As humans, we communicate with each other readily and effortlessly, transmitting generations' worth of hard-won knowledge in a single conversation. Both children and adults use these abilities frequently, and so **it is natural to assume that we are quite good at teaching.** But while some recent work indicates that both children and adults accurately teach in situations where they must select one (and only one) piece of information to show to learners,[1,2,3] teachers' abilities to provide maximally informative data seems to decline over longer interactions. For example, tutors often overestimate how much correct knowledge learners possess, and thus have difficulties assessing and closing knowledge gaps.[4] Because prior work has not directly assessed the reasoning underlying teachers' responses, **it is still unclear why humans teach optimally in some situations and not in others**. My research will integrate computational and behavioral methods in order to investigate both the cognitive mechanisms at the root of our impressive teaching abilities, as well as pitfalls that may cause us to teach suboptimally.

Pilot: In order to establish a method where there is **variance in learner outcomes** (and thus to examine the mechanisms underlying human teaching), I developed a short causal learning task. Adult participants (n = 11) recruited from Amazon Mechanical Turk (mTurk) assumed the role of teachers. They were presented with 5 blocks, each painted a different color and lettered A-E, and a "light-up machine". Subjects then learned a B*and*E rule: only the presence of both B *and* E caused the machine to activate. Then, we asked participants to teach this rule to a hypothetical learner, by demonstrating whether combinations of blocks turned the machine on or off. We showed the teaching demonstrations of these teachers to learners (n = 11), and tested their understanding of the B*and*E rule. Less than half of our participants learned the B*and*E rule from the teachers' demonstrations. This result is likely a combination of some teachers' failures to provide effective data, as well as some learners' misinterpretation of the data provided. Given variance in participants' performance, my current research will address both possibilities, investigating the reasoning underlying the failures of both teachers and learners.

Study 1: Why do some teachers fail to convey the B*and*E rule, and some succeed? In order to teach efficiently, teachers need to not only infer a learner's knowledge, but also decide how to act to further inform the learner. Do teachers fail to accurately *capture* changes in a learner's hypothesis space? Or, are teachers able to infer and represent the learner's hypothesis space, but fail to *use* this information to decide what to teach?

Methods will be similar to the pilot, except that an adult in-lab sample will be assigned to either a *pilot replication*, *hypothesis generation*, or *hypothesis given* condition. In order to investigate whether some of the variance in teaching performance is actually a function of *learners'* abilities, each teacher's demonstrations will be shown to multiple learners. After selecting each demonstration for learners, *generation* participants will indicate which hypotheses a learner might still hold to be true. This provides a measure of teachers' abilities to capture learners' shifting hypothesis spaces. Because it is difficult to smoothly pair a teacher and learner in real-time while also asking each about their teaching choices/current hypotheses, in the *given* condition, we will use a **computational model** to *give* teachers the learner's current hypothesis space. This provides a measure of teachers' abilities to *use* correct information, once they have it. As before, teachers' demonstrations will be shown to a separate set of learners, who will specify which hypotheses they are still considering after seeing each demonstration. This will provide a measure of how the learner's hypothesis space is *actually* changing over the course of the teaching demonstration, and should dovetail with the predictions of our model.

Teachers' results will be compared to the *pilot replication* baseline condition, where teachers were not given or asked to generate learners' hypotheses. If teachers improve in the *generation* condition, this pattern of results will provide evidence that asking teachers to reflect upon their learner's knowledge states is an effective tool in improving teacher performance. If teachers improve in the *given* condition, this pattern will indicate that some teachers might have difficulties generating the learner's hypothesis space (but are able to act upon the information, once they have it). If teachers fail to improve in both conditions, this result will indicate that some other factor underlies gaps in teachers' abilities.

Another possibility is that teachers might differ in the extent to which they engage their Theory of Mind (TOM) capacities as they teach. Perhaps some teachers rely upon simpler heuristics (e.g., "this is how I was taught, so I will teach this way"), whereas others truly attempt to reflect upon the conclusions their learners might draw from data. Given that very young children do not have an explicit TOM, Study 2's developmental sample will not only shed light upon the development of our abilities, but also the cognitive capacities that may underlie them.

Study 2: What cognitive capacities support our teaching abilities, and how do these abilities develop across the lifespan? Young children are impressive learners, and there is evidence that they are efficient teachers in simple situations.[e.g., 3] Are teaching abilities (and thus the source of teaching errors) static over the life span, or do children display a different set of deficits, perhaps tied to their lack of explicit TOM?

We will replicate Study 1 with children aged 4-7 years, a developmental sample with a different range of explicit TOM abilities. Child participants will be asked to teach a puppet, who in the *given* condition will "go away" under the table when the experimenter explains the puppet's knowledge. Children's explicit TOM understanding will also be assessed.

We will then test whether children's performance on these explicit TOM measures predicts some of the variance in children's teaching performance. If children without an explicit TOM produce errors of a different type, this will indicate that TOM might support our teaching abilities. We will also investigate *which* abilities are supported by TOM; for example, TOM might only affect participants' abilities to *infer* others' knowledge (but not *use* knowledge they are given). In this case, we would expect children in the *given* condition to perform better than children in the *generate* and *pilot replication* conditions. In addition, we will explore whether some adult participants' errors might stem from a lack of TOM engagement. For example, if some adult teachers tend to make the same type of systematic errors as children who lack explicit TOM capacities, this result will provide support for the hypothesis that some adults may fail to engage their full TOM abilities when they teach. If, however, teachers produce the same frequency and types of errors across development, irrespective of other factors such as TOM, this will indicate that the ability to teach develops very early, and does not rely upon explicit TOM.

Conclusion: There is a divide in the pedagogical literature: in some situations, humans are impressive teachers, and in others, appear to have difficulty teaching effectively. In **future research** I will continue investigating the roots of teachers' (and learners') difficulties in order to understand exactly what factors underlie these disparate findings. This research will also explore whether there are **strategies** that can target these sources of difficulty, thus **improving teacher and learner performance**. The results of this work will not only inform our understanding of effective pedagogy, but will also deepen our understanding of humans' TOM (which appears to support many of our unique cognitive abilities) more generally.

1) Shafto, Goodman & Griffiths (2014). *Cog Psych*, *71*, 55-89. **2)** Rhodes, Gelman & Brickman (2010). *Dev Sci, 13*(3) 421-429. **3)** Gweon, Shafto, & Schulz (2014). *Proc Annu Conf Cog Sci Soc*. **4)** Chi, Siler & Jeong (2004). *Cog and Instr, 22*(3), 363-387. **5)** Gopnik & Sobel (2000). *Child Dev, 71*(5), 1205-1222. **6)** Ruggeri & Lombrozo (2015). *Cognition,* 143, 203-216. **7)** Gershman, Horvitz & Tenenbaum (2015). *Science, 349*(6245), 273-278.