.5. Instead, we suggest using a form of two-stage design called the random saturation design.

The ideal design would first collect information about peer-groups. Then, if K denotes the number of peer groups and Z_{ik} the treatment indicator of unit i in group k , an example of random saturation design is:

$$\begin{aligned} Z_{ik} &\sim \text{Bernoulli}(\pi_k) \\ \pi_k &= S_k \cdot \tilde{h}^{low} + (1 - S_k) \cdot \tilde{h}^{up} \\ S_k &\sim \text{Bernoulli}(0.5). \end{aligned}$$

In words, each group $k = 1, \dots, K$ is first assigned to a low saturation probability ($\pi_k = \tilde{h}^{low}$) or high saturation probability ($\pi_k = \tilde{h}^{up}$), then each unit i is assigned to treatment independently with the saturation probability corresponding to its group. We then have the following result:

Proposition 3. *We have:*

$$P(Z_i = 0, H_i = \tilde{h}_i^{high} \mid Z_i = 0) = P(Z_i = 0, H_i = \tilde{h}_i^{low} \mid Z_i = 0) \approx \frac{1}{2}$$

Therefore if we take $\tilde{h}^{low} < h^{low} < h^{high} < \tilde{h}^{high}$, then most units with $Z_i = 0$ have $H_i \leq h^{low}$ or $H_i \geq h^{up}$, so no exposure is “wasted”. In this case, we can consider the average exposure effect estimand of Equation 2: the associated Horvitz-Thompson estimator of Equation 3 will be well-behaved.

4 Conclusion and future work

We have discussed some key challenges posed by the estimation of spillover effects in an experiment designed to study direct effects. We developed a few strategies for dealing with some of the issues, and proposed a design that would circumvent an important limitation of this study, in future experiments.

Our work leaves a number of open research directions. First, more work is needed on how to best design experiments in this setting: random saturation designs are a interesting direction. Second, we have not investigated the sensitivity of our results to different choices of exposure and different constructions of peer groups – how to do this in a principled work is an open question.

Finally, although our work finds no evidence of spillover effects, additional experiments designed specifically to study spillovers are needed to answer the question conclusively.

References

- Aronow, P. M., C. Samii, et al. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics* 11(4), 1912–1947.
- Basse, G. and A. Feller (2018). Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association* 113(521), 41–55.
- Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association* 103(482), 832–842.
- Yeager, D. S., P. Hanselman, G. M. Walton, J. S. Murray, R. Crosnoe, C. Muller, E. Tipton, B. Schneider, C. S. Hulleman, C. P. Hinojosa, et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 1–6.