How Does School Choice Effect Measures Of School Performance?

by blind-review

Abstract

Measuring school performance has become an important activity within our education systems. Today, most developed countries have some metric to rank and compare schools — Adequate Yearly Progress (AYP) of No Child Left Behind (NCLB), Progress 8 scores, National Assessment Program - Literacy and Numeracy (NAPLAN), Trends in International Mathematics and Science Study (TIMSS), and the Programme for International Student Assessment (PISA).

Yet these measures are only substitutes for what we are really interested in — how much do schools influence student achievement. Because we can't measure the causal impact of schooling on student achievement directly, we can't know what impact schools have with any certainty. Instead, we must infer the causal impact of schooling on student achievement by way of proxy measures such as standardised student test scores which presents us with a significant epistemic challenge.

This paper introduces a novel approach to computer simulation to argue that the presence of choice or selectivity within a school system undermines measures of school performance based on student acheivement. In it, I develop a simple model in order to investigate the robustness of the inference from student achievement to school performance. I then demonstrate how the introduction of school choice or selectivity leads to a complete failure of the inference mechanism.

My claim is simple — if an inference is unreliable in an ideal setting, then it will certainly be unreliable in a non-ideal one. By eliminating measurement errors and abstracting away confounders, computer simulation allows us to create the ideal conditions from which to assess the quality of this inference. Thus, if the conditions in this model are present in the real world, then we should be very sceptical inferring anything about school performance from student achievement.

Introduction

There are a variety of ways we might measure school performance. One way is to *measure the things schools do* — the practices, pedagogies, and processes they apply. If their application of a process is faithful to some standard, then we might judge them to be performing well. If not, then we judge them poorly. But a more common approach is to *measure the things schools produce* —- how many students qualified for university, how many found well paying jobs, or especially, how well students performed in tests.

Measuring educational outputs is much simpler than measuring processes. However we measure the academic achievement of students — whether by *declarative knowledge*, the learning and appropriate recall of particular facts; or by *procedural knowledge*, the application of skill and know-how — we have an abundance of existing data available to us. The tools needed to obtain this data are also relatively easy to administer — standardised literacy and numeracy tests like NAPLAN (NAPLAN, 2015); the Peabody Picture Vocabulary Test (L. M. Dunn, Dunn, Bulheller, & Häcker, 1965); IQ scores or the Wechsler Intelligence Scale for Children (WISC-IV) (Wechsler, 2003). Yet the metrics from these tools are just proxies for what we really want to know — what is the educational impact of schooling upon students?

Accurately measuring school performance by proxy however, is hard. The challenges associated with inferring school performance from student achievement have long been documented. Ecological models, for example, stress the important role non-school factors like family and neighbourhood have on student performance (Bronfenbrenner, 1994), (Zubrick, Silburn, Williams, & Vimpani, 2000). Further clouding our understanding are the often competing causal theories that attempt to explain how schools affect student achievement - peer effect (Hanushek, Kain, Markman, & Rivkin, 2003), class size policies (Ehrenberg, Brewer, Gamoran, & Willms, 2001), or teacher training and qualifications (Kosgei, Mise, Odera, & Ayugi, 2013). How and to what degree schools themselves impact student achievement remains an open question. How much the changes in student achievement can be attributed to school performance is, therefore, also uncertain.

Measuring school performance is hard because measuring causation is hard. The limits of causal knowledge have been well documented in science and philosophy (see Schaffer (2016), Hitchcock (2019)). Causation cannot be observed directly — it can't be seen, heard, or touched. Neither can it be known *a priori*. Instead, causal connections must be *inferred* from their observable, posited effects (Hume, 1748). In order to discover the

causes of effects, we try to hold all but a few variables fixed, and observe the covariance between them in order to identify causality.

In complex systems, accurate causal inferences are especially challenging. Common causes, feedback loops, under-determination, over-determination, and causal indeterminacy all strain the certainty of our inference mechanisms despite the best controls, protocols, and experimental design we might put in place. Causal inference is difficult. Inferring the causal impact of a school on student achievement in a complex education system with multiple confounders is especially difficult. Yet infer it we must. So an important question we need to ask is *just how warranted is this inference from student achievement to school performance*?

In some scenarios, inferring school performance from student achievement might be perfectly justified. Changes in student achievement might be largely or even wholly explainable by a school's causal impact. In these instances, when students results improve, we can justifiably say that because the school's causal impact is largely responsible, the school performance has improved. In many other scenarios however, we might have serious grounds for scepticism. Confounders such as parental age and socio-economic status (Caro, McDonald, & Willms, 2009), birth weight, neighbourhood characteristics (Nghiem, Nguyen, Khanam, & Connelly, 2015), or even a student's breakfast consumption (Adolphus, Lawton, & Dye, 2013) might explain a great deal about differences in student achievement. In these cases, when the extent of a school's causal impact is uncertain, then we should question whether we can infer anything at all about school performance from changes in student achievement.

If inferring from observable effects to their causes is difficult, then judging the quality of the inference mechanism is even harder. Doing so requires some standard against which we can make comparisons but our knowledge of this standard is limited by the same problem of inference. If we could somehow know independently the causal impact of schools upon students, then we could judge the qualify of our inference from empiric data. But how can we judge the quality of the inference from student achievement to school performance if we can't be certain what is the cause of student achievement in the first place? We lack the epistemological foundations to properly ground our second order judgements.

This is a problem that empiric approaches cannot overcome. If we are to use empiric data from student achievement to make causal inferences about school performance, then we can't use that same data to judge the quality of the causal inference mechanism. If however, we are to use the same empiric data to judge the quality of the inference,

then we need to already know the cause — but if we already knew the cause, we wouldn't need to infer school performance via the proxy of student achievement in the first place. We need another way.

Methodology

Computer simulation offers us one way out of this problem. Simulation allows us to not only model how we think the world *is*, but to also specify how we think the world *should be*. It allows us to know the causal relationships *within* the model with certainty because they are explicitly stipulated in code *ex ante*. With causality known by stipulation *a priori*, we can then observe the empiric data the simulation generates *a posteriori* and assess the quality of the inference mechanism. We can judge how good the inference mechanism from student achievement to school performance really is - in our model at least.

The purpose of this simulation is therefore not to explain or predict the real world but to create a yard stick by which we can judge our causal inferences in the real world. The model simplifies and abstracts away all possible confounding causes of student performance. Importantly, it eliminates any errors when measuring student performance and stipulates that schooling is the *only* cause of changes in student performance.[1] This creates ideal conditions for assessing the quality of our second order judgements — the inference from student achievement to school performances.

Because the causal impact of a school on student achievement in the simulation is known by stipulation, we can assess how accurately the empiric data about student achievement generated from the simulation maps to the stipulated causal impact of the school — to our belief about school performance. If the inference mechanism from student acheivement to school performance is warranted in the simulation when the causal mechanism is known by stipulation, then we have some grounds to be confident about the inference mechanism in the real world, additional confounders not withstanding. *But if the inference isn't warranted when the causal mechanism is known, then it can't be warranted when the causal mechanism is unknown.* This simulation will therefore allow us to say when the student-acheivement-to-school-performance inference *might* be warranted, and when it *cannot be* warranted. It will give us a yard stick against which we can judge an inference as *possibly* or *definately not* warranted.

The methodological approach of this article is different from typical philosophical fare, even philosophical work involving computer simulation. Written in Literate

[Coffeescript](), this article is simultaneously a philosophical argument and a computer simulation that demonstrates the claims of the argument from within it. Literate Programming (Knuth, [1984]()) involves embedding computer code within a written argument and has a great deal to offer scholarly writing. The compter code and written argument are one.

Firstly, it ensures that all assumptions of the model are explicit. Computer programs are deterministic, so all the instructions necessary for the simulation to run have to be explicitly documented. Secondly, it allays concerns often directed at simulation — the challenges of validation and replication. Often, simulations are 'black boxes' of code, opaque to the reader and reviewer alike. Because all the code necessary for the running simulation is embedded in the article, replication is as simple as following the installation instructions in the appendix.

Lastly, it enhances the persuasiveness of argument. Compared to traditional static argument, the dynamic simulation offered in this paper allows the reader to experience the argument evolve from intial definitions to final conclusion as more premises are added. This is especially useful when dealing with counter-intuitive results that might require considerable concentration and reflection to accept.[2]

I make no assumptions about the reader's knowledge of computer programming. Simulation code is indicated by `monospace font and indented blocks`, is written to be as informative as possible, and is fully described in the surrounding text. Any code that is not germane to the argument itself has been placed in the appendix. Coffeescript is used because it runs in any modern web browser and has a syntax close to natural language. Best viewed in HTML at http://blind-review.github.io/school-performance/ to take advantage of the interactive visualisations, a static but less engaging version of this article is available in PDF or print.

A Model of School Performance

We begin by defining a model of students. Students have a level of academic achievement measured from a minimum of `0.0` to a maximum of `1.0`. Academic achievement is broadly construed and can be interpreted as any attribute of value that can be influenced by schooling. This achievement is randomly generated with a uniform distribution and is centred around a mean of `0.5`.

```
class Student
```

```
  constructor: () ->
    @achievement = Math.random()
```

Next we model schools. Schools are collections of students upon whose acheivement they causally impact. In short, *schools teach students*. The causal impact of a school on a student's academic performance ranges from `-1.0` (a strong negative impact) to `1.0` (a strong positive impact). This impact is the same for all students and is fixed for the duration of the simulation. Again, no claim is made about how educational impact is transmitted to student achievement in the real world. Policy, pedagogy, or the school environment are all possible causal mechanisms compatible with this model. The key stipulation is that `impact` is the *only* causal mechanism at work.

```
class School
  constructor: (@id, @impact) ->
```

Finally, we create a class to encapsulate the simulation itself. We will construct each simulations with a profile describing the desired attributes of the schools and student distribution. We then create 1000 students of random achievement and assign them in the schools according to the stipulated profile.

```
class Simulation
  constructor: (@profile) ->
    @schools = for school, id in @profile.schools
      new School(id, school.impact)
    @students = [1..1000].map () =>
      student = new Student()
      student.school = randomly_assign(student, @schools, @profi
      student
```

The initial assignment of students between schools is determined by a skew factor ranging from `0.0` (all students of below average achievement are in the first school) to `1.0` (all students of above average achievement students are in the first school). The default value of `0.5` results in an even distribution of students by achievement. Subsequent enrolments will be random.

```
randomly_assign = (student, schools, skew=0.5) ->
  [first, second] = if student.achievement > 0.5 then [0,1] else
  if Math.random() < skew then schools[first] else schools[secon
```

Events in the simulation will occur during a generic time period called a `tick`. A tick can represent any fixed period of time such as a term, semester, or year. During each tick schools will `teach` students, some students will `graduate`, and some new students

will `enrol`.

```
Simulation::tick = () ->
  @teach()
  @graduate()
  @enrol()
```

Schools causally impact a student's achievement — schools teach and students learn. Again, the actual causal mechanism the model describes could be a variety of influences such as pedagogy, policy, or classroom environment. By stipulation however, schooling is the *only* mechanism by which student achievement can change within the model. The causal impact on student achievement is assumed to be linear and a constraint is applied to ensure achievement remains between `0.0` and `1.0`.

```
Simulation::teach = () ->
  transference=0.2
  @students.forEach (student) ->
    student.achievement = student.achievement * (student.school.
    student.achievement = Math.min student.achievement, 1.0
```

During each tick, a percentage of students will graduate and be replaced by new students enrolling. We will set the graduation rate at 20% per tick.

```
Simulation::graduate = (graduation_rate=0.2) ->
  @graduates = @students.filter (student) ->
    Math.random() < graduation_rate
```

Identifying the 'best' school requires us to rank schools based on student performance so we create a simple league table that all students have knowledge of.

```
Simulation::rank_schools = () ->
  @measure_schools()
  [best, worst] = if @schools[0].score > @schools[1].score
  then [@schools[0], @schools[1]]
  else [@schools[1], @schools[0]]
```

Measuring school performance is then done by calculating the average student achievement in each school. This acts as an error free method of student assessment which removes concerns about measurement validation from the model. School performance will therefore perfectly track the academic achievement of the students enrolled.

```
Simulation::measure_schools = () ->
  @schools.map (school) =>
    students = @students.filter (student) ->
      student.school.id is school.id
    total = students.map (student) ->
        student.achievement
      .reduce (prev, curr) ->
        prev + curr
    school.score = total / students.length
```

Finally, we will introduce an optional *selectivity* parameter. If present, students will prefer to enroll in the best performing school, and schools will prefer to admit the best achieving students, so that all schools have roughly even numbers of students.

```
Simulation::enrol = () ->
  [best, worst] = @rank_schools()

  @graduates.forEach (student) =>
    student.achievement = Math.random()
    if Math.random() < @profile.selectivity
      student.school = if student.achievement > 0.5
      then @schools[best.id]
      else @schools[worst.id]
```

We now have a complete model of students with initial uniformly random `achievement`; schools whose stipulated `impact` is the only mechanism for affecting student achievement; a perfect measure of school performance as average student achievement; methods for some student to graduate and new ones to enrol; and a means of varying how much *selectivity* is present in the school system as well as the initial *skew* of above average students. In short, we have created perfect conditions for measuring a school's causal impact in order judge the quality of the inference from student achievement to school performance.
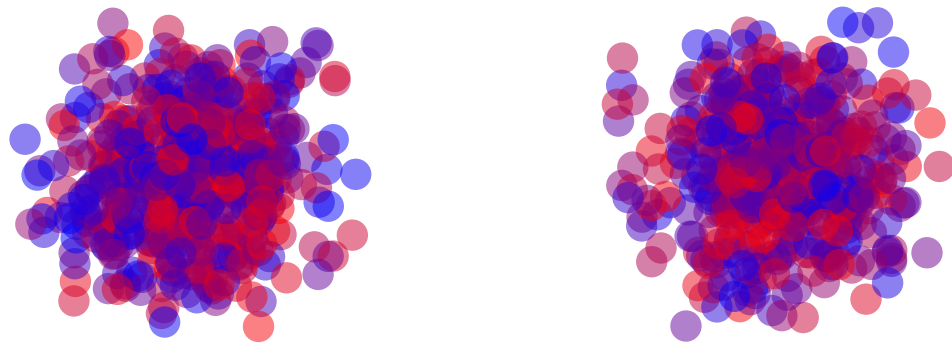
Simulations

With the model defined, we begin with a test to ensure the code is working as expected and ground our confidence in the simulation. We will begin with students split evenly between the two schools. We stipulate that schools have no causal impact on student achievement, and that there is no selectivity within the system. Because enrolment is completely random and there is no causal impact from schools, what we *should* observe is no significant change within and between schools.

The movement in the simulation visually represents a `tick`. Student achievement is

represented in colour - blue for > 0.5 and red for < 0.5 achievement. The profile of the simulation is expressed in the code below and the simulation can be started or stopped by clicking on it.

```
sanity_check_1 = {
  schools: [{impact: 0.0}, {impact: 0.0}],
  selectivity: 0.0
}
```
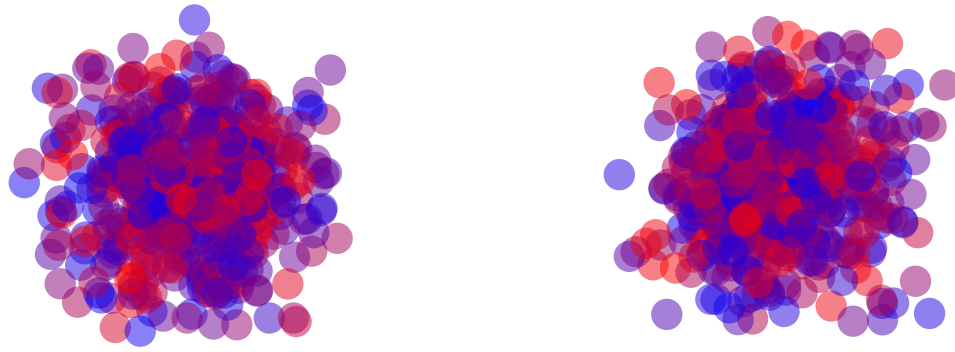

Simulation 1: No impact, no selectivity.

Running the simulation generates the expected results. During each tick, the `teach` method has no impact on student achievement because both schools have `0.0` impact. Some students graduate and are replaced with new students of random achievement. As such, school performance, measured by average student achievement, remains close to 0.5.

Next, we will model two schools with different causal impact - one positive and one negative - but again with no selectivity. While both schools will start with similar student achievement levels, the `teach` method *should* see student achievement in the first school increase while it decreases in the second.

```
sanity_check_2 = {
  schools: [{impact: 0.5}, {impact: -0.5}],
  selectivity: 0.0
}
```
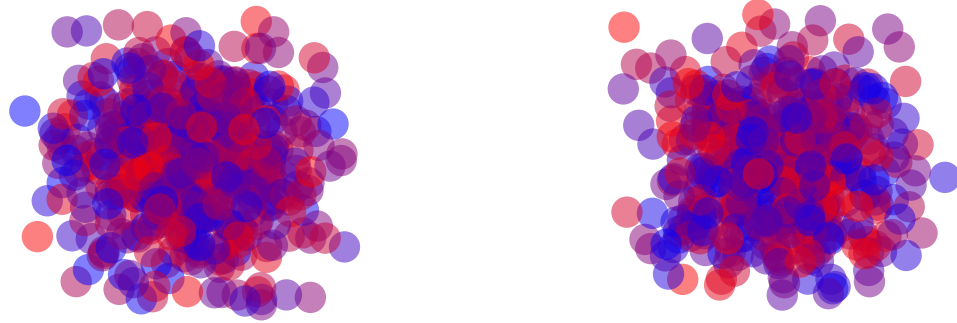
Simulation 2: School impact, no selectivity.

Again, the simulation generates the expected results with one school's performance stabilising at approximately `0.66` and the other at `0.33`. In both cases, the simulation performed as expected and our inference from student achievement to school performance is warranted. In these scenarios we can reliably infer school performance from student achievement.

School Performance as Shifting Averages

Most schools systems have some degree of selectivity. This might be selectivity on the part of the school where only the top achieving students are admitted; or it might be selectivity on the part of the students or parents who explicity choose one school over another. Selectivity can also be implicit when for example, parents choose a 'good' neighbourhood with 'good' schools to live in.

In the next scenario, we take parameters from Simulation 1 where schools have no causal impact on student achievement but introduce selectivity into the model. Because both schools are identical in terms of their causal impact, and school `impact` is the only causal mechanism in the model, *any changes in school performance must be the result of selectivity*.

```
shifting_averages = {
  schools: [{impact: 0.0}, {impact: 0.0}],
  selectivity: 0.5
}
```

Simulation 3: No impact but with selectivity.

While the performance of both schools is approximately `0.5` when the simulation begins, any minor imbalance in relative performance caused by random variation in new student achievement results in a run-away effect. As soon as one school is perceived to perform better than others, *selectivity* ensures that students who do selectively enrol, choose the school with the higher percentage of high achievement students. After a few ticks, the simulation stabilises with one school performing at approximate `0.63` and the other at `0.37`.

Recall however that the impact of schools is stipulated as zero - both schools are identical. Individual student achievement never changes because schools have no impact in this scenario. The large differences in school performance are therefore completely explained by *selectivity*. As selectivity and the percentage of students enrolling non-randomly increases, so too does the performance difference. Differences in school performance as measured by student achievement are the result of nothing more than shifting averages of student achievement. When selectivity is present, the inference from student achievement to school performance is unreliable.

Performance is Relative

The previous simulation showed how significant performance differences can arise even when schools are causally identical. The next simulations demonstrate how a school with a low or negative impact can appear to perform well above what it should. We begin with two schools having differing levels of negative impact but with no selectivity. In this case the average performance of both schools *should* decrease, propped up only by new enrolments whose average achievement is `0.5`.

```
relative_1 = {
  schools: [{impact: -0.25}, {impact: -0.5}],
  selectivity: 0.0
}
```
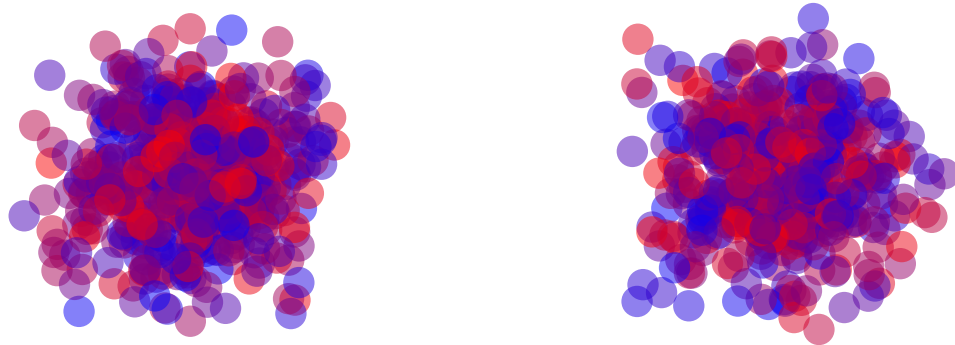

Simulation 4: Negative impact, no selectivity.

Again, the simulation results are as expected. Both schools performed poorly but the school with the lowest causal impact performed the worst. Next however, we introduce selectivity to the same schools and students.

```
relative_2 = {
  schools: [{impact: -0.25}, {impact: -0.5}],
  selectivity: 0.5
}
```
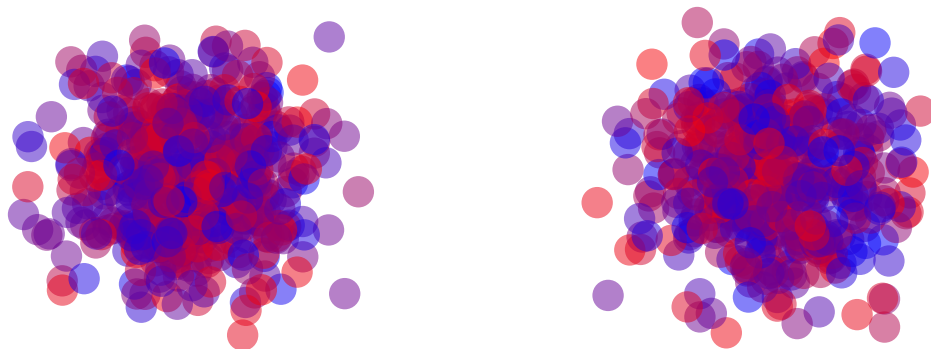

Simulation 5: Negative impact with selectivity.

When selectivity is present, the inference from student achievement and school performance breaks down again. The school with the negative but better causal impact performs far better than its causal impact says it should - increasing average student

achievement from `0.4` in Simulation 4 to `0.5` in Simulation 5. Meanwhile, the school with the worst causal impact performs much worse than it should - decreasing average student achievement from `0.33` to `0.22`. The schools remained the same in both simulations. The only change in the simulations, and therefore only possible cause of the change, was the introduction of selectivity.

As selectivity increases, so too does the disconnect between perceived performance and causal impact. The mere presence of a school with lower causal impact on student achievement combined with selectivity results in the higher achievement students enrolling in the *least worst* school, thereby *inflating* that school's apparent performance. Thus, school performance as measured by student achievement is determined not only by the causal impact of a school but also by the relative performance of other schools. This is not to say that school performance *is* relative, but rather that the performance of one school is affected by the performance of another, whenever selectivity is present. Again, the inference from student achievement to school performance is unreliable.

Initial Conditions Matter

In an egalitarian utopia, every child would have the same opportunities as every other. The biological lottery wouldn't affect educational outcomes. Obviously, this is not our reality — initial conditions matter. Until now, we have looked at scenarios where both schools started with similar initial random distributions of students. In the next simulations, we add a new parameter, `skew`, that alters the initial enrolment of students.

We continue with the same 1000 students of random achievement (mean `0.5`) but set the initial distribution between schools. A `skew` value of `0.75` here means that 75% of the above average, and therefore only 25% of the below average students, will initially be enrolled in the first school. The default value of `0.5` means there is an equal distribution of achievement.

In the next simulation, we stipulate that schools have different causal impact but the initial distribution of student achievement is skewed in favour of the school with the lowest impact. No selectivity is present so we should expect the initial mismatch between school impact and performance to be corrected over time.

```
head_start_1 = {
  schools: [{impact: -0.25}, {impact: 0.25}],
  selectivity: 0.0,
  skew: 0.75
```

```
}
```
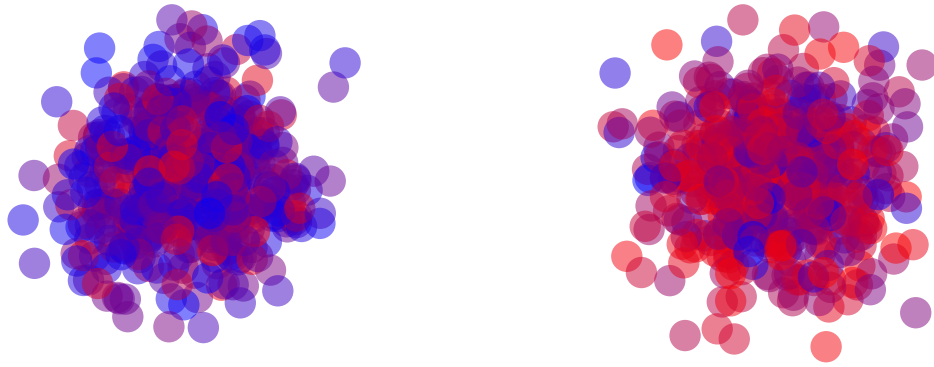

Simulation 6: Mixed impact with skew.

As expected, any imbalance in the initial distribution when selectivity is absent is eliminated given enough time. In each tick, the causal impact of schools affects student achievement, and every enrolling cohort normalises school performance to a degree. The first school, while appearing to perform strongest initially, eventually performs poorly, with the opposite occurring in the second school.

Next, we add selectivity to the same parameters and see how this changes performance.

```
head_start_2 = {
    schools: [{impact: -0.25}, {impact: 0.25}],
    selectivity: 0.5,
    skew: 0.75
}
```
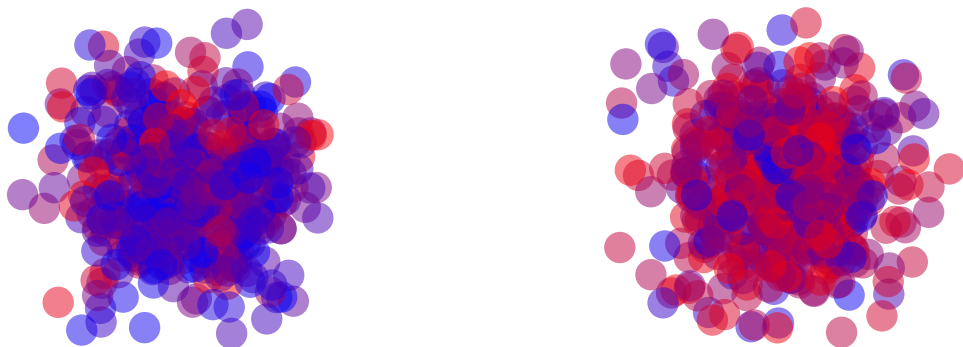

Simulation 7: Mixed impact with skew and selection.

Again, when selectivity is present school performance no longer reflects a school's

stipulated causal impact. A skewed initial distribution of students and selectivity results in a school with a negative causal impact performing better than a school with a postive causal impact. Given a sufficiently skewed initial distribution, schools with negative causal impact can continue to outperform schools with positive impact. Yet again, the presence of selectivity in the school system undermines the student achievement to school performance inference mechanism.

Discussion

Measuring school performance is important but we cannot measure it directly. Instead, we typically rely on student achievement as a proxy, and *infer* school performance. This of course, raises a number of epistemic challenges. Few people if any would claim that current school performance measurement regimes that rely on student achievement as proxy data from standardised tests such as NAPLAN are perfect. What these simulations demonstrate however, is that *if selectivity is present, then the inference from student achievement to school performance is unwarranted*. If a causal inference mechanism isn't warranted when causes are known with certainty, then it seems highly impausible that the inference can be warranted when causes aren't known or observable.

The inference from student achievement to school performance is poor even under the very best epistemic conditions. The model above allows us to stipulate the behaviour of much that is uncontrollable in the real world. Within the model, our knowledge of student achievement is perfect and not affected by test errors, strategic teaching, or exam stress. So too is our knowledge of the causal mechanism. As a non-ecological model, we have stipulated that the only way student achievement can change in the simulation is via school impact. We could not ask for more perfect conditions from which to *infer* causal impact yet even here the link between proxy evidence and actual cause breaks down. *We should therefore be very skeptical about inferring school performance from student achievement*.

This is of course, a purely theoretical argument. It remains an open empiric question just how well the simulation models reality, and how much choice or selectivity occurs within our school systems. Social scientists are likely to be better placed answering this question than philosophers are, but a few observations about the Australian context are germane. Approximately 35% of all students currently attend a non-government school (Australian Bureau of Statistics, 2014) indicating a minimum level of explicit selectivity. Then there is selectivity within government schools. Data on what percentage of students attend their *nearest* school is difficult to find, as is data on parents moving to different school catchment zones because of perceived school performance, but anecdotal

evidence suggests this too occurs.

One might object to my conclusion by claiming that the model described doesn't accurately reflect the real world. This is certainly true but not grounds to reject the conclusion. All models are abstractions and simplifications of reality - that is their strength. In this case, the model described presents a best case epistemic scenario for measuring school performance. As such, the simplification of the model *strengthens* the claims for epistemic skepticism towards school performance.

Despite its simplicity, the model also appears to conform with empiric observation. In the sanity check simulations, the model performed exactly as expected. When selectivity was introduced, the simulation was compatible with recent Australian data. Nghiem et al. (2015) analysis of the Longitudinal Study of Australian Children showed that non-government (and therefore selectively enrolled) schools had higher average student NAPLAN results but once controlled, performed equally with non-government schools or worse in the case of Catholic ones.

Some might also simply dispute my claim that student achievement is typically used for measuring school performance. The NAPLAN, Peabody, and Wechsler tests for exapmle, are intended for measuring student progress, not school performance. Longitudinal (Zubrick et al., 2000), predictive (Lavin, 1965), and psychological studies of academic performance (Pintrich & De Groot, 1990) for example, all rely on these measures without making claims about school performance. In short, one might argue that while my conclusion may be entailed, it is poorly targeted. Yet while it is certainly true that student achievement can be used for much more than assessing school performance, student achievement is still used to infer school performance by parents, teachers, and policy makers alike. NAPLAN as just one example is explicit on this - "[s]chools can gain detailed information about how they are performing, and they can identify strengths and weaknesses which may warrant further attention." (NAPLAN, 2015)

The impact of student intake and composition has long been acknowledged to play an important role in determining school performance (Thrupp, 1995, Thrupp (1999), Thrupp & Lupton (2006)). What this simulation so graphically demonstrates is how these common scenarios further complicate research into school performance by undermining our inference mechanisms even in ideal settings.

In short, we should be very skeptical about inferring school performance from student

achievement, especially when selectivity is present.

Appendix

This article is written in Literate Coffeescript. Literate programming has a great deal to offer the humanities, not least of which is that it makes replication available to all readers. To build the simulation from the raw paper, download the project from the repo and type `make paper` in the command line. Make sure you've got Coffeescript and Pandoc installed first.

Running a simulation with your own parameters is easy. You just need to specify the parameters and call `display(target-location, parameters)` and add a target location in HTML `id="target-location"`

Browser Code

Much of the browser code required in this simulation is not germane to the argument and so has been extracted here to the appendix. It provided here so that the entire simulation is self contained within the paper itself.

To make things look pretty, we will use the D3.js data visualisation library by Mike Bostock. We will also set some global variables from the browser such as height and width.

```
d3      = require 'd3'
width   = window.innerWidth || 600
height  = width * 0.4
```

Because we will be running multiple simulations in the browser, we will need a way of creating different ones. Here we define a `display` method for creating a simulation and binding it to a canvas with click events. When a simulation canvas is clicked, the interval runner starts and calls the `tick` method every 1000 milliseconds.

```
display = (id, params) =>
  runner = false
  simulation = new Simulation params
  canvas = d3.select("##{id}")
    .append("svg:svg")
    .attr "height", "auto"
    .attr "width", width
    .attr "preserveAspectRatio", "xMidYMid meet"
    .attr "viewBox", "0 0 #{width} #{Math.max(width * 0.4, heigh
    .on "click", () ->
      if runner
        clearInterval runner
        runner = false
      else
        runner = setInterval () ->
          tick simulation
        , 1000
  canvas.append('text')
    .attr "y", () -> height * .85
    .attr "x", () -> width * .33
```

In every tick cycle, we run call simulation's `teach` and `graduate` methods. We then `render` the simulation to calculate the x & y coordinates for the students, and update the canvas.

```
tick = (simulation) ->
  simulation.tick()
  render simulation
  draw canvas.selectAll "circle"
```

We then `draw` the students on the canvas. We will represent our students as coloured circles and schools by student proximity. We will also append the school averages in text form.

```
draw = (students) ->
  students.transition()
    .duration 1000
    .style "fill", (d) -> colour d, 'achievement'
    .style "opacity", 0.5
    .attr "r", 8
    .attr "cx", (d) -> d.x
    .attr "cy", (d) -> d.y
  canvas.select "text"
    .text "#{simulation.schools[0].score.toFixed(5)}
      - Average Student achievement
      - #{simulation.schools[1].score.toFixed(5)}"
```

Finally, we setup the initial rendering.

```
setup = () ->
```

```
      render simulation
      students = canvas.selectAll "circle"
        .data simulation.students
      students.enter()
        .append "circle"
        .style "fill", (student) ->
          colour student, 'achievement'
        .style "opacity", 0.5
        .attr "r", 8
        .attr "cx", (d) -> d.x
        .attr "cy", (d) -> d.y

    setup()
```

In the browser code above, we have relied on a few helper methods. The first of these represents student achievement graphically using colour. Blue represents high achievement and red low achievement. With a little bit of maths, we can convert achievement on a range of 0.0 to 1.0 to a hexadecimal representation of Red-Green-Blue colour.

```
colour = (d, attribute) ->
  red = Math.floor( (1-d[attribute])*255 ).toString 16
  blue = Math.floor( d[attribute]*255   ).toString 16
  red = "0#{red}" if red.length is 1
  blue = "0#{blue}" if blue.length is 1
  "##{red}00#{blue}"
```

To display students and indicate school enrolment by student proximity, we create a Gaussian overlay so all students in the same school clump together. Each student is assigned a randomised position centred on their school.

```
render = (simulation) ->
  simulation.schools.map (school) ->
    school.x = width * (0.3 + 0.5 * school.id)
    school.y = height * 0.6
  simulation.students.map (student) ->
    student.x = gausian(width/1.2) + student.school.x
    student.y = gausian(height/0.5) + student.school.y

gausian = (range) ->
  Math.random()*range/8 + Math.random()*range/8 + Math.random()*
```

Finally, we trigger the various simulations outlined in the article once all the scripts have loaded.

```
window.onload = () ->
  display 'sanity-check-1', sanity_check_1
  display 'sanity-check-2', sanity_check_2
  display 'shifting-averages', shifting_averages
  display 'relative-1', relative_1
```

```
display 'relative-2', relative_2
display 'head-start-1', head_start_1
display 'head-start-2', head_start_2
```

Bibliography

Adolphus, K., Lawton, C. L., & Dye, L. (2013). The effects of breakfast on behavior and academic performance in children and adolescents. *Frontiers in Human Neuroscience*, *7*.

Australian Bureau of Statistics. (2014). *4221.0 - schools, australia, 2014 dataset*. Retrieved from http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup /4221.0Main+Features12014?OpenDocument

Bronfenbrenner, U. (1994). Ecological models of human development. *Readings on the Development of Children*, *2*, 37–43.

Caro, D. H., McDonald, J. T., & Willms, J. D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education/Revue Canadienne de L'éducation*, *32*(3), 558–590.

Dunn, L. M., Dunn, L. M., Bulheller, S., & Häcker, H. (1965). *Peabody picture vocabulary test*. American Guidance Service Circle Pines, MN.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, 1–30.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*(5), 527–544.

Hitchcock, C. (2019). Causal models. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 2019). Retrieved from https://plato.stanford.edu/archives/sum2019 /entries/causal-models/

Hume, D. (1748). *An enquiry concerning human understanding*.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Knuth, D. E. (1984). Literate programming. *The Computer Journal*, *27*(2), 97–111.

Kosgei, A., Mise, J. K., Odera, O., & Ayugi, M. E. (2013). Influence of teacher characteristics on students' academic achievement among secondary schools. *Journal of Education and Practice*, *4*(3), 76–82.

Lavin, D. E. (1965). *The prediction of academic performance.*

NAPLAN. (2015, October). Why nap. Retrieved from http://www.nap.edu.au/about/why-nap.html

Nghiem, H. S., Nguyen, H. T., Khanam, R., & Connelly, L. B. (2015). Does school type affect cognitive and non-cognitive development in children? Evidence from australian primary schools. *Labour Economics*, *33*, 55–65.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33.

Schaffer, J. (2016). The metaphysics of causation. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2016). Retrieved from https://plato.stanford.edu/archives/fall2016/entries/causation-metaphysics/

Thrupp, M. (1995). The school mix effect: The history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education*, *16*(2), 183–203.

Thrupp, M. (1999). *Schools making a difference: School mix, school effectiveness, and the social limits of reform*. McGraw-Hill Education (UK).

Thrupp, M., & Lupton, R. (2006). TAKING school contexts more seriously: THE social justice challenge. *British Journal of Educational Studies*, *54*(3), 308–328. https://doi.org

[/10.1111/j.1467-8527.2006.00348.x](/10.1111/j.1467-8527.2006.00348.x)

Wechsler, D. (2003). Wechsler intelligence scale for children–Fourth edition (wisc-iv). *San Antonio, TX: The Psychological Corporation*.

Zubrick, S., Silburn, S., Williams, A., & Vimpani, G. (2000). *Indicators of social and family functioning: Final report*. Commonwealth Department of Family and Community Services, Canberra. Available at http://www. facs. gov. au/internet/facsin ternt. nsfl aboutfac si programs/families/isff. htm,[2000, 6 Sept.].

---

1. Again, the claim is not that schooling is the only causal factor in student performance in the *real world*, but that it is the only causal factor in the model because it has been programmed as such.↵

2. This is not to say that persuasion should trump logical entailment but that a persuasive and valid argument is better than a valid one alone. Kahneman (2011) (p9) for example attributes the broad appeal of his and Tversky's work on heuristics biases to the inclusion of demonstrations that the reader could experience themself.↵