Running head: NEWTONIAN PHYSICS

Learning Newtonian Physics with Conversational Agents and Interactive Simulation

Arthur C. Graesser, Don Franceschetti, Barry Gholson, and Scotty Craig

University of Memphis

Send correspondence to:

Art Graesser
Department of Psychology & Institute for Intelligent Systems
202 Psychology Building
University of Memphis
Memphis, TN 38152-3230
901-678-2146
901-678-2579 (fax)
a-graesser@memphis.edu

Chapter targeted for book edited by N. Stein (Ed), *Developmental and Learning Sciences Go to School: Implications for Education and Public Policy*.

Learning Newtonian Physics with Conversational Agents and Interactive Simulation

Newtonian physics is an area of science that is part of the curriculum in middle and high schools throughout the country. It is easy to set up demonstrations of physics principles that are easy to observe and provocative to talk about. Students drop objects of different size or weight and wonder whether they hit the pavement at the same time. They observe how objects float in water, how objects collide, and how moving objects land at particular locations. Some of what the students see is counterintuitive, which ideally stimulates them to ask questions and perform mini-experiments to answer the questions. Students discuss what they see and learn with their peers and teachers.

Consider, for example, the following conceptual physics question.

PHYSICS QUESTION: If a lightweight car and a massive truck have a head-on collision, upon which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion, and why?

When students answer this question, they would hopefully communicate principles such as a 1 and 2 below. However, they might be misled by the two misconceptions.

Principle 1: The magnitudes of the forces exerted by A and B on each other are equal.

Principle 2. If A exerts a force on B, then B exerts a force on A in the opposite direction.

Misconception 1: A lighter/smaller object exerts no force on a heavier/larger object.

Misconception 2: Heavier objects accelerate faster for the same force than lighter objects.

The hope is that they will acquire more correct principles and fewer misconceptions over the course of learning. Research on physics learning has shown, however, that it is very difficult to correct many of the misconceptions, particularly those that are entrenched in a student's everyday experiences (Chi, 2005; diSessa, 1993; Hunt & Minstrell, 1996; Ploetzner & VanLehn, 1997). For example, misconception 1 would appear to be confirmed perceptually when a child throws a rubber ball against a wall or when a ping pong ball is hit by a paddle.

The mechanisms of Newtonian physics are arguably more visible than the mechanisms of other forms of physics (such as thermodynamics), chemistry, and biology. It is true that there are abstract constructs like force in Newtonian physics, but there are many other constructs that can be readily observed, such as motion, volume, and distance. The objects are typically large enough to be manipulated and perceived. In contrast, there are many small and minimally visible entities in chemistry, such as atoms and molecules; a theoretical model is needed to magnify and characterize (or indeed caricature) the components to make them visible. In biology, there are many parts of plants and animals that are visible, but the scientific mechanisms of biology are much more complex and difficult to visualize and manipulate, as in the case of genetics and DNA. There is considerable jargon to memorize in biology when students learn taxonomies of living organisms and the parts of intricate anatomies.

The tight correspondence of Newtonian physics principles to everyday experience has both advantages and liabilities. The advantages lie in (a) the natural correspondence between the constructs of physics and everyday actions, events, and perceptions, (b) the ease of setting up demonstrations to illustrate many physics principles, and (c) the comparative low density of jargon to memorize. The disadvantages lie in (a) the clash between some mental models that are grounded in everyday experience and the proper mental models of Newtonian physics and (b) the difficulty of correcting mental models that are entrenched in everyday experiences.

It is illuminating to explore the mental models of students by asking them to generate explanations while solving the conceptual physics problems (Gentner & Stevens, 1983; VanLehn et al., 2007), such the example above or the example below.

> When a car without headrests on the seats is struck from behind, the passengers often suffer neck injuries. Why do passengers get neck injuries in this situation? Explain why.

The nature of the students' mental models is often manifested by asking them to draw sketches of the processes. Many college students believe that a rear-end collision will directly push the head of the victim forward, sometimes unfortunately through the windshield. Their mental model is that the head goes forward much like a billiard ball goes forward when hit from behind. They may have a memory of a person's head going through the windshield from a movie or personal experience. However, this is flawed reasoning. The head first goes backwards after the impact of the collision by virtue of the forces underlying Newton's laws, which explains whiplashes in accidents. The head subsequently goes forward after recoil. Knowledgeable students identify the initial stage of the head going back when asked to draw pictures, but students with shallow understanding miss this step. It is not until the student can reason with abstract vectors, forces, and Newton's laws that a correct answer emerges. The generation of verbal explanations and pictures are excellent tasks for diagnosing misconceptions.

Methods of Training Newtonian Physics

It is beyond the scope of this chapter to summarize all of the traditional and inventive ways that Newtonian physics has been taught. Instead, this section briefly addresses some of the typical learning environments, whereas subsequent sections focus on our own work involving human and computer tutoring. Most of our research on the learning of Newtonian physics has been on college students, but we have collected some data from 9th and 11th graders. Our expertise is not in early child development and elementary school education, however, so our remarks on physics research would only be speculative at the younger age and grade levels. It is a very important research question whether the findings from middle school and above could be extrapolated to a child in the second grade or even younger (Stein, Hernandez, & Gamez, 2007). The learning of science at young ages is a foundational question for this edited volume so we will offer some remarks in the final section of this chapter.

(1) *Reading textbooks*. Textbooks are routinely assigned to students in physics courses so one would hope that this conventional approach would promote learning. However, students normally have low background knowledge of science so it is a struggle for students to comprehend the text and stay motivated as they struggle (Otero, Leon, & Graesser, 2002; Vitale & Romance, 2007). They no doubt acquire shallow knowledge from reading, such as definitions of terms, facts, lists of properties of entities, and historical details. However, it is more difficult for students to acquire deep knowledge, such as causal explanations, dynamic processes, and complex mechanisms. Indeed, the ability of students to acquire deep knowledge from reading is somewhat doubtful. We have assessed how much deep knowledge is acquired from an excellent textbook (*Conceptual Physics*, Hewitt, 1998) when college students read the textbook for several hours and are subsequently tested on deep knowledge with a multiple choice test similar to the highly regarded Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992). No significant differences were found in posttest scores when we compared a Textbook Reading condition to a Do Nothing condition (Graesser, Jackson et al. 2003; VanLehn et al., 2007). Scores were significantly higher in various tutoring conditions that will be described later in this chapter. The counterintuitive claim here is that students get shallow knowledge but not deep knowledge from reading physics texts, so it is not sufficient to merely assign a textbook and expect the students to learn physics very deeply. This claim would no doubt apply to young children, particularly those who are struggling readers.

(2) *Teachers lecturing*. Students presumably learn well from the lectures of award-winning teachers. However, it is hardly a secret that most of the physics science teachers in schools are not

adequately trained in science. Teachers also vary in pedagogical methods. One of the drawbacks to such lecturing environments is that the training is not tailored to individual students.

(3) *Discovery learning environments*. Students would no doubt be fascinated with a museum of physics artifacts that allow them to discover physics principles on their own. However, available evidence is that little is learned unless there is an instructor or tutor who assigns activities and mediates the learning process (Klahr, 2002). Researchers have carefully created inquiry learning environments that stimulate students to ask questions, to generate hypotheses, and to plan activities for experimenting in pursuit of answers (Goldman et al., 2003; White & Frederiksen, 1998). However, it is extremely difficult to design such environments to guarantee effective inquiry and discovery. Students need guidance.

(4) *Human tutoring*. Human tutoring is an effective method of teaching science topics so the process of human tutoring has been investigated in depth (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Graesser, Person, & Magliano, 1995; VanLehn et al., 2007). Meta-analyses show learning gains from human tutors that vary in expertise of 0.42 sigma (effect size in standard deviation units) compared to classroom controls and other suitable controls (Cohen, Kulik & Kulik, 1982). More will be said about tutoring later in the chapter.

(5) *Animation*. Computers can be designed to present animations that exhibit physics principles through event sequences that unfold over time. However, the animations run a number of risks: not being easy to understand, being transient, moving too quickly, presenting distracting material, placing demands on working memory, and depicting processes in a fashion other than what the learner would otherwise actively construct (Hegarty, 2004). Therefore, animations have failed to improve learning in a large percentage of systematic studies (Ainsworth, 2008; Tversky, Morrison, & Betrancourt, 2002).

(6) *Multimedia*. Physics lessons can be delivered in different presentation modes (verbal, pictorial), sensory modalities (auditory, visual), and delivery media (text, video, simulations). The impact of different forms of multimedia has been extensively investigated by Mayer and his colleagues (Mayer, 2005). Meta-analyses by Dodds and Fletcher (2004) report an effect size of .50 sigma for multimedia learning, whereas the meta-analyses reported by Mayer (2005) is considerably higher, more like 1.00. Mayer (2005) has documented and empirically confirmed a number of psychological principles that predict when different forms of multimedia will facilitate learning. Among these are the principles of multimedia, modality, spatial and temporal contiguity, coherence, and redundancy. In many of these studies, retention, problem solving, and transfer of training is facilitated by multimedia because the separate modalities offer multiple codes, conceptually richer and deeper representations, and multiple retrieval routes. However, it is important that the multimedia presentation does not present a large cognitive load and split the learner's attention (Kalyuga, Chandler, & Sweller, 1999).

(7) *Interactive simulation.* Interactive simulation is expected to enhance learning because the learner can actively control input parameters and observe the results on the system. The learner can slow down animations to inspect the process in detail, zoom in on important subcomponents of a system during the course of a simulation, observe the system from multiple viewpoints, and systematically relate inputs to outputs (Kozma, 2000). Interactive simulation has indeed shown a positive impact on science learning in several studies but others have shown no gains compared with various control conditions (Jackson, Olney, Graesser, & Kim, 2006; Van der Meij & de Jong, 2006). These learning environments may have complex content and human-computer interfaces that are unfamiliar to learners, so there are difficulties getting started, managing the interface, and strategically interacting with the simulation to advance learning.

Our research group has investigated interactive simulation in microworlds that captured principles of Newtonian physics (Jackson et al., 2006). We developed an interactive simulation world with people, objects, and the spatial setting associated with the conceptual problems that were illustrated earlier. The college students could manipulate parameters of the situation (e.g., mass of objects, speed of objects, distance between objects) and then ask the system to simulate what will happen. The results showed only a small nonsignificant increase in learning compared with a conversation-based computer tutoring system called AutoTutor, which will be discussed later. The good news is that there were substantial gains for those students who ran several simulations for a problem. However, most students ran only one simulation, which is entirely inadequate for tracing the impact of variables on system behavior. For example, in the example collision problems, a student might want to compare the impact of the mass of the vehicles and air resistance on the displacement of vehicles in the collision. In order to discover that the mass of vehicles is important, but air resistance is not, it is necessary to run 4 simulations that cross large versus small mass with high versus zero air resistance (see Klahr, 2002). Students rarely performed such a systematic comparison. Examples like these illustrate that students need to be trained on how to effectively use interactive simulations.

(8) *Intelligent tutoring systems*. ITS's track the knowledge states of learners in fine detail and adaptively respond with activities that are sensitive to these knowledge states. The processes of tracking knowledge (called user modeling) and adaptively responding to the learner incorporate computational models in artificial intelligence and cognitive science, such as production systems, case-based reasoning, Bayes networks, theorem proving, and constraint satisfaction algorithms. Successful systems have been developed for Newtonian physics, such as Andes and Why/Atlas (VanLehn et al., 2002; VanLehn et al., 2007). These systems show

impressive learning gains (.5 to 1.00 sigma), particularly for deeper levels of comprehension. Children might benefit from the verbal ITSs (such as Why-Atlas) whereas the mathematical ITSs (Andes) would have mathematical constructs that are beyond the children's zone of proximal development.

(9) *Games*. The game industry has captured the imagination of the current and next generations of students. Serious teenage gamers play games over 20 hours per week. There are many types of games, even first person shooter games, that would have the potential of incorporating physics principles. The challenge of combining entertainment and pedagogical content is the foundational question of serious games (Gee, 2003; O'Neil, Wainess, Baker, 2005; Ritterfeld, Cody, & Vorderer, in press; Vorderer, Bryant, Pieper, & Weber, 2006). Presumably, the success of a game can be attributed to such factors as feedback, progress markers, engaging content, fantasy, competition, challenge, uncertainty, curiosity, control, and other factors that involve cognition, emotions, motivation, and art. The technical and psychological components of games have been analyzed at considerable depth, but there has been very little research on the impact of these components on learning gains, engagement, and usability (Malone & Lepper, 1987; O'Neil et al., 2005; Virvou, Katsionis, & Manos, 2005).

(10) *Animated conversational agents*. Animated conversational agents play a central role in some of the recent advanced learning environments (Atkinson, 2002; Baylor & Kim, 2005; Graesser, Chipman, Haynes, & Olney, 2005; McNamara, Levinstein, & Boonthum, 2004; Moreno & Mayer, 2004; Reeves & Nass, 1996). These agents interact with students and help them learn by either modelling good pedagogy or by holding a conversation. The agents may take on different roles: mentors, tutors, peers, players in multiparty games, or avatars in the virtual worlds. The students communicate with the agents through speech, keyboard, gesture, touch panel

screen, or conventional input channels. In turn, the agents express themselves with speech, facial expression, gesture, posture, and other embodied actions. Intelligent agents with speech recognition essentially hold a face-to-face, mixed-initiative dialogue with the student, just as humans do (Graesser, Jackson, & McDaniel, 2007; Johnson & Beal, 2005). Single agents model individuals with different knowledge, personalities, physical features, and styles. Ensembles of agents model social interaction.

From the standpoint of learning, there are at least three fundamental reasons why these agents would be effective in facilitating knowledge construction. First, it is well documented that one-to-one tutoring is one of the most effective methods of helping students learn. Second, the computer-generated agents can consistently and reliably apply the tutoring strategies, unlike teachers and human tutors. Third, agents can demonstrate (i.e., model) good learning activities and strategies. Students rarely have the opportunity to observe other students exhibiting good learning strategies in the classroom and other typical settings in school systems. Both single agents and ensembles of agents can be carefully choreographed to mimic virtually any activity or social situation: curiosity, inquiry learning, negotiation, interrogation, arguments, empathetic support, helping, and so on. Agents not only enact these strategies, individually or in groups, but can also think aloud while they do so.

Two Example Learning Environments with Agents: AutoTutor and iDRIVE

This section describes two computer learning environments with agents that we believe have some potential for helping young children learn Newtonian physics. The two systems (AutoTutor and iDRIVE) have already proven to be effective for college students, as well as middle and high school students. Although they have been untested in young children, we believe they hold some promise because an agent can simulates face-to-face tutorial conversations with the learner or two agents can model good conversations with each other. AutoTutor adapts to the idiosyncratic knowledge states of the student and has been integrated with 3-D interactive simulation environments. iDRIVE is scripted, rather than adaptive, but the script involves a dialogue between two agents that models excellent question asking and inquiry. This section describes these two systems in the context of Newtonian physics and reports learning gains.

AutoTutor

AutoTutor is an intelligent tutoring system that helps students learn through tutorial dialogue in language (Graesser et al., 2005; Graesser, Lu et al., 2004; VanLehn et al., 2007). AutoTutor's physics dialogues are organized around the example conceptual physics problems that require reasoning and explanations in the answers. These questions require the learner to construct approximately 3-7 sentences in an ideal answer and to exhibit reasoning in natural language. It takes a conversation to answer each one of these questions, typically 30 to 100 conversational turns between AutoTutor and the student.

Table 1 illustrates AutoTutor in a conversation that was extracted from an actual tutoring session. This session was with a relatively verbose, knowledgeable, college student so the conversation is comparatively short. Lower ability students would take many more conversational turns before a good answer would be constructed. When students are asked these challenging questions, their initial answers are typically only 1 or 2 sentences in length. However, 1-2 sentences provide insufficient information to adequately answer the question so tutorial dialogue is needed to flesh out a complete answer. AutoTutor engages the student in a mixed-initiative dialogue that draws out more of what the student knows and that assists the student in the construction of an improved answer.

AutoTutor can be viewed as a proof of concept that a computer tutor can manage a reasonably smooth and pedagogically effective conversation. It is beyond the scope of this chapter to discuss the theoretical and computational mechanisms of AutoTutor, but some highlights are in order. The structure of the dialogue in both AutoTutor and human tutoring (Chi et al., 2001; Graesser et al., 1995; VanLehn et al., 2007) can be segregated into three structural components: (1) expectation and misconception-tailored dialogue, (2) a 5-step dialogue frame, and (3) composition of a conversational turn. These three levels can be automated and produce respectable tutorial dialogue.

Expectation and misconception tailored dialogue. Both AutoTutor and human tutors typically have a list of expectations (anticipated good answers) and a list of anticipated *misconceptions* associated with each main question. For example, expectations E1 and E2 (corresponding to principles P1 and P2) and misconceptions M1 and M2 are relevant to the example physics problem that was presented earlier. AutoTutor guides the student in articulating the expectations through a number of dialogue moves: *pumps* (what else?), *hints*, and *prompts* for the student to fill in missing words. Hints and prompts are carefully selected by AutoTutor to produce content in the answers that fill in missing content words, phrases, and propositions. For example, a hint to get the student to articulate expectation E1 might be “What about the forces exerted by the vehicles on each other?”; this hint would ideally elicit the answer “The magnitudes of the forces are equal.” A prompt to get the student to say “equal” would be “What are the magnitudes of the forces of the two vehicles on each other?” As the learner expresses information over many turns, the list of expectations is eventually covered and the main question is scored as answered. Complete coverage of the answer requires AutoTutor to have a pool of hints and prompts available to extract all of the content words, phrases, and propositions in each expectation. AutoTutor adaptively selects those hints and prompts that fill missing constituents and thereby achieves pattern completion.

AutoTutor is dynamically adaptive to the learner in other ways than coaching them to articulate expectations. There is the conversational goal of correcting misconceptions that arise in the student's talk. When the student articulates a misconception, AutoTutor acknowledges the error and corrects it. There is the conversational goal of giving feedback to the student on their contributions. AutoTutor gives short feedback on the quality of student contributions: positive (*very good, yeah*), neutral (*uh huh, okay*), or negative (*not quite, not really*). AutoTutor accommodates a mixed-initiative dialogue by attempting to answer the student's questions. The answers to the questions are retrieved from glossaries or from paragraphs in textbooks via intelligent information retrieval. AutoTutor asks counter-clarification questions (e.g., I don't understand your questions, so could you ask it in another way?) when it does not understand the students' questions.

Five-step dialogue frame. This dialogue frame is prevalent in human tutoring (Graesser & Person, 1994; VanLehn et al., in press) and is implemented in AutoTutor. The 5 steps of the dialogue frame are:

(1) Tutor asks main question.

(2) Student gives initial answer.

(3) Tutor gives short feedback on the quality of the student's answer in #2.

(4) Tutor and student collaboratively interact via expectation and misconception tailored dialogue.

(5) Tutor verifies that the student understands (e.g., Do you understand?).

Students often answer that they understand in step 5, when most do not. A good tutor would press the student further by asking more questions to verify the student's understanding, but even good tutors rarely do this. Most tutors end up giving a summary answer to the main question and then select another main question. A good tutor would ask the student to provide the summary (as in the example dialogue in Table 1) rather than it being provided by the tutor, but even good tutors rarely do that.

Managing one conversational turn. Each turn of AutoTutor in the conversational dialogue has three information slots (i.e., units, constituents). The first slot of most turns is short feedback (positive, neutral, or negative) on the quality of the student's last turn. The second slot advances the coverage of the ideal answer with either prompts for specific words, hints, assertions with correct information, corrections of misconceptions, or answers to student questions. The third slot is a cue to the student for the floor to shift from AutoTutor as the speaker to the student. For example, AutoTutor ends each turn with a question or a gesture to cue the learner to do the talking. Discourse markers (*and, also, okay, well*) connect the utterances of these three slots of information within a turn.

The three levels of AutoTutor go a long way in simulating a human tutor. AutoTutor can keep the dialogue on track because it is always comparing what the student says to anticipated input (i.e., the expectations and misconceptions in the curriculum script). Pattern matching operations and pattern completion mechanisms drive the comparison. These matching and completion operations are based on latent semantic analysis (Landauer et al., 2007) and symbolic interpretation algorithms (Rus, McCarthy, McNamara, & Graesser, in press) that are beyond the scope of this article to address. AutoTutor cannot interpret student contributions that have no matches to content in the curriculum script. This of course limits true mixed-initiative dialogue. Thus, AutoTutor cannot explore the topic changes and tangents of students as the students introduce them. However, available studies of naturalistic tutoring (Chi et al., 2001; Graesser et al., 1995) reveal that (a) human tutors rarely nurture true mixed-initiative dialogue when students change topics that steer the conversation off course and (b) most students rarely change topics, rarely ask questions, and rarely take the initiative to grab the conversational floor. Instead, it is the tutor that takes the lead and drives the dialogue. AutoTutor and human tutors are very similar in these respects.

Versions of AutoTutor. We have created many versions of AutoTutor that were designed to incorporate particular pedagogical goals and cover different topics. So far, the topics have covered computer literacy, physics, biology, tactical planning, and critical thinking. In most versions of AutoTutor, the students type in their contributions via keyboard, whereas recent versions allow spoken input. The interface of AutoTutor has a number of different windows that show: the conversational agent, the main question to work on, the information entered by the student, the dialogue history (optional), and an external picture or interactive simulation (optional).

Figure 1 shows the interface of a version of AutoTutor that has an interactive simulation. This *AutoTutor-3D* version guides learners on using interactive simulations of physics microworlds (Graesser, Chipman et al., 2005; Jackson et al., 2006). The student manipulates parameters of the situation (e.g., mass of objects, speed of objects, distance between objects) and then asks the system to simulate what will happen. Students are also prompted to describe what they see. Their actions and descriptions are evaluated with respect to covering the expectations or matching misconceptions. AutoTutor manages the dialogue with hints and suggestions that scaffold the learning process with dialogue.

**** INSERT FIGURE 1 ABOUT HERE ****

We are currently working on a version of AutoTutor that is sensitive to the students' emotions. AutoTutor is augmented with sensing devices and signal processing algorithms that classify affective states of learners. Emotions are classified on the basis of dialog patterns during tutoring, the content covered, facial expressions, body posture, and speech intonation (D'Mello, Picard, & Graesser, 2006). The primary emotions that occur during learning with AutoTutor are frustration, confusion, boredom, and flow (engagement), whereas surprise and delight occasionally occur (Graesser et al., 2008). It should be noted that trained human judges are not much more reliable that AutoTutor's algorithms when classifying these emotions. We are currently investigating whether learning gains and impressions of AutoTutor are influenced by dialogue moves of AutoTutor that are sensitive to the learner's emotions. For example, if the student is extremely frustrated, then AutoTutor presumably should give a good hint or prompt that directs the student in a more positive learning trajectory. If the student is bored, AutoTutor should give more engaging, challenging, and motivating problems. If the student is very absorbed and satisfied, then AutoTutor should be minimally directive. The emotions exhibited by AutoTutor is also an important consideration, just as it is for human tutoring (Lepper & Woolverton, 2002). Should AutoTutor be empathetic to a frustrated student, or be earnest, forceful, or upbeat? Answers to such questions await future research.

Learning gains with AutoTutor. The learning gains of AutoTutor have been evaluated in 15 experiments conducted during the last 9 years. Assessments of AutoTutor on learning gains have shown effect sizes of approximately 0.8 standard deviation units in Newtonian physics (VanLehn, Graesser et al., 2007), which is on par with human tutors. The two primary measures used in assessing learning are (1) multiple choice questions on deep knowledge (see Force Concept Inventory of Hestenes et al., 1992) and (2) the quality of answers to essay questions that involve near or far transfer from the training problems. A variety of comparison conditions to AutoTutor have uncovered the following findings.

(1) *AutoTutor versus reading a textbook*. Learning gains with AutoTutor are superior to reading from a textbook on the same topics for an equivalent amount of time.

(2) *Reading a textbook versus doing nothing*. Learning gains are zero in both of these conditions when the tests tap deeper levels of comprehension (as discussed earlier).

(3) *AutoTutor versus expert human tutors*. Comparisons were made between AutoTutor and accomplished human tutors via computer mediated communication. The learning gains were equivalent for students with a moderate degree of physics knowledge, but the expert human tutors prevailed when the students had low physics knowledge and the dialogue was spoken.

(4) *Zone of proximate development*. AutoTutor is most effective when there is an intermediate gap between the learner's prior knowledge and the ideal answers of AutoTutor. AutoTutor is not particularly effective in students with high domain knowledge or when students with low knowledge receive problems that do not push them to new levels of understanding.

(5) *Carefully prepared texts versus AutoTutor*. AutoTutor shows few if any advantages when compared to texts that succinctly answer the physics questions.

iDRIVE (Instruction with Deep-level Reasoning questions In Vicarious Environments)

AutoTutor interacts with students one-to-one in a fashion that attempts to be sensitive to the learners' cognitive profiles. Learning environments can also have pairs of agents (dyads) and larger ensembles of agents that exhibit ideal learning strategies and social interactions (McNamara et al., 2004; Millis et al., 2006). iDRIVE has agent dyads train students to learn science content by modeling deep reasoning questions in question-answer dialogues. A student agent asks a series of deep questions about the science content and the teacher agent immediately answers each question. There is evidence that learning improves when learners have the mindset of asking deep questions (*why, how, what-if, what-if-not*) that tap causal structures, complex systems, and logical justifications (Craig, Gholson, Ventura, & Graesser, 2000; Rosenshine, Meister, & Chapman, 1996). However, the asking of deep questions and inquiry does not come naturally (Graesser, McNamara, & VanLehn, 2005) so the process needs to be trained or modeled by agents or humans (Azevedo & Cromley, 2004; Goldman et al., 2003). The iDRIVE system models the asking of deep questions with dialogues between agents. Virtually any content (information in a textbook, ideas in a lecture, a list of expectations that answer our example physics problems) can be augmented by preceding each chunk of content with a deep question that motivates the content. When the content is laced with a large amount of deep questions, the learner acquires the mindset of thinking deeply.

Most of the iDRIVE studies have been conducted on college students in the area of computer literacy (Craig, Driscoll, & Gholson, 2004; Craig, Sullins, Witherspoon, & Gholson, 2006). Learning gains on the effectiveness of iDRIVE on question asking, recall of text, and multiple-choice questions have shown effect sizes that range from 0.56 to 1.76 compared to a condition in which students listen to the monologue on the same content without questions . We have recently conducted a study on 9^{th} and 11^{th} graders who learned Newtonian physics. The students were randomly assigned to one of three conditions: iDRIVE dialogues, AutoTutor, and monologues (Gholson et al., in press). Attempts were made to achieve information equivalence by covering the same set of expectations in all conditions. The results of the study showed pre-to-posttest effect sizes of .95, .43, and .58 sigma in the iDRIVE, AutoTutor, and monologue conditions, respectively. The fact that the iDRIVE dialogues showed the highest learning, even higher than AutoTutor, supports the claim that modeling deep question asking with agents can have a powerful impact on physics comprehension. We consider this encouraging news for researchers who are exploring how agent technologies can improve science learning in young children.

Learning the Principles and Shedding the Misconceptions

One way of viewing Newtonian physics is that there are a set of physics principles that constrain the reasoning of students as they work on toy problems or real-world applications (Hunt & Minstrell, 1996; VanLehn et al., 2007). Students should follow these principles or articulate them when asked to explain their reasoning. Their actions and explanations should not exhibit the misconceptions that reflect everyday intuitions and faulty mental models.

We have identified approximately 40 principles and a similar number of misconceptions that are affiliated with Newtonian laws of motion and mechanics. This content can be clustered in the following broad categories that arguably can be ordered on complexity and a prerequisite continuum. An example principle (P) and misconception (M) is provided for each category.

1. Meanings of displacement, velocity, acceleration, kinematic relationships for single moving bodies.
   P: Constant velocity implies zero acceleration.
   M: Zero acceleration implies zero velocity.

2. Newton's first law (inertia)
   P: Air resistance can be ignored under certain circumstances.
   M: Air resistance can never be ignored.

3. Newton's second law (net F = ma, net force equals mass times acceleration)
   P: The same force will accelerate a less massive object more than a more massive object.
   M: Accelerations of both objects are equal during interaction.

4. Gravity and contact forces
   P: All objects in freefall have the same acceleration.

M: Heavier objects fall faster**.**

5. Newton's third law (action and reaction forces)
   P: When A and B exert a force on each other, the magnitude of the two forces is equal.
   M: A smaller object exerts no force on a larger object.

Many physics books order the chapters or lessons in the above order. It is unclear whether this is mere convention, a historical accident, or a custom motivated by pedagogy. It is intuitively obvious that there should be some ordering on prerequisites. It is difficult to imagine how Newton's laws could be understood without at least a preliminary understanding of the definitions of displacement, velocity, and acceleration. However, it is not intuitively obvious that Newton's first law must precede the second and third laws. We are convinced that the ordering of these lesson categories needs to be justified by a pedagogical theory that is strongly rooted in a development science. This volume, if course, was inspired by this proclamation.

Aside from the ordering of the principles in the sequence of lessons, it is important to assess how reliably the students apply these principles when they solve problems. In an ideal world, the students would consistently apply the principles correctly as they receive hundreds of problems (toy or real-world) over time. They would also discontinue being seduced by misconceptions and faulty mental models. However, the ideal world is a very long distance from reality. The best we can hope for is a statistical progression over time and the learning history that eventually increases the probability of correct application of principles and decreases the probability of misconceptions. Occasionally the students manifest all-or-none learning, but that is more of a special case than the normal state of affairs.

In either case, one of the challenges for a developmentally inspired pedagogy is to select example problems in an intelligent manner that optimizes learning. If a student has already mastered a principle consistently, it is presumably a waste of time to select problems that address that principle. Instead, the selected problems should address the deficits of the individual student. Computer technologies will undoubtedly be required to precisely track the learning history of individual students and to generate problems that optimize learning. It would be much too tedious for a human to keep track of such details.

We remain optimistic about the pedagogical value of learning environments with conversational agents, interactive simulations, and serious games. These should be the important routes to accelerated learning of physics in children. The agents are firmly grounded in social interaction, can model excellent learning, can intelligently respond to individual learners, and are captivating to young children. The interactive simulations make difficult concepts perceptually visible and allow the learner to actively manipulate the environment. The games keep the students motivated for hours on end as they simultaneously acquire difficult conceptualizations with academic and cultural value. All we need to do now, at this magic moment in the history of technology, is to build the systems, test them, and disseminate the successful ones to relevant educational communities. The learning, cognitive, discourse, and developmental sciences seem ready to hunker down and take up the challenge. As we take up the challenge, we will continue to remind ourselves of the long parade of Don Quixote’s in education.

Author Notes

The research on AutoTutor was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918), the Institute of Education Sciences (R305H050169, R305B070349), and the DoD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, DoD, or ONR. Requests for reprints should be sent to Art Graesser, Department of Psychology, 202 Psychology Building, University of Memphis, Memphis, TN 38152-3230, a-graesser@memphis.edu.

References

Ainsworth, S. (2008). How do animations influence learning? In D.H. Robinson and G. Schraw (Eds.), *Recent innovations on educational technologies that facilitate student learning* (pp. xxx-xxx). Charlotte, NC: Information Age Publishing.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*, 416-427.

Azevedo, R., & Cromley, J.G. (2004). Does training on self-regulated learning faciliate students' learning with hypermedia. *Journal of Educational Psychology*, 96, 523-535.

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*, 95-115.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161--199.

Chi, M. T. H., Siler, S. A., Jeong, H. (2004). Can tutors monitor students' understanding accurately? *Cognition and Instruction, 22*, 363-387.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471-533.

Cohen, P. A., Kulik, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.

Corbett, A.T. (2001). Cognitive computer tutors: Solving the two-sigma problem. *User Modeling: Proceedings of the Eighth International Conference, UM 2001, 137-147.*

Craig, S. D., Driscoll, D., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia, 13*, 163-183.

Craig, S. D., Sullins, J., Witherspoon, A. & Gholson, B. (2006). Deep-level reasoning questions effect: The role of dialog and deep-level reasoning questions during vicarious learning. *Cognition and Instruction, 24*, 565-591.

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., & the Tutoring Research Group (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education*, *11*, 242-253.

De Jong, T. & Van Joolingen, W.R. (1998) Scientific discovery learning with computer simulations of conceptual domains *Review of Educational Research*, *68*, 179-201.

di Sessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10(2 & 3), 105-225.

D'Mello, S. K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., and Graesser, A. C. (2008). Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User-Adapted Interaction*, 18(1-2), 45-80.

D'Mello, S.K., Picard, R., & Graesser, A.C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems, 22*, 53-61.

Dodds, P., & Fletcher, J. D. (2004). Opportunities for new "smart" learning environments enabled by next-generation web capabilities. *Journal of Educational Multimedia and Hypermedia, 13*(4), 391-404.

Gee, J. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gentner, D., & Stevens, A.L. (1983)(Eds.). *Mental models*. Hillsdale, NJ: Erlbaum.
Gholson, B, Witherspoon, A., Morgan, B., Brittingham, J., Coles, R., Graesser, A.A., Sullins, J., & Craig, S.D. (in press). Exploring the deepl-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer literacy and Newtonian physics. *Instructional Science*.
Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S., & Tzou, C. T. (2003). Science inquiry in a digital age: Possibilities for making thinking visible. In H. van Oostendorp (Ed.), *Cognition in a digital world.* (pp. 253-284) Mahwah, NJ: Erlbaum.
Graesser, A.C., Chipman, P., Haynes, B.C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education, 48*, 612-618.
Graesser, A.C., D'Mello, S.K., Craig, S.D., Witherspoon, A., Sullins, J., McDaniel, B., & Gholson, B. (2008). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research*, 19, 293-312.
Graesser, A.C., Jackson, G.T., Mathews, E.C., Mitchell, H.H., Olney, A.,Ventura, M., Chipman, P., Franceschetti, D., Hu, X., Louwerse, M.M., Person, N.K., & TRG (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In R. Alterman and D. Hirsh (Eds.), *Proceedings of the 25rd Annual Conference of the Cognitive Science Society* (pp. 1-6). Boston, MA: Cognitive Science Society.
Graesser, A.C., Jackson, G.T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology, 47*, 19-22.
Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers, 36*, 180-193.
Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist, 40*, 225-234.
Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104-137.
Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*, 359.1-28.
Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14*, 343-351.
Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141-158.
Hewitt, P. G. (1987). *Conceptual Physics*. Menlo Park, CA: Addison-Wesley.
Hunt, E., & Minstrell, J. (1996). A collaborative classroom for teaching conceptual physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and the classroom*. Cambridge, MA: MIT Press.
Jackson, G.T., Olney, A., Graesser, A.C., Kim, H.J. (2006). AutoTutor 3-D Simulations: Analyzing user's actions and learning trends. In R. Son (Ed.), *Proceedings of the 28th Annual Meetings of the Cognitive Science Society* (pp. 1557-1562). Mahwah, NJ: Erlbaum.
Johnson, W.L., & Beal, C. (2005). Iterative evaluation of a large-scale intelligent game for language learning. In C. Looi, G. McCalla, B. Bredeweg, and J. Breuker (Eds.), *Artificial Intelligence in Education: Supporting learning through intelligent and socially informed technology* (pp. 290-297). Amsterdam: IOS Press.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Klahr, D. (2002). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Kozma, R. (2000). Reflections on the state of educational technology research and development. *Educational Technology Research and Development, 48*, 5-15.

Landauer, T., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.) (2007). *Handbook on Latent Semantic Analysis.* Mahwah, NJ: Erlbaum.

Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.). *Aptitude, learning and instruction. Volume 3: Conative and affective process analyses* (pp. 223-253). Hillsdale, NJ: Erlbaum.

Mayer, R.E. (2005). *Multimedia Learning*. Cambridge, MA: Cambridge University Press.

McNamara, D.S., Levinstein, I.B. & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers, 36*, 222-233.

Millis, K.K., Magliano, J., Britt, A., Wiemer-Hastings, K., Halpern, D., & Graessere, A.C. (2006). Acquiring research evaluative and investigative skills (ARIES) for scientific inquiry. Unpublished manuscript, Northern Illinois University.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology, 97*(1), 117-128.

O'Neil, H.F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*, 455 – 474.

Otero, J., Leon, J.A., & Graesser, A.C. (2002)(Eds). *The psychology of science text comprehension*. Mahwah, NJ: Erlbaum.

Ploetzner, R., & VanLehn, K. (1997). The acquisition of informal physics knowledge during formal physics training. *Cognition and Instruction, 15*, 169-206.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, televisions, and new media like real people and places*. Cambridge, U.K.: Cambridge University Press.

Ritterfeld, U., Cody, M. & Vorderer, P. (in press)(Eds.), *Serious games: Mechanisms and effects*. Mahwah, NJ: Routledge, Taylor and Francis.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181-221.

Rus, V., McCarthy, P.M., McNamara, D.D., & Graesser, A.C. (in press). A study of text entailment. *International Journal on Artificial Intelligence Tools*.

Stein, N.L., Hernandez, M., & Gamez, P.. Making the invisible visible: The conditions for the early learning of physics. Symposium presentation at the Society for Text and Discourse, Glasgow, Scotland.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*, 247-262.

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*(3), 199-212.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31, 3-62.

VanLehn, K., Lynch, C., Taylor, L., Weinstein, A., Shelby, R. H., Schulze, K. G., et al. (2002). Minimally invasive tutoring of complex physics problem solving. In S. A. Cerri, G. Gouarderes & F. Paraguacu (Eds.), *Intelligent Tutoring Systems, 2002, 6th International Conference* (pp. 367-376). Berlin: Springer.

Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology and Society, 8*, 54-65.

Vitale, M. R., & Romance, N. R. (2007). A knowledge-based framework for unifying the content-area reading comprehension and reading comprehension strategies. In D. McNamara (Ed.), *Theories of text comprehension: The importance of reading strategies to theoretical foundations of reading comprehension* (pp. 73-104). Mahwah, NJ: Erlbaum.

Vorderer, P., Bryant, J., Pieper, K.M., & Weber, R. (2006). The (computer) games people play. In P. Vorderer and J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 1-7). Mahwah, NJ: Erlbaum.

White, B. Y & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction, 16(1)*, 1998, 3-118.

Table 1. *Example dialogue with AutoTutor*.

TUTOR: Now let's move on. A lightweight car and a massive truck have a head-on collision. On which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion? Explain why.

STUDENT: The lightweight vehicle experiences the greater impact and the greater change in motion because the mass of the lighter vehicle is less, therefore having less inertia to resist the change in motion. The momentum of the truck is greater than that of the car.

TUTOR: Right! What does Newton's third law of motion say about the forces that bodies exert on each other?

STUDENT: The forces must be equal in magnitude and opposite in reaction.

TUTOR: Yes, correct! Can you add to that?

STUDENT: The forces must be equal and opposite reactions acting on opposite bodies.

TUTOR: Ya! We've almost covered it, but there's one part I think is still missing. I am sure you can get this. During an interaction, each body exerts an equal but opposite force on the other ____?

STUDENT: body

TUTOR: Yes, right! on the other body. So when two balls, one plastic and the other iron, hit each other, does the heavier or neither ball exert the greater impact force on the other?

STUDENT: neither

TUTOR: Yes, correct! Neither, the forces are equal in magnitude. Let's take a moment for general physics questions. Ask me one and I'll do my best to answer it.

STUDENT: no

TUTOR: Alright. To recap what we have just discussed, please give me your complete answer to the original question, as presented at the top of the screen.

STUDENT: Although the lighter car experiences a greater change in motion, both vehicles exert an equal impact force, but opposite in direction. The larger object has a larger mass, and therefore a larger inertial resistance than that of the car.

TUTOR: That is right! Putting everything together, um, According to Newton's third law of motion, if one body exerts a force on the other body then the other body must exert an equal and opposite force on the first body. Therefore, the car and the truck both experience the same impact force. By Newton's second law, the car will experience the greater change in its motion because its mass is much smaller. This question is finished so let’s go on to the next question.

Figure Caption Page

Figure 1: A computer screen of AutoTutor on conceptual physics with interactive 3D simulation.