12

What Do Human Tutors Do?

Kurt VanLehn

Abstract

People teaching an agent or robot might use the same methods that they use when tutoring a human student. Because teaching agents and robots is a central topic of this Ernst Strüngmann Forum, this chapter reviews research that characterizes human tutoring. Most of this research was done to improve the design of computer-based tutoring systems, which were assumed to be inferior to human tutors. However, it turns out that human tutors and a certain class of tutoring systems actually behave quite similarly, and their effectiveness is about the same. This chapter begins with a description of prototypical human tutoring behavior before discussing some common hypotheses about human tutoring behavior, which turn out to be unsupported by studies. It concludes with an attempt to synthesize these descriptions and apply them to the goals set forth at this Forum.

Introduction

Our discussion at this Ernst Strüngmann Forum focused both on a person teaching a bot (a robot, softbot, or agent) and a bot teaching a person. To understand both approaches, we need to have a general understanding of how a person behaves when teaching another person. Thus, in this chapter I review what is known about human–human tutoring.

I use the term "tutor" to refer to an adult domain expert who works with a single student, but note that in some literatures, "tutor" refers to someone working with a small group of students while in others, *peer* tutoring indicates a situation where one student teaches another. Here I review only studies where the tutors are adults who have expertise in the knowledge to be taught, and they are working one-on-one with a human student. This seems the most likely analogue to a human teaching a bot or a bot teaching a human. Further, I will discuss both expert and novice tutors. Both are domain experts, but an expert tutor has considerably more experience *as a tutor* than a novice.

I will refer only to studies that have sought general properties of human tutoring. These have appeared mostly in the cognitive science literature or in the intelligent tutoring systems literature. In contrast, many disciplines have studies that address the effectiveness of human tutoring compared to ordinary instruction in that discipline (e.g., Bausell et al. 1972; Marston et al. 1995; Scruggs and Richter 1985; Shanahan 1998; Wasik 1998; Wasik and Slavin 1993) or the occurrence of discipline-specific tutor behaviors. As an example of the latter, Juel (1996) analyzed dialogues of human reading tutors to find out, among other things, how frequently they asked tutees to sound out words.

What do human tutors do? This very question requires a description as its answer. Although some of the research reported here formulates and tests specific hypotheses, most of the research is qualitative and descriptive. Thus, this review is qualitative and descriptive as well. I begin by describing, in broad strokes, what human tutors do. Thereafter I analyze what a few exceptional tutors sometimes do, but most tutors do not do, before comparing the effectiveness of human versus computer tutors. Finally, I attempt a synthesis of learning and tutoring, and speculate on its implications for instruction of bots.

A Prototypical Session with a Human Tutor

A prototypical tutoring session can be viewed as a setting in which participants work through a sequence of *tasks*. A task could be solving a problem, studying an example, or answering a question. A task could even involve the student reading a text, while the tutor watches and helps.

A key fact is that most human tutoring supplements classroom instruction. Because the tutees are also taking a course in school, the student's teacher (i.e., not the tutor) usually determines the curriculum, so students often arrive at a tutoring session with a list of tasks to do (i.e., their homework) or a topic that requires further explanation. If they bring only a topic, then the tutor often selects tasks from a standard resource, such as a textbook, rather than inventing new tasks on the spot. In short, tutors prefer not to do on-the-spot instructional design, so they seldom invent new tasks. In this respect, human tutors play a very different role than a human instructing a bot. Presumably, bot instructors would need to invent their own tasks, because a school for bots with homework assignments seems highly unlikely!

Once the tasks to be accomplished during the session have been determined, the tutor takes control. After analyzing 72 one-hour tutoring sessions, Graesser et al. (1995:500) concluded that "the students never set the agenda for the tutoring session. Thus, the tutors carried the burden of setting the agenda, introducing subtopics, and proposing problems to solve." Typically the tutor works through one task after another. What the tutor does during each task depends on the nature of the task. Let us consider several common tasks.

Reading

Let us assume that the student is asked to read a short text, such as a one-page description of mitosis. Although the tutor could stop the student after each paragraph to ask questions (Chi et al. 1994), tutors typically let the student read silently until the end of the reading.

Answering Complex Questions

Now let us assume that the student is given a complex question to answer. Such questions are commonly assigned as homework. Tutors sometimes interrupt other tasks to ask the student complex questions. The students' behavior appears to be similar regardless of whether the question is a task in itself or an interruption of a task.

Graesser et al. (1995) characterize human tutoring of such questions as exhibiting a five-step pattern:

1. The tutor asks the student an open-ended question, such as "What happens when a computer boots up?"
2. The student gives an initial answer: "It starts up the windows."
3. The tutor gives some brief feedback: "Yes. Good."
4. The tutor then conducts a subdialogue to extract a better answer from the student:
    - Tutor: "What is the software called that has those windows?"
    - Student: "Windows"
    - Tutor: "I meant the type of software."
    - Student: "Do you mean operating system?"
    - Tutor: "Yes. So when you boot the computer, it starts the operating system. From where does it get the operating system?"
5. The tutor typically ends the dialogue with an evaluative comment or question:
    - "So do you understand 'booting up' now?"
    - Student: "Yes"

Although the five-step frame is common, novice tutors have a tendency to cut it short and replace step four with a brief lecture (Glass et al. 1999; Kim et al. 2005).

Studying Worked Examples

Studying a worked example (i.e., a multistep process needed to perform a task or solve a problem) is another task type. Consider the following example in physics:

Problem: Suppose a roller coaster is at rest on a track 200 meters above ground. The track tilts, and the coaster zooms down the first drop. Its descent is so steep and the wheels are so well lubricated that friction can be ignored. What is its speed when it reaches ground level?

Solution: This problem can be solved by applying conservation of mechanical energy.

At the start of the fall, the coaster's mechanical energy is entirely gravitational potential energy, because its velocity is zero, so

$$E_1 = m \times g \times h.$$

At the end of the fall, the coaster's mechanical energy is entirely kinetic energy, so

$$E_2 = m \times v^2.$$

By conservation of mechanical energy,

$$E_1 = E_2$$

so

$$m \times g \times h = m \times v^2$$
$$g \times h = v^2$$
$$v = \text{sqrt}\ (g \times h) = \text{sqrt}\ (9.8 \times 200) = 44.7\ \text{m/sec}.$$

Each line in the solution can be considered a step in solving the problem. If students explain each line thoroughly, they will learn well from studying the example (Chi et al. 1989). Although human tutors can be taught to prompt for such explanations after each line, most prefer to lecture instead (Chi et al. 2001).

Solving Problems

Tutors often spend most of their time helping students solve multistep problems, like the physics problem explained above. Their behavior varies with the competence of the student in a pattern widely known as *model-scaffold-fade* (Collins et al. 1989). When students lack the competence to solve a problem themselves, then tutors "model" how to do it by executing all or most of the steps. As they do a step, tutors usually explain it and then ask: "Do you understand?"

Once students gain even a little competency, tutors will let them solve the problem while giving them several kinds of "scaffolding." Forms of scaffolding include the following:

- The *prompting* of steps, especially difficult steps, through hints given before the student has even attempted to do the step. This is a natural means of scaffolding the task performance of bots.
- Giving *feedback* after a student's initial try, by indicating whether the attempt was correct or incorrect. The tutor may elaborate on this feedback and explain why the attempt was correct or incorrect. When giving negative feedback, tutors seldom say "wrong" or some other explicit negation. Instead, they use circumlocutions or even just pause a

moment before starting in on an explanation (Fox 1991, 1993; Rose et al. 2003). This particular habit, of giving negative feedback in a subtle, face-saving way, may be problematic when humans tutor bots.

- Providing *hints* when the student asks for one, or when it appears that the student is struggling to enter a step but has not yet asked for a hint.
- Asking students open-ended *questions*. Here a tutor might ask the student to explain a correct step or a concept related to it. In one study of tutored problem solving, 18% of the tutor turns were open-ended questions (Rose et al. 2003).

As students gain competence, the tutor removes or "fades" the scaffolding. Prompting is often the first form of scaffolding to disappear, but feedback may be withdrawn as well. Tutors can often determine, after just a few minutes of tutoring, how much scaffolding their tutee needs (Siler and VanLehn 2015).

When a problem is finished, tutors will sometimes review the solution with the student (Cho et al. 2000; Katz et al. 2003). For instance, if the student makes an incorrect step during the tutoring session, and the tutor believes it was due to a serious misunderstanding, the tutor may simply correct the step for the student and then, during the post-solution review, discuss the step and the misunderstanding.

Novice and expert human tutors vary in how they tutor. Novices tend to do more modeling (explaining) whereas experts tend to let the student do as much of the work as possible (Glass et al. 1999; Wood et al. 1976). Novice tutors sometimes give too much prompting, which reduces errors and learning (Kim et al. 2005; VanLehn et al. 2003).

What Human Tutors Seldom Do

The behaviors described above strike many people, including some participants at this Forum, as surprisingly mundane. People often think that human tutors engage in much more sophisticated behaviors, such as the ones listed below. Although at least a few tutors performed as hypothesized, research has shown that their behavior was not common enough to account for the effectiveness of human tutoring.

Detailed Diagnostic Assessments

In a few early studies of human tutoring (Collins 1977; Stevens and Collins 1977; Stevens et al. 1979), tutors debugged students. These studies involved tutors who went through a script that asked the student about the causes of heavy rainfall in various parts of the world. When a tutor uncovered a student's misunderstanding, the tutor conducted a misconception-specific subdialogue that resulted in removal of the misconception. This tutoring strategy is often called "diagnose and remediate." This strategy, however, appears to be uncommon, and is perhaps even unique to inquiry instruction, which is the instructional technique studied by Collins and Stevens. Let us consider, in turn, diagnosis remediation.

Although human tutors in other domains usually know which *correct* knowledge components their tutees have not yet mastered, tutors rarely know about their tutees' *misconceptions, false beliefs,* and *buggy skills* (Chi et al. 2004; Jeong et al. 1997; Putnam 1987). Moreover, human tutors rarely ask questions that could diagnose a student's specific misconception (McArthur et al. 1990; Putnam 1987). Thus, although tutors can try to uncover misconceptions, they rarely do so, and thus rarely know what misconceptions their tutees harbor.

Several studies asked whether human tutors could remediate when given a diagnosis. When human tutors were told which *correct* knowledge components were not yet mastered by their tutees, their behavior changed and their tutoring became more effective (Wittwer et al. 2010). On the other hand, tutors typically do *not* change their behavior nor become more effective when given detailed diagnostic information about their tutee's misconceptions, bugs, and false beliefs (Sleeman et al. 1989). Siler and VanLehn (2015) found that human tutors who worked with the same student for an extended period, and could thus diagnose their tutee's strengths, weaknesses, and preferences, were *not* more effective than when they rotated and had little familiarity with their tutees.

Human tutors do not seem to infer an assessment of their tutee that includes misconceptions, bugs, or false beliefs, nor do they seem to be able to use such an assessment when it is given to them. Instead, they often infer an assessment of which *correct* conceptions, skills, and beliefs the student has not yet mastered, and are able to use such an assessment when it is given to them. In short, tutors attempt to fill in the missing correct knowledge of their tutees, but do not attempt to remove their incorrect "knowledge." In this respect, human tutors operate just like computer tutors that use an *overlay* model of the student, where the overlay model represents only the presence or absence of correct knowledge (VanLehn 1988, 2008).

Adaptive Task Selection

Another hypothesis is that human tutors select tasks adaptively. That is, they give the student a task that is just what that student needs at that time. Studies suggest, however, that when human tutors are not working off a set of tasks brought to the tutoring session by their tutee, the human tutors select tasks using a *curriculum script*, which is a sequence of tasks ordered from simple to difficult (Chi et al. 2008; Graesser et al. 1995; Putnam 1987). Human tutors use their assessment of the student's mastery of correct knowledge to regulate how fast they move through the curriculum script. For example, tutors typically move to the next problem type after students correctly answered two or three problems of the current type (Putnam 1987).

Sophisticated Tutorial Strategies

A common early belief was that the power of human tutoring lay in their use of sophisticated strategies, such as Socratic irony (Collins and Stevens 1982), wherein students who give an incorrect answer are led to see that their answer entails an absurd conclusion. Other sophisticated strategies include reciprocal teaching (Palinscar and Brown 1984), inquiry (Collins and Stevens 1982), or authentic anchored cases (Goldman et al. 1993).

Studies of human tutors in many task domains with many degrees of expertise indicate, however, that these sophisticated strategies are rarely used (Cade et al. 2008; Chi et al. 2001; Cho et al. 2000; Core et al. 2003; Evens and Michael 2006; Fox 1991, 1993; Frederiksen et al. 2000; Graesser et al. 1995; Hume et al. 1996; Katz et al. 2003; McArthur et al. 1990; Merrill et al. 1995; Merrill et al. 1992; Ohlsson et al. 2007; VanLehn 1999; VanLehn et al. 2003).

A related belief is that human tutors sometimes let students make mistakes so that the students can practice finding them. However, tutors rarely let this happen (Fox 1991, 1993; Frederiksen et al. 2000; Merrill et al. 1995). If they ignore a student's error, it is often because the error is trivial and would have no impact on subsequent problem solving (Merrill et al. 1995).

These are statements about frequency and not about capability. For instance, from the Collins and Stevens (1982) studies it is clear that human tutors can, and sometimes do, use sophisticated strategies. Chase et al. (in prep.) gave human tutors the job of scaffolding an invention task and observed that they frequently employed more sophisticated tutoring practices than typical tutors. The demands of the task may have a large impact on the type of tutoring observed.

Questions Asked by Students

Human tutoring allows mixed initiative dialogues, as students can ask questions or even change the topic. This contrasts with most computer-based tutoring systems, where student initiative is highly constrained. For instance, although students can ask a typical tutoring system for help on a step, they are not able to ask other questions, nor can they cause the tutor to veer from solving the problem. On the other hand, students are free to ask any question of human tutors and to negotiate topic changes with the tutor.

Analyses of human tutorial dialogues show that although students take this initiative more often than they do in classroom settings, the frequency is still low (Chi et al. 2001; Core et al. 2003; Graesser et al. 1995). For instance, Shah et al. (2002) found only 146 student initiatives in 28 hours of typed human tutoring (i.e., tutor and student communicated over a chat connection), and in 37% of these 146 instances, students were simply asking the tutor whether their statement was correct (e.g., by ending their statement with "right?") or establishing common ground (e.g., "Did you mean condition two?"). That is, there were about 3.3 nontrivial student questions per hour. Participants were medical students being tutored as part of a high-stakes physiology course, so apathy is not a likely explanation for this low rate. With high school and undergraduate students, Graesser et al. (1995) found that 71% of the student questions were trivial, so there were only 7.7 nontrivial questions per hour. In short, even though students can ask nontrivial questions of human tutors, they seldom do.

Motivational Comments

Human tutoring is often believed to increase the motivation of students. Episodes of tutoring that aim to increase students' motivation certainly do occur in human tutoring (Cordova and Lepper 1996; Lepper and Woolverton 2002; Lepper et al. 1993; McArthur et al. 1990), but their effect on student learning is unclear.

For instance, consider praise, which Lepper et al. (1993) identified as a key tutorial tactic for increasing motivation. One might think that a human tutor's praise increases motivation, which increases engagement, which increases learning. However, the effect of human tutors' praise on tutees is actually quite complex (Boyer et al. 2008; Henderlong and Lepper 2002). In Kluger and DeNisi's (1996) meta-analysis of hundreds of feedback interventions, praise had a small negative effect (d = –0.17). Kluger and DeNisi (1996:275) conclude:

> The debilitating effects of praise on performance received some direct experimental support both in the laboratory and in the field and were explained, respectively, by a model of self-attention (Baumeister et al. 1990) and by control theory (Waldersee and Luthans 1994). These findings are also consistent with a review of field studies (many of which did not qualify for the meta-analysis) that concluded that "praise may not be widely effective as a reinforcer."

There are many interventions that increase motivation, such as convincing students that their intelligence is malleable rather than fixed (Dweck 1986). The effects are often quite large compared to the brevity of the intervention (Lazowski and Hulleman 2016). However, most of these studies involve interventions with large groups of students, such as a whole class. It is unclear what the impacts would be if these interventions were implemented by human tutors (or computer tutors). This is clearly an area where more research is needed.

Habits of Tutors That Bots Should Be Designed to Handle

Human tutors, especially novice human tutors (Glass et al. 1999), utilize certain habitual dialogue to support the self-monitoring and self-correction processes in a student. Looking toward future encounters between a human tutor and a bot, designers need to be aware of this and reflect it in the design of bots.

Self-Monitoring: "Do You Understand?"

Tutors often ask students: "Do you understand?" Such questions often occur at the end of the five-step frame (see above), when students have finished a reading, or when they have finished studying an example (Glass et al. 1999; Graesser et al. 1995).

A student's answer reflects their ability to monitor their own understanding (self-monitoring). Chi et al. (1989) divided their students into Good and Poor, based on learning gains, and found that while the self-monitoring statements of Good students were 45% negative and 54% positive, the self-monitoring statements of Poor students were 15% negative and 85% positive. This suggests that the Good students were good at self-monitoring and thus accurate at reporting their degree of understanding, but that the Poor students were not good at self-monitoring and their default was to assume that they understood. Graesser et al. (1995) found the same pattern when self-monitoring statements were elicited by the tutor's question. These findings reflect the general challenge that many learners have with self-monitoring (Glenburg et al. 1982).

Because bots are likely to get such questions from tutors, they should have good self-monitoring capabilities. Moreover, they should be able to act like Chi et al.'s (1989) Good students, whose negative self-monitoring statements were quite specific (e.g., "I'm wondering whether there should be acceleration due to gravity?") compared to the Poor student's self-monitoring statements (e.g., "What should I do now?").

Negative Feedback That Saves Face and Encourages Self-Correction

Although human tutors are quite accurate at spotting errors in the students' work (Evens and Michael 2006; Merrill et al. 1995), they often will not give explicit negative feedback. For example, Graesser et al. (1995) found that tutors directly acknowledged student errors only 24% of the time. Fox (1993) and Lepper et al. (1993) found that human tutors prefer dialogue filled with pauses, prompts, and questions which encourage the student to notice and repair the error themselves. In addition to giving the student practice in self-correction, they hypothesized that the circumlocutions minimize the impact of negative feedback on motivation and allow the student to "save face."

Merrill et al. (1995) found that the proportion of explicit versus implicit negative feedback depended on the type of error. If the error was clearly an unintentional slip, such as a typo, and the students did not themselves detect it, then the tutor explicitly indicated the correction. If, however, the error was plausibly due to a misunderstanding or lack of domain knowledge, then the tutor often used more indirect dialogue to allow the student to do as much of the error detection and correction process as possible.

This is not to say that human tutors fail to give negative feedback. McArthur et al. (1990) found that *every* error was followed by some kind of remedial dialogue. Merrill et al. (1995) found that 95% of student errors which were not immediately caught by the student were followed by some kind of remedial dialogue by the tutor.

Looking toward human tutor–bot interactions, a bot can count on human tutors giving negative feedback on almost every error, but the *way* in which tutors give the negative feedback might make it hard for a bot to recognize it as such.

The Effectiveness of Human Tutors

It is often thought that human tutors are the most effective form of instruction on the planet. For instance, Bloom (1984) claimed that human tutors produced a two-standard deviation effect size, which is extremely large. However, a review of experiments comparing human tutors with computer tutors suggests that human tutors are only slightly more effective than step-based tutoring systems (VanLehn 2011); that is, tutoring systems which can give feedback and hints after each step in the student's problem solving. Both human tutors and step-based tutors were around 0.75 standard deviations more effective than no tutoring.

Discussion

So much research has been done on human learning, tutoring, and instruction that it is easy to lose track of the bigger picture. Here, I will summarize my "big picture." I include the four most important factors that affect, in my opinion, human learning from instruction, and then use them to explain why human and computer tutoring are so effective. Finally, some implications are considered for instruction in the context of interactive task learning.

The "Big Four" Factors Affecting Human Learning

The first of the "big four," *engagement*, is arguably the most important factor when explaining the success or failure of students in classroom instruction. In laboratory studies, students are almost always engaged in doing the tasks assigned to them, but in classrooms, students often disengage and "go off task."

The second most important factor concerns the *qualitative nature of the learning curve*. A learning curve is a quantitative display of students' progress in learning a specific piece of knowledge. The horizontal axis corresponds to the episodes where the student tried to apply that piece of knowledge. The vertical axis is either the error probability or the duration of the episode. The curve descends in a ~~logarithmic~~ or power-law shape. That is, as people practice the same thing over and over, their errors decrease and their speed increases. However, learning curves only look nice and smooth when they represent averages over hundreds of students. The errors and durations of a single student's episodes are hardly smooth. This is because students' behavior and thinking varies qualitatively as they practice. Many theorists like to use the following stages as an approximation of the qualitative changes that underpin the learning curve (Anderson 1982; Fitts and Posner 1967; Koedinger et al. 2012; VanLehn 1996):

- *Sense making*: During the initial phase, students are just trying to understand information presented to them about the target knowledge. During the first attempt to apply the target knowledge, they might have to refer back to the text or example that comprised their original instruction. They may also need help in finding out what they need to know or in interpreting it. Such an episode can take minutes. If the second attempt immediately follows the first and the task is almost the same, then the episode might go quickly and smoothly, but the result may be incorrect. There is high variability in errors and duration during the sense-making phase.
- *Refinement*: During the middle phase, students understand the core or basic elements of the target knowledge, so their learning covers the fine points, or what programmers call "corner cases." Thus, students make errors and struggle on unusual tasks, but perform smoothly and rapidly on tasks that are similar to those previously experienced. Variability in errors and duration is less in the refinement phase than the sense-making phase.
- *Fluency building*: During the final phase, students fully understand the target knowledge, so their errors are only unintentional ones. As they practice, the probability of such slips decreases and their performance speed increases.

Different theorists have used different terms to describe these stages; the terms above are from Koedinger et al. (2012). The important point is that the cognitive and overt behavior of learners changes qualitatively as they practice. Instruction that is optimal at one time may not be optimal at a different time. This explains why model-scaffold-fade is so popular and effective. It is an effective way to make instruction adaptive.

The third important factor in human learning is the *type of behavior students engage in when learning a* ~~*task*, well~~ captured by the ICAP framework (Chi 2009; Chi and Wylie 2014; Fonseca and Chi 2011; Menekse et al. 2013). The acronym stands for four types of student behavior, listed below from least to most effective:

- *Passive:* The student is paying attention to the instruction, but does not move or do anything overtly.
- *Active:* In addition to paying attention to the instruction, the student is displaying overt (visible) behavior that suggests she is evaluating and

selecting portions of the information presented. Highlighting phrases in text, taking verbatim notes during a lecture, or answering multiple-choice questions are often active student behaviors.

- *Constructive:* To qualify as behaving constructively, students must generate information that is not contained in the instruction. That is, they are constructing/generating *new* information (e.g., inferences, examples, judgments).
- *Interactive:* This type of behavior occurs only when two or more students are working together; it is not an individual behavior. Interactive students are co-constructive. That is, both students are constructive but, critically, each student's construction builds upon the information that their partner generated.

The ICAP framework aptly predicts learning gains during the sense-making phase of learning and perhaps also during the refinement phase. However, it probably does not make the right predictions for fluency building. ICAP is also ordered by learning gains, not by duration. Passive activities generally go faster than the others. Interactive activities are probably the slowest, given that all four types of activity "cover" the same knowledge. Thus, for the very first exposure of students to a piece of knowledge, a passive activity might be a better choice than an interactive one just because it takes less time. Thus, ICAP may often be compatible with model-scaffold-fade. Although ICAP is a classification of student behavior, it is often used to classify instructional tasks. Thus, for example, a task is classified as interactive if the instructions for doing the task ask that students behave interactively. This does not mean that all students will behave interactively, which brings us to the last major factor.

The fourth major factor influencing human learning is *feedback* that compares the student's actual behavior to the desired ICAP behavior, pointing out discrepancies and suggesting ways to improve the student's behavior (Hattie and Timperley 2007; Narciss 2007; Shute 2008; VanLehn 2016). The feedback may be immediate or delayed, and it might be given directly to the student or conveyed via a teacher. The feedback might come from the task itself, as students discover what doesn't work or uncover contradictions in their beliefs.

Another type of feedback, although it is not usually recognized as such, is deciding when a student has passed a module, a course, or a grade level based on assessing the student's competence. When this feedback loop is applied to modules or units of a course it is called "move on when ready," "gated instruction," or "mastery learning" (Bloom 1984). Such feedback is another way in which instruction can be adaptive.

What Makes Tutoring Effective

Tutoring is effective because it addresses each of the "big four" factors listed above:

- Tutors (especially human tutors) *keep students engaged*.
- Tutors can determine from the student's behavior *where they are located on the learning curve* for each piece of target knowledge.
- Tutors can modify episodes within a task along the ICAP dimension so that the episode becomes *appropriate for the students' position along the learning curve*. For instance, if the tutor determines that a piece of target knowledge is completely unfamiliar to the student, the tutor may explain it to the student rather than watch silently as the student tries to construct it for themselves.
- Tutors can *provide feedback* that helps students approach optimal behavior for the task. During the refinement and fluency phases, tutors can encourage students to take over the feedback loop and self-regulate their behavior. Tutors that are allowed to control when a student passes a module can use *mastery learning* so that students move on to the next modules only when they have achieved mastery of the current module.

Implications for Teaching Bots

Discussions at this Forum usually assumed that someone instructing a bot would first explain how to do a task, then monitor the bot's initial performance of the task, and help the bot debug itself if there were errors. How well this works depends on both the human instructor and the bot. The discussion above suggests what our expectations of humans should be for each phase:

- *Initial explanation*: Human tutors love to give explanations, but human students rarely master a concept or skill from just an explanation, so it is appropriate to assume that bots will probably need all three stages (i.e., initial explanation, monitoring, and debugging) to master new knowledge. It is worth pointing out that some people think that if they explain something very clearly to students, then the students will understand it perfectly and that will suffice. Even professors who have been lecturing for years harbor this belief. Such people may become frustrated if the bot makes mistakes or asks for clarifications even after receiving a "perfectly clear" initial explanation.
- *Monitoring performances*: Human tutors are quite good at comparing a student's performance to what they would do in the same situation and spotting differences. They almost always give negative feedback but sometimes in an implicit way, which makes it hard for a learner to recognize that an error has been made.
- *Debugging*: Human tutors and learners are not good at debugging. Tutors are not good at asking diagnostic questions, nor are they able to use diagnostic information when it is given to them. Students, especially poor learners, often do not know when they fail to understand. Students seldom ask nontrivial questions that might help them debug

> their knowledge. Instead, when errors occur, the tutor simply teaches the correct knowledge again, and this seems to suffice for most students (Sleeman et al. 1989).

The strength of human tutoring lies in monitoring learners. Bots can trust their human tutors to point out errors, albeit in an oblique way sometimes. Since human tutors are not good at debugging, the bot may need to invent its own test cases or deep questions, and use them to infer on its own how to fix the errors pointed out by the tutor.

Humans can do this as well, and this behavior is referred to as self-regulated learning (Zimmerman 2008). Self-regulated learners constantly test their performance against desired performance, note discrepancies, and correct their errors. In the expert-novice literature, this is often called deliberate practice (Ericsson and Lehmann 1996). This is the kind of student that every tutor would be proud to teach because the tutor merely has to point out the occasional failure to exhibit perfect performance, and the student does the rest. If future ITL systems are designed for deliberate practice, tutoring bots may turn out to be infinitely more satisfying than tutoring humans.

Acknowledgments

Preparation of this chapter was supported by the ES Forum, NSF grants IIS-1628782 and DUE-1525197, grant OPP1061281 from the Bill and Melinda Gates Foundation, and the Diane and Gary Tooker Chair for Effective Education in Science, Technology, Engineering and Math.

Note: The book will contain a compiled bibliography. Listed below are just those entries cited in your chapter. Chapters in which an entry appears are coded in square brackets. When an entry is qualified as "a" or "b," this distinguishes citations for the entire volume -- not just a single chapter.

Anderson, J. R. 1982. The Acquisition of Cognitive Skill. *Psychol. Rev.* **89**:369–406. [12]

Bausell, R. B., W. B. Moody, and F. H. Walzl. 1972. A Factorial Study of Tutoring versus Classroom Instruction. *Am. Educ. Res. J.* **9**:591–597. [12]

Bloom, B. S. 1984. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educ. Res.* **13**:4–16. [12]

Boyer, K. E., R. Phillips, M. Wallis, M. Vouk, and J. C. Lester. 2008. Balancing Cognitive and Motivational Scaffolding in Tutorial Dialogue. In: Intelligent Tutoring Systems. Its 2008. Lecture Notes in Computer Science, ed. B. P. Woolf et al., vol. 5091, pp. 239–249. Heidelberg: Springer. [12]

Cade, W. L., J. L. Copeland, N. Person, and S. K. D'Mello. 2008. Dialogue Modes in Expert Tutoring. In: Intelligent Tutoring Systems. Its 2008. Lecture Notes in Computer Science, ed. B. P. Woolf et al., vol. 5091, pp. 470–479. Heidelberg: Springer. [12]

Chase, C. C., Marks, J., Bernett, D., Bradley, M., & Aleven, V. (2015). Towards the development of the invention coach: A naturalistic study of teacher guidance for an exploratory learning task. In Artificial Intelligence in Education 2015. ed. C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo, pp. 558-561). Switzerland: Springer International. [12]

Chi, M. T. H. 2009. Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Top. Cogn. Sci.* **1**:73–105. [12]

Chi, M. T. H., M. Bassok, M. Lewis, P. Reimann, and R. Glaser. 1989. Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cogn. Sci.* **15**:145–182. [12]

Chi, M. T. H., N. de Leeuw, M. H. Chiu, and C. LaVancher. 1994. Eliciting Self-Explanations Improves Understanding. *Cogn. Sci.* **18**:439–477. [12]

Chi, M. T. H., M. Roy, and R. G. M. Hausmann. 2008. Observing Tutorial Dialogues Collaboratively: Insights About Human Tutoring Effectiveness from Vicarious Learning. *Cogn. Sci.* **32**:301–342. [12]

Chi, M. T. H., S. Siler, and H. Jeong. 2004. Can Tutors Monitor Students' Understanding Accurately? *Cogn. Instr.* **22**:363–387. [12]

Chi, M. T. H., S. Siler, H. Jeong, T. Yamauchi, and R. G. Hausmann. 2001. Learning from Human Tutoring. *Cogn. Sci.* **25**:471–533. [12]

Chi, M. T. H., and R. Wylie. 2014. Icap: A Hypothesis of Differentiated Learning Effectiveness for Four Modes of Engagement Activities. *Educ. Psychol.* **49**:219–243. [12]

Cho, B.-I., J. A. Michael, A. A. Rovick, and M. W. Evens. 2000. An Analysis of Multiple Tutoring Protocols. In: Intelligent Tutoring Systems. Its 2000. Lecture Notes in Computer Science, ed. G. Gauthier et al., vol. 1839, pp. 212–221. Heidelberg: Springer. [12]

Collins, A. 1977. Processes in Acquiring Knowledge. In: Schooling and the Acquisition of Knowledge, ed. R. C. Anderson et al., pp. 339–363. Hillsdale, NJ: Lawrence Erlbaum. [12]

Collins, A., J. S. Brown, and S. E. Newman. 1989. Cognitive Apprenticeship: Teaching the Craft of Reading, Writing and Mathematics. In: Knowing, Learning and Instruction: Essays in Honor of Robert Glaser, ed. L. B. Resnick, pp. 453–494. Hillsdale, NJ: Lawrence Erlbaum. [12]

Collins, A., and A. Stevens. 1982. Goals and Strategies for Inquiry Teachers. In: Advances in Instructional Psychology, ed. R. Glaser, vol. 2, pp. 65–119. Hillsdale, NJ: Lawrence Erlbaum. [12]

Cordova, D. I., and M. R. Lepper. 1996. Intrinsic Motivation and the Process of Learning: Beneficial Effects of Contextualization, Personalization, and Choice. *J. Educ. Psychol.* **88**:715–730. [12]

Core, M. G., J. D. Moore, and C. Zinn. 2003. The Role of Initiative in Tutorial Dialogue. In: Proc. 10th Conf. of the European Chapter of the Association for Computational Linguistics, ed. A. Copestake and J. Hajic, vol. 1, pp. 67–74. Stroudsburg, PA: Assn. of Computational Linguistics. [12]

Dweck, C. S. 1986. Motivational Processes Affecting Learning. *Am. Psychol.* **41**:1040–1048. [12]

Ericsson, K. A., and A. C. Lehmann. 1996. Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints. *Annu. Rev. Psychol.* **47**:273–305. [12]

Evens, M., and J. Michael. 2006. One-on-One Tutoring by Humans and Machines. Mahwah, NJ: Lawrence Erlbaum. [12]

Fitts, P. M., and M. I. Posner. 1967. Human Performance. Belmont, CA: Brooks/Cole. [12]

Fonseca, B., and M. T. H. Chi. 2011. The Self-Explanation Effect: A Constructive Learning Activity. In: The Handbook of Research on Learning and Instruction, ed. R. E. Mayer and P. Alexander, pp. 296–321. New York: Routledge. [12]

Fox, B. A. 1991. Cognitive and Interactional Aspects of Correction in Tutoring. In: Teaching Knowledge and Intelligent Tutoring, ed. P. Goodyear, pp. 149–172. Norwood, NJ: Ablex. [07, 12]

———. 1993. The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems. Hillsdale, NJ: Lawrence Erlbaum. [07, 12]

Frederiksen, N., J. Donin, and M. Roy. 2000. Human Tutoring as a Model for Computer Tutors: Studying Human Tutoring from a Cognitive Perspective. In: Modelling Human Teaching Tactics and Strategies: Workshop W1 of ITS'2000, ed. B. du Boulay. Heidelberg: Springer. [12]

Glass, M. S., J. H. Kim, M. Evens, J. Michael, and A. Rovick. 1999. Novice vs. Expert Tutors: A Comparison of Style. In: Proc. 10th Midwest Artificial Intelligence and Cognitive Science Conf. (MAIS-99). Bloomington: AAAI Press. [12]

Glenburg, A., A. Wilkinson, and W. Epstein. 1982. The Illusion of Knowing: Failure in the Self-Assessment of Comprehension. *Mem. Cognit.* **20**:597–602. [12]

Goldman, S., J. Pellegrino, and J. D. Bransford. 1993. Assessing Programs That Invit Thinking. In: Technology Assessment: Estimating the Future, ed. H. F. O'Neil and E. L. Baker, pp. 199–230. Hillsdale, NJ: Lawrence Erlbaum. [12]

Graesser, A. C., N. Person, and J. Magliano. 1995. Collaborative Dialog Patterns in Naturalistic One-on-One Tutoring. *Appl. Cogn. Psychol.* **9**:359–387. [12]

Hattie, J., and H. Timperley. 2007. The Power of Feedback. *Rev. Educ. Res.* **77**:81–112. [12]

Henderlong, J., and M. R. Lepper. 2002. The Effects of Praise on Children's Intrinsic Motivation: A Review and Synthesis. *Psychol. Bull.* **128**:774–795. [12]

Hume, G., J. Michael, A. Rovick, and M. Evens. 1996. Hinting as a Tactic in One-on-One Tutoring. *J. Learn. Sciences* **5**:23–49. [12]

Jeong, H., S. Siler, and M. T. H. Chi. 1997. Can Tutors Diagnose Students' Understanding? In: Proc. 19th Annu. Conf. of the Cognitive Science Society, ed. M. G. Shafto and P. Langley, p. 959. Mahwah, NJ: Lawrence Erlbaum. [12]

Juel, C. 1996. What Makes Literacy Tutoring Effective? *Read. Res. Q.* **31**:268–289. [12]

Katz, S., D. Allbritton, and J. Connelly. 2003. Going Beyond the Problem Given: How Human Tutors Use Post-Solution Discussions to Support Transfer. *Int. J. Artific. Intel. Educ.* **13**:79–116. [12]

Kim, J. H., H. M. Chae, and M. S. Glass. 2005. Expert and Novice Algebra Tutor Behaviors Compared. Proc. 27th Annu. Conf. of the Cognitive Science Society, ed. B. G. Bara et al., p. 2499. Mahwah, NJ: Lawrence Erlbaum. [12]

Kluger, A. N., and A. DeNisi. 1996. The Effects of Feedback Intervention on Performance: A Historical Review, a Meta-Analysis and a Preliminary Feedback Intervention Theory. *Psychol. Bull.* **112**:254–284. [12]

Koedinger, K. R., A. Corbett, and C. Perfetti. 2012. The Knowledge-Learning-Instruction (KLI) Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cogn. Sci.* **36**:757–798. [12]

Lazowski, R. A., and C. S. Hulleman. 2016. Motivation Interventions in Education: A Meta-Analytic Review. *Rev. Educ. Res.* **86**:602–640. [12]

Lepper, M. R., and M. Woolverton. 2002. The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In: Improving Academic Achievement: Impact of Psychological Factors on Education, ed. J. Aronson, pp. 135–158. New York: Academic Press. [12]

Lepper, M. R., M. Woolverton, D. L. Mumme, and J.-L. Gurtner. 1993. Motivational Techniques of Expert Human Tutors: Lessons for the Design of Computer-Based Tutors. In: Computers as Cognitive Tools, ed. S. P. Lajoie and S. J. Derry, pp. 75–105. Hillsdale, NJ: Lawrence Erlbaum. [12]

Marston, D., S. L. Deno, D. Kim, K. Diment, and D. Rogers. 1995. Comparison of Reading Intervention Approaches for Students with Mild Disabilities. *Except. Child.* **62**:20–37. [12]

McArthur, D., C. Stasz, and M. Zmuidzinas. 1990. Tutoring Techniques in Algebra. *Cogn. Instr.* **7**:197–244. [12]

Menekse, M., G. S. Stump, S. Krause, and M. T. H. Chi. 2013. Differentiated Overt Learning Activities for Effective Instruction in Engineering Classrooms. *J. Engineer. Educ.* **102**:346–374. [12]

Merrill, D. C., B. J. Reiser, S. K. Merrill, and S. Landes. 1995. Tutoring: Guided Learning by Doing. *Cogn. Instr.* **13**:315–372. [12]

Merrill, D. C., B. J. Reiser, M. Ranney, and J. G. Trafton. 1992. Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *J. Learn. Sciences* **2**:277–306. [12]
Narciss, S. 2007. Feedback Strategies for Interactive Learning Tasks. In: Handbook of Research on Educational Communications and Technology (3rd Edition), , ed. J. M. Spector et al., pp. 125–144. Mahwah, NJ: Erlbaum. [12]
Ohlsson, S., B. Di Eugenio, B. Chow, et al. 2007. Beyond the Code-and-Count Analysis of Tutoring Dialogues. In: Artificial Intelligence in Education, ed. R. Luckin et al., pp. 349–356. Amsterdam: IOS Press. [12]
Palinscar, A. S., and A. L. Brown. 1984. Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cogn. Instr.* **1**:117–175. [12]
Putnam, R. T. 1987. Structuring and Adjusting Content for Students: A Study of Live and Simulated Tutoring of Addition. *Am. Educ. Res. J.* **24**:13–48. [12]
Rose, C. P., D. Bhembe, S. Siler, R. Srivastava, and K. VanLehn. 2003. The Role of Why Questions in Effective Human Tutoring. In: Artificial Intelligence in Education, ed. U. Hoppe et al. Amsterdam: IOS Press. [12]
Scruggs, T. E., and L. Richter. 1985. Tutoring Learning Disabled Students: A Critical Review. *Learn. Disab. Q.* **8**:286–298. [12]
Shah, F., M. Evens, J. Michael, and A. Rovick. 2002. Classifying Student Initiatives and Tutor Responses in Human Keyboard-to-Keyboard Tutoring Sessions. *Discourse Process.* **33**:23–52. [12]
Shanahan, T. 1998. On the Effectiveness and Limitations of Tutoring in Reading. *Rev. Res. Educ.* **23**:217–234. [12]
Shute, V. J. 2008. Focus on Formative Feedback. *Rev. Educ. Res.* **78**:153–189. [12]
Siler, S., and K. VanLehn. 2015. Investigating Micro-Adaptation in One-to-One Tutoring. *J. Experiment. Edu.* **83**:344–367. [12]
Sleeman, D., A. E. Kelly, R. Martinak, R. D. Ward, and J. L. Moore. 1989. Studies of Diagnosis and Remediation with High School Algebra Students. *Cogn. Sci.* **13**:551–568. [12]
Stevens, A., and A. Collins. 1977. The Goal Structure of a Socratic Tutor. In: Proc. Natl. ACM Conf., pp. 256–263. New York: ACM. [12]
Stevens, A., A. Collins, and S. E. Goldin. 1979. Misconceptions in Student's Understanding. *Int. J. Man Mach. Stud.* **11**:145–156. [12]
VanLehn, K. 1988. Student Modeling. In: Foundations of Intelligent Tutoring Systems, ed. M. Polson and J. Richardson, pp. 55–78. Hillsdale, NJ: Lawrence Erlbaum. [12, 13]
———. 1996. Cognitive Skill Acquisition. *Annu. Rev. Psychol.* **47**:513–539. [12]
———. 1999. Rule Learning Events in the Acquisition of a Complex Skill: An Evaluation of Cascade. *J. Learn. Sciences* **8**:179–221. [12]
———. 2008. Intelligent Tutoring Systems for Continuous, Embedded Assessment. In: The Future of Assessment: Shaping Teaching and Learning, ed. C. A. Dwyer, pp. 113–138. New York: Lawrence Erlbaum. [12]
———. 2011. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems and Other Tutoring Systems. *Educ. Psychol.* **46**:197–221. [12]
———. 2016. Regulative Loops, Step Loops and Task Loops. *Int. J. Artific. Intel. Educ.* **26**:107–112. [12]
VanLehn, K., S. Siler, C. Murray, T. Yamauchi, and W. B. Baggett. 2003. Human Tutoring: Why Do Only Some Events Cause Learning? *Cogn. Instr.* **21**:209–249. [12]
Wasik, B. A. 1998. Volunteer Tutoring Programs in Reading: A Review. *Read. Res. Q.* **33**:266–291. [12]
Wasik, B. A., and R. E. Slavin. 1993. Preventing Early Reading Failure with One-to-One Tutoring: A Review of Five Programs. *Read. Res. Q.* **28**:178–200. [12]
Wittwer, J., M. Nuckles, N. Landmann, and A. Renkl. 2010. Can Tutors Be Supported in Giving Effective Explanations? *J. Educ. Psychol.* **102**:74–89. [12]
Wood, D. J., J. S. Bruner, and G. Ross. 1976. The Role of Tutoring in Problem Solving. *J. Child Psychol. Psychiatry* **17**:89–100. [12]
Zimmerman, B. 2008. Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments and Future Prospects. *Am. Ed. Res. J.* **45**:166–183. [12]