How does the Human Mind Think and Learn about Feedback Loops?

Kim A. Kastens

Lamont-Doherty Earth Observatory of Columbia University
(kastens@ldeo.columbia.edu)

and

Thomas F. Shipley

Department of Psychology Temple University
thomas.shipley@temple.edu

2023 Systems Dynamics Society Conference

Abstract

This paper considers how concepts and approaches from cognitive and learning sciences might shed light on how humans think and learn about feedback loops. Cognitive biases that might interfere with one's ability to accurately form individual causal links include tendencies to perceive covariation that is direct rather than inverse, to overweight information about presence rather than absence, to weight a plausible mechanism over empirical evidence, and to fail to account for what other influences might matter. Cognitive limitations that might compromise one's ability to merge individual links into a closed loop include working memory limitations, tendency to see sequences as linear chains, tendency to look for explanations of phenomena at the level of components, and difficulty comprehending exponential growth. On the more optimistic side, cognitive affordances that make feedback loop thinking possible include analogical reasoning, language and categorization, ability to create runnable mental models, and distributed cognition. The paper concludes that the realm of how humans think and learn with and about feedback loops is ripe for further cognitive, learning science, and neuroscience research, and suggests research questions that have the potential to both advance systems education practice and elucidate under-researched capabilities of the mind.

Problem Statement:

The authors have not been able to locate a body of literature or a community of scholars in the domain of discipline-based education research (DBER) focused on research into how the human mind thinks and learns about feedback loops or other key aspects of systems thinking. DBER combines rigorous methods of cognitive and learning science research with deep grounding in discipline-specific knowledge, practices, priorities and worldview of a scientific discipline (National Research Council, 2012). This paper aims to catalyze further DBER-style research on feedback loop thinking, and to surface researchers who are already working on this or related topics in hopes of coalescing a community of practice (Kastens & Manduca, 2017).

Methodological Approach

Discipline-based Education Research (DBER) is now well established in physics, chemistry, life sciences and geosciences. This paper's authors were among the founders and early advocates for the field of geoscience education research (GER), with a particular focus on spatial thinking (e.g. Kastens, et al, 2009; Shipley, et al, 2013; Resnick, et al, 2018). Although each of the DBERs followed a somewhat different historical pathway to its current status, each grew from a body of "practitioners' wisdom," comprising informal observations shared among highly-motivated educators about common student misconceptions, effective teaching strategies, student attitudes, and so on. The Creative Learning Exchange (http://www.clexchange.org) is rich in such practitioners' wisdom for systems thinking in K-12 education.

During its formative years, each of the DBERs benefited from a process of introspection and collaborative discourse, during which were identified aspects of thinking and practice that were important in the discipline and that could be conceptualized in terms of known capabilities of the mind. Kastens & Ishikawa (2006) and Kastens & Manduca (2012) are exemplars of this approach and the model for the current paper.

We focus this paper around feedback loops, because they are extremely important in systems that are consequential for humanity, are central to systems dynamics, and present interesting cognitive challenges. We begin with cognitive limitations that may make it hard to think about feedback loops, move on to cognitive affordances that nonetheless make it possible, and conclude with ideas towards an ambitious research agenda of foundational and applied research on feedback loop thinking. Our intent is not to be all-encompassing, but rather to make the case that there is an underpopulated research domain with fascinating research questions and the potential to both improve systems education practice and elucidate under-appreciated capabilities of the mind.

Cognitive Limitations:

Readers who sometimes wear an instructor's hat will probably have encountered learners who struggle to understand the concept of feedback loops in general, or to distinguish between reinforcing and balancing loops, or to recognize feedback loops in real world contexts, or to use the feedback loop concept to explain real world phenomena, or to use feedback loops to make predictions about how a system will behave under not-yet-observed conditions — or all of the above. This section of the paper explores some of the cognitive limitations that may make such thinking challenging.

We begin with cognitive biases and limitations that interfere with the ability to create an isolated A-->B causal link that is veridical and plausible, and then expand to problems that emerge when people try to think about a feedback loop in its entirety. Further problems surely emerge when trying to think about multi-loop systems, but those are beyond the scope of this paper.

Problems and biases in considering single A-->B causal links

Every feedback loop is made up of multiple individual links of the form A causes or influences B. In reaching a conclusion that A influences B, humans rely on three sources of insight: observations that A and B consistently covary, pre-existing knowledge about plausible causal mechanisms, and counterfactual reasoning about what else might be responsible for B. Each of these sources of information can go astray.

Tendency to perceive covariation that is direct rather than inverse: Kareev (1995) published both a theoretical analysis and empirical evidence that humans find it easier to discern a positive (direct) correlation than a negative (inverse) correlation between observable attributes within a series of events or a group of objects. In the objects task, he asked participants to draw envelopes, one by one, from an opaque bag. Each envelope was colored red or green, and each contained a small coin marked with either an X or O. After drawing the envelope, the participant had to guess whether it would contain an X or O coin. They were then allowed to open the envelop and look at the coin. Across the 56 student participants, the experimenters varied the relative frequency of the X's and O's (symmetry) and the strength and sign of the relationship between the predictor (envelope color) and the criterion (X/O). Over the course of 128 envelopes, participants began to pick up the nature of the system they were exploring and guess correctly more often. Across the full range of relationships, from strongly negative, through zero, to strongly positive, the participants' guesses in the final quartile of the data collection indicated that they had inferred a more positive correlation between color and coins than the actual correlation they had experienced. Might this cognitive bias make it harder for learners to perceive "opposite" links (aka "O" or "-" links) than "same" links (aka "S" or "+" links)?

Tendency to overweight information about presence rather than absence: Events that happened or objects that are present are more salient than events that didn't happen or objects that are not present (Spellman & Mandel, 2005). A famous example is in the Sherlock Holmes story, *The Adventure of Silver Blaze:* "Inspector Gregory of Scotland Yard: Is there any other point to which you would wish to draw my attention? Holmes: To the curious incident of the dog in the night-time. Gregory: The dog did nothing in the night-time. Holmes: That was the curious incident." Gregory, and everyone else in the story, had failed to note that the dog had not barked. But Holmes did notice this detail, and thus deduced that the horse-thief had been someone that the dog knew well, the horse's trainer (Doyle, 1894). This bias might make it easier to perceive links that enable an effect and harder to perceive links that disable an effect.

Tendency to weight a plausible mechanism over empirical evidence: The strongest causal links are those that are underlain by both a plausible mechanism that the influence *should* occur and observational evidence that the influence *does* occur. However, these two kinds of information come by different pathways, may solidify at different times, and may be handled differently by the mind. A robust body of research by psychologist Deanna Kuhn and her students makes the case that when humans judge whether one phenomenon is or is not a cause of another phenomenon, they tend to weight plausible theory more heavily than empirical evidence.

Kuhn's experimental design is to present a group of participants with information about an idealized system in which there is one observable outcome and multiple potentially causal factors. Kuhn's team has presented variants of this experiment to all sorts of people: children, undergraduates, jurors, people sitting in the waiting room of a train station (Kuhn, 2001, 2004, 2007, 2010; Kuhn, et al, 2000). They have varied the nature of the system, using both natural phenomena (e.g. floods, earthquakes) and human-made (e.g. parties, factories). The generalizable finding is that if a person believes a plausible, logical story about why A *should* cause B, they are likely to maintain that belief even in the presence of data and observations more consistent with an interpretation that A *does not* cause B. Conversely, if a person has a plausible mechanism for why A *should not* cause B, they are likely to maintain that belief even in the presence of data and observations supporting that A *does* cause B. Does this bias come into play when deciding which links have sufficient empirical evidence to deserve to be included in an emerging conceptual model?

Failure to account for what other influences might matter: Ruth Byrne has extensively researched (summarized in Byrne, 2007) how people imagine what might have happened under varying circumstances (counterfactuals). Imagining what might have happened under varying circumstances is what systems thinkers do when they envision a nudge to one node in a causal loop diagram and then step around the loop, imagining the result of each link. With this in mind, several of Byrne's findings are relevant. She reports that approximately a quarter of participants see only one possibility and don't explore what other influences might matter. When prompted in various ways to imagine what else might have had causal influence, Byrne's participants tend to cite things over which they have agency rather than things over which they have no agency, and socially acceptable actions rather than socially unacceptable actions.

Problems and limitations in considering a loop in its entirety

After a reasoner has a clear grasp of each individual A-->B causal link within a loop, a new set of limitations emerge when it comes time to merge the individual links into a feedback loop. First, the working memory capacity of the human mind is severely limited, unable to hold the nodes and linkages of a loop in mind for any but the simplest feedback loop. Secondly, humans (at least in western cultures) seem to prefer to envision a sequence of events as a linear chain, and resist envisioning them as a loop. Finally, perhaps as a consequence of the first limits, people are notably poor at anticipating system behavior when it exibits exponential growth or decay.

Working memory limitations: Working memory refers to information that can be held in mind in a readily accessible form, and used in the execution of cognitive tasks. Working memory facilitates planning, comprehension, reasoning and problem solving (Burmester, 2017; Cowan, 2014). Working memory is active, interacting with the world to allow one to coordinate actions and plans. However, working memory capacity is extremely limited, in the range of 3 to 5 items for most people. The simplest possible feedback loop contains two nodes and two links for a total of four items. Even the simplest feedback loop would seem to push the limits of what a person can think about in working memory. Many of the things that one might want to use feedback loop thinking to accomplish--such as inferring mechanisms, predicting behaviors, and evaluating alternatives--are working memory intensive tasks.

People see temporal cycles as linear chains, not loops: Stillings (2012) hypothesized that learners' difficulties with feedback loops may arise, in part, because their notions of causality are intermingled with their notions of time. One important notion of causality is the expectation that if A happened before B, then A can have caused or influenced B, but B cannot have caused or influenced A. But feedback loops, in which C connects back to A, defy this expectation. A body of work by psychologist Barbara Tversky and students has established that people have difficulty envisioning a series of events in time as having a looped or cyclical structure (Noel & Tversky, unpublished; Kessel, 2008; Tversky, 2019; Tversky & Jamalian, 2019). In one illustrative task, Tversky's team presented undergraduates with a verbal description of four events that formed a temporal sequence (e.g. Break eggs in bowl; Add milk; Stir; Cook in frying pan) or a temporal cycle (e.g. Wake up; Go to work; Come home; Go to sleep.) Students were asked to draw a simple diagram conveying the information presented. Drawings almost all showed pictures or boxes, connected by lines or arrows. Drawings were coded according to whether the sequence of lines or arrows ended or returned to the beginning to indicate a repeating cycle. In response to the temporal *sequence* stimulus, students overwhelmingly (>95%) drew an end diagram. For the temporal *cycle*, most students also drew an end diagram; only 20% drew a repeat diagram. The researchers then tried several manipulations to make the cyclic nature of the repeating sequences more salient. Even for the strongest manipulation, more than 40% of participants produced a linear diagram. Since concepts of time and causality are so tightly coupled, it might be that reluctance to envision a temporal cycle as a looped structure spills over into reluctance to envision a causal chain as a looped structure. In our own pilot studies, when we asked undergraduates to draw a diagram depicting the information in a short loop-containing narrative, the most common failure mode was failing to "close the loop," similar to Tversky's participants.

People look for explanations of phenomena at the level of components: According to a body of research by two psychologists and a philosopher, people tend to gravitate towards explanations at the component level rather than the systems level (Hopkins, et al, 2016, 2019; Weisberg, et al, 2018). In one illustrative sets of experiments, the researchers gave adult participants narrative explanations of a research finding and asked them to rate the quality of the explanation on a 7-point scale. Explanations in the language of physics were more highly ranked than explanations in the language of chemistry, which in turn were more highly ranked than cell biology explanations, and so on. The researchers summarized their body of research as indicating a bias towards explanations that are at a more "reductive" level. If their finding generalizes, a bias towards more reductive explanations could interfere with people's willingness to look for explanations at systemic levels. For example, a bias towards component-level explanations might incline a person to interpret an observation that over time the poor get poorer while the rich get richer as a consequence of individual effort or lack of same, overlooking the Matthews Effect reinforcing feedback loop.

People struggle to comprehend exponential growth: Psychologist Daniel Kahneman (2020) whose career has been spent thinking how humans make good and poor decisions, describes in an interview his own failure to comprehend the behavior over time of a reinforcing feedback loop. Interviewed early in the Covid-19 pandemic, he said: " This is an exponential event: what is we see things doubling every two days, every three days, every four days. And people, certainly including myself, don't seem to be able to think straight about exponential growth. I knew that there were 100 cases in France, and I was about to fly to France. I also knew that epidemics are exponential. I didn't even consider the fact that in a month it would have increased by a factor of 1000…. All of this is I think beyond intuitive human comprehension. This is interesting, that we are in a situation that we are simply not equipped to understand…"

Cognitive Affordances:

If there are so many obstacles, limitations, and biases that make it difficult for the human mind to think with and about feedback loops, how then do we do it? The mind has a toolbox of cognitive affordances that can be called in play. Using analogical reasoning, humans can use what they know about a system they partially understand to make inferences about a system they don't yet understand. They can imagine worlds that they have not personally experienced, including worlds of the future and worlds that operate according to different ground rules. Humans have the capacity to create "mental models," which are internal representations of an aspect of the world, fully fashioned with a suite of interacting elements. And humans can "run" those mental models, set those elements to interacting with each other, and watch the results play out. When confronted with puzzles too complex for one mind to handle, humans can distribute cognition across multiple minds, or across minds plus tools and artifacts.

Analogical reasoning:

In cognitive science, "analogs" are two or more situations or events or phenomena that are similar to each other in important ways. "Analogical reasoning" refers to the process by which the human mind can notice important similarities between analogs and use then use those similarities to generate new inferences (Gentner & Smith, 2012). This powerful cognitive capacity is one of the ways that humans can create new hypotheses and new explanations--new knowledge that goes beyond what they have learned from other humans.

Verbal analogy: The simplest and most familiar form of analogical reasoning is "verbal analogy", when one person tells another person that a phenomenon or situation under consideration is like something that the listener already understands. For example, a journalist for the Economist writes: "Even profound changes in what it means to be a Democrat or Republican seem to return the parties to their equilibrium, *as though obeying some thermostat*" (emphasis added) (Lexington, 2023). Verbal analogies require an explaining person to offer the analogy as a gift to the person in need of understanding. The next two forms of analogy enable a person in need of understanding to create their own understandings even in the absence of an explaining mentor/author/teacher.

Projective analogy: In this form of analogical reasoning, the reasoner begins with a fairly solid understanding of one analog. The reasoner identifies ways in which another situation is somewhat aligned with the known situation. They can then project understandings from one analog onto the other to create new inferences which fill in places where there had been gaps in their understanding (Kurtz, Miao, & Gentner, 2001). For example, a social worker may be familiar with one situation in which the rich get richer while the poor stay poor. If they encounter a new situation with similar behavior, they may be able to use their understanding of the first situation to more quickly build an understanding of the second situation. As they learn more about the second situation, they may then also be able to project those understandings to better understand the first situation.

Extracting the Schema: In this most powerful level of analogical reasoning, two or more phenomena under consideration have a similar suite of interconnected and important relationships. The reasoner discerns and articulates the nature of these relationships, called the "schema." The extracted schema can then be used to solve problems or understand phenomena any time the reasoner is confronted by a novel situation that has the same underlying schema. In a classic cognitive psychology experiment (figure 1), participants were asked to imagine that they are a doctor faced with patient who has a malignant tumor in his stomach (Gick & Holyoak, 1980, 1983). Participants were given certain information about a form of radiation that could destroy the tumor at a certain dosage--but would also destroy healthy tissue. Their task was to come up with a procedure that would destroy the tumor without destroying surrounding healthy tissue. The key was to realize that the radiation had to be directed at the tumor from multiple directions; where the rays converged, the tumor would be destroyed. The experimenters gave

The challenge for all participants: Radiation Problem

Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. It is impossible to operate on the patient, but unless the tumor is destroyed the patient will die. There is a kind of ray that can be used to destroy the tumor. If the rays reach the tumor all at once at a sufficiently high intensity, the tumor will be destroyed. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At low intensities the rays are harmless to healthy tissue, but they will not affect the tumor either. What type of procedure might be used to destroy the tumor with the rays, and at the same time avoid destroying the healthy tissue?

Analog 1: The General

A small country was rule from a strong fortress by a dictator. The fortress was situated in the middle of the country, surrounded by farms and villages. Many roads lead to the fortress through the countryside. A rebel general vowed to capture the fortress. The general knew that an attack by his entire army would capture the fortress. He gathered his army at the head of one of the roads, ready to launch a full-scale direct attack. However, the general then learned that the dictator had planted mines on each of the roads. The mines were set so that small bodies of men could pass over them safely, since the dictator needed to move his troops and workers to and from the fortress. However, any large force would detonate the mines. Not only would this blow up the road, but it would also destroy many neighboring villages. It therefore seemed impossible to capture the fortress.

However, the general devised a simple plan. He divided his army into small groups and dispatched each group to the head of a different road. When all was ready, he gave the signal and each group marched down a different road. Each group continued down its road to the fortress so that the entire army arrived together at the fortress at the same time. In this way, the general captured the fortress and overthrew the dictator.

Analog 3: Red Adair

An oil well in Saudi Arabia exploded and caught fire. The result was a blazing inferno that consumed an enormous quantity of oil each day. After initial efforts to extinguish it failed, famed firefighter Red Adair was called in. Red knew that the fire could be put out if a huge amount of fire retardant foam could be dumped on the base of the well. The was enough foam available at the site to do the job. However, there was no hose large enough to put all the foam on the fire fast enough. The small hoses that were available could not shoot the foam quickly enough to do any good. It looked like there would have to be a costly delay before a serious attempt could be made.

However, Red Adair knew just what to do. He stationed men in a circle all around the fire, with all of the available small hoses. When everyone was ready all of the hoses were opened up and foam was directed at the fire from all directions. In this way, a huge amount of foam quickly struck the source of the fire. The blaze was extinguished, and the Saudis were satisfied that Red had earned his three million dollar fee.

Diagram seen by some participants

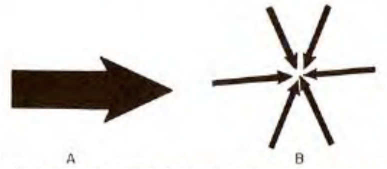

Figure 1: Participants were most successful on the radiation task when they had prepared by reading two narratives with problem-solving structure analogous to the task ahead of them and then written a description of how the two stories were similar. This preparation activated the analogical reasoning process. (Figure modified from Gick & Holyoak, 1983).

variants of the same task to many groups of participants. Far and away the most successful participants were those who were first given two narratives to read in which other protagonists had faced other problems that could be solved by convergence upon the problem from multiple directions at once. After reading the two narratives, they were asked to describe in writing, as clearly as possible, how the two stories were similar. By carefully comparing those two analogous narratives, successful participants "extracted the schema" of directing elements of the solution (be it radiation, or soldiers, or firefighting foam) from multiple directions onto a centrally situated problem area. Reading one narrative rarely did the trick. The key was that as participants looked for similarities between the two narratives, surficial features dropped away leaving exposed the underlying shared structure of convergence.

We think that skilled systems thinkers have extracted robust schema for balancing and reinforcing feedback loops from their long experiences with multiple loop-containing systems. A foundational learning objective for feedback loop education is for learners to extract these schema for themselves, and then be able to project their understandings from this generalized schema into new situations. We are currently testing curriculum materials that guide undergraduates through the process of extracting the underlying feedback loop schema from short narratives about systems with differing surficial features.

Language and categorization:

Analogical reasoning makes it possible to do feedback loop thinking without using or even knowing the term "feedback loop." The journalist in the "thermostatic" verbal analogy example offered the reader his insights about politics while avoiding the term "feedback loop." The social worker in our projective example was able to build their own deeper insight, even if they didn't know the term. We suspect that an intuitive understanding of feedback loops existed in human culture many generations before engineers began building them or scientists began researching them. Folklore provides a clue. For example, the Windigo tales of Algonquian-speaking Native Americans describe a human who has transformed into a monster trapped in a positive feedback loop of greed and desire (Kimmerer, 2015). The Windigo eats human flesh, and the more he eats the more he craves. The Windigo tale is taught to children with the intent that they will project their understanding of the Windigo dynamic onto novel situations to help them make prudent decisions when tempted by greed or desire.

However, language helps enormously. Gentner (2010) makes the case that analogical reasoning in humans benefits from a reinforcing feedback system with language. Language supports analogical reasoning in multiple ways: Use of the same words (i.e. "leads to") can make parallelisms more obvious. If a type of relation is called by a consistent label (i.e. "causal link"), one thinks of it as a consistent thing. Use of a term (i.e. "feedback loop") can reify an abstraction and make it easier to recall. Conversely, analogical reasoning supports language learning in children, including acquisition of individual words and of grammar.

Although terms are powerful, giving a learner the term "feedback loop" or the term plus a definition does not substitute for the active-learning process of extracting the schema through interactions with multiple instances of varied loops. The term "feedback loop" encompasses a complicated category of phenomena. As recounted by George Richardson (1991, see especially his figure 3.1), humanity collectively took centuries to build the modern concept of "feedback loop," from the engineered water flow systems of ancient Greece, through various parallel threads of thought in math, econometrics, engineering, social sciences, biology, and logic.

Sorting related aspects of reality into categories, and assigning verbal labels to the categories, helps the human mind cope efficiently with the complexity of the experienced world. If a novel phenomenon can be placed into an established category, a previously learned set of insights about members of the category can be brought into play. In other fields, including medical diagnosis and natural sciences, there has been research on how to support learners in categorizing phenomena that manifest as complex and subtle observations (Papa, et al., 2007; Wahlheim, et al, 2012; Nosofsky, et al., 2019). One robust finding is that if the goal is for learners to be able to recognize not-before-seen instances of a category, it is more effective to train them on a suite of examples that span the full range of attributes of the category rather than on a similar number of training examples that cluster close to the "center" of the category. Discovering what are the attributes and where is the center of psychologically-important dimensions of the categories in other disciplines has been a research project in and of itself. If applied to feedback loop instruction, we speculate that the "center" would be occupied by reinforcing feedback loops with no "opposite" (negative) links and balancing feedback loops with only one "opposite" link.

Runnable mental models:

Another relevant mental capacity that characterizes human cognition is a prodigious capacity to think about times and places that we have not personally experienced--such as the inside of a cell, or a faraway country, or the future. Thinking about the future is one of the most compelling reasons to engage in feedback loop thinking, grappling with questions such as how fast will the pandemic spread, or what will the climate be like in 2050.

The mind constructs a narrative about a situation of interest by imagining and combining objects, events, thoughts, and feelings. Such imagining is supported by memories of past experiences, but also includes elements that may not exist in the present and indeed may never have existed. To handle new situations, humans have evolved the capacity to extract the critical parts of memories and recombine them in new ways.

Memories are believed to be stored in pieces (Dumper, et al, n.d.; diSessa, 2018). The components that make up our memories of people, places, and things are a combination of elements and relationships among elements. For example, a memory of a house may include the elements of doors and furniture, but also the relationships among the pieces of furniture. The relationships are extractable from memories in a way that allows them to act in an abstract manner to represent relations among new elements. For example, having seen many living rooms allows a human to extract an abstract schema of living rooms that can be used to imagine what a living room could look like if furnished with novel objects, for example, rocks in the *Flintstones* television show (figure 2).

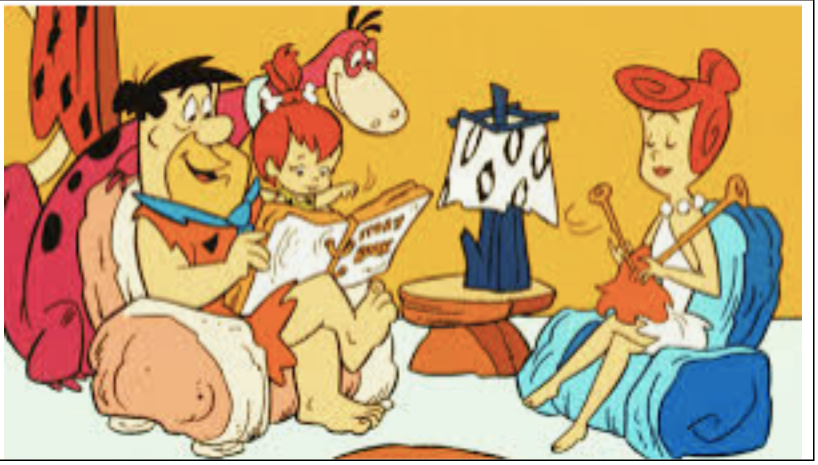

Figure 2: The creators of The Flintstones television series were able to imagine living room furniture made out of rocks, by combining attributes of "rock" with attributes of "living room" in a new way.

When a new object is imagined in a familiar scene the mind draws by analogy on known attributes about both object and scene. The living room schema in the Flintstones example retains the familiar spatial relationships among husband's chair, wife's chair, side table and lamp, as well as the geometry of seat/back/arms within the chair object. The rock schema features massiveness, rough surface, and lack of square corners. To imagine a rock-chair, aspects of the rock schema are slotted into the living room schema, to create an imagined world with never-before-seen furniture.

Imagining never-before-seen objects and rooms is only the beginning of what human minds can do. They can populate these scenes with agents who have feelings and motivations and make decisions and take consequential actions. They can combine objects, scenes and characters into entire elaborate scenarios in which events play out over time. They can contemplate more than one possible future scenario, and judge whether one is more desirable than another. Crucially for feedback loop thinking, the characters and objects in mental models often — although not always — respond to the same influences and act by the same causal mechanisms as did the characters and objects in the memories from which the model was constructed.

We refer to this entire collection of imagined objects, scenes, events, scenarios, and characters as "mental models," and use the term "runnable mental model" to refer to models that can be used to mentally simulate behavior over time. When a feedback loop thinker steps their way around a causal loop diagram explicating what they think is happening at each link, and then describes the net outcome of the loop as a whole, they are running mental models.

Unfortunately, the process by which the mind makes mental models remains deeply mysterious. However, the finding that mental models are created from fragments of remembered knowledge, often in a generalized or abstracted form, could provide a valuable clue. What are the elements, the abstractable fragments of knowledge, that experts and novices draw on when creating a mental model of a feedback loop? Can we hypothesize that memories of being caught up in a traumatizing feedback loop (e.g. pandemic, addiction) might flavor a learner's mental models of other feedback loops?

Distributed cognition:

Although the human mind is capable of amazing feats, the world presents many problems that are bigger than one human can solve alone. In response, humans have evolved the capability to distribute cognition across multiple individuals and artifacts (Hutchins, 2005).

Distributing cognition across multiple minds: Distributing cognitive load across multiple human minds may be the easiest form to recognize, because we have a normal English word for it: "collaboration." Collaborators bring different cognitive resources to the table, such as different experiences, different knowledge, different skill sets, or access to different perceptions. One form of powerful alliance can be with people who have deeper knowledge of different parts of the system: for example, people who know about the human part of a system and people who know about the technology-enabled or ecological part. Another powerful form of alliance can be among people who have lived experience with different instances or cases of broadly similar phenomena. A feedback loop model that successfully represents multiple instances will be more generalizable and useful than one that only accounts for one idiosyncratic set of circumstances. Interestingly, collaboration itself can be amplified by reinforcing feedback loops (Kastens & Manduca, 2017). An increase in the capacity of one collaborator enables an increase in the capacity of the group, which in turn enables a further increase in the capacity of the individual, and so on.

*Offloading cognitive load onto a visual representation****:*** A "visualization" or "visual representation" is a type of model in which visible objects or symbols are used to represent aspects of a real or imagined system of interest. Like mental models, visualizations represent selected aspects of the target system in a simplified and useful way. But visual representations are out in the material world, where they can be shared, discussed, modified, critiqued, archived and otherwise manipulated by their creator or by others (Latour, 1986). Creating a causal loop diagram lets the reasoner take on a manageable chunk of cognition, think it through, and then pin it down in a concrete, stable configuration where it won't squirm away or evaporate while the reasoner goes onto the next bit of cognition. The reasoner can contemplate A-->B: is there a plausible mechanism for A to influence B? would it be a direct or inverse relationship? is there empirical evidence for such a relationship? Then A-->B can be placed into the visualization for safekeeping, clearing out cognitive resources for the reasoner to tackle B-->C. Externalizing a mental model into a visualization can improve on the mental model that was in the originating mind — especially if multiple minds are brought to bear, each bringing different skills and knowledge bases. Pointing to the causal loop diagram, one discussant can specify where he thinks there is a gap or a mistake in the causal logic. Another discussant can sketch in an additional exogenous influencer coming into the loop from outside. A third discussant can sketch her hypothesis of what will happen to the system under certain conditions as a behavior over time graph. Through a process known as "discourse over materials," meaning-making emerges in a complicated way through the interplay among materials, spoken words, and gesture.

Leveraging cognition embodied in software: This paper is about the claim that the human mind is capable of powerful feats of feedback loop thinking, alone or with low tech aids such as pencil and paper. However, we hasten to admit that the world is full of problems that call for collaborating with even more powerful cognitive aids, computational aids. Systems dynamics software embodies the cognition of people who had deep conceptual and mathematical knowledge of systems dynamics, along with hard-learned lessons from people who used such computational approaches to tackle a wide range of complex systems. By collaborating with such software, a person with deep knowledge of a system of concern can create quantitative simulations of the behavior of the systems they care about, even if they lack deep knowledge of systems dynamics. How the mind holds up its end of this collaboration is beyond the scope of this paper, but is surely a fascinating and researchable topic.

Conclusions: Ideas for future research

The entire realm of how humans think and learn with and about feedback loops is ripe for further cognitive science, learning science, and neuroscience research.

Recognizing that one cannot improve what one cannot measure, the authors and collaborators have developed an instrument to quantify learners' ability to distinguish short narratives that contain feedback loops from similar narratives without a closed loop. As presented at the 2022 SDS conference (Kastens & Shipley, 2022), we use signal detection analysis to quantify both participants' ability to distinguish loops from non-loops (d') and their bias (c) towards classifying narratives as either loops or non-loops. We are in the early stage of using this instrument to test the effectiveness of feedback loop instruction in four college level courses spanning environmental science, psychology, neuroscience, and gender studies. Our instructional approaches are designed to leverage the power of analogical reasoning and of visual representation. We welcome inquiries from other researchers who might want to try this loop detection instrument.

Here are some other potential research questions triggered by this review.

- How do cognitive biases manifest in the context of feedback loops? For example,
    - Does the tendency documented by Kuhn's group — to weight a plausible mechanism over observational evidence when making judgements about causality — hold true when novices and/or experts are building A-->B links for conceptual systems models?
    - Do the tendencies identified by Byre's group — to think of certain types of causal influences (e.g. things over which they have agency) while overlooking other types — come into play when individuals or groups are brainstorming about systems?
- Does a learner's understanding of feedback loops depend on the context in which they had their first exposure to the concept? Our preliminary data hint that students who learned about feedback loops in anatomy/physiology classes have more difficulty recognizing positive loops than other students in our participant pool, and we hypothesize that that might be because they learned about feedback loops in a disciplinary context where most feedback loops are the balancing loops of homeostasis.
- Can we develop evidence-informed learning progressions for teaching about feedback loops, acknowledging that the starting point will be in different disciplinary contexts for different learners?
- What is the extended learning progression from complete novice to experienced expert? Before that can be tackled, what are the dimensions along which the feedback loop understanding of novices versus experts needs to be characterized?

- How does a learner's understanding of structure, function and behavior of a feedback loop system coalesce from knowledge-in-pieces into an integrated whole?
- How does lived experience as a participant in a feedback loop system (e.g. addiction, viral marketing) influence one's understanding of feedback loops?
- More generally, how do people make the analogical mapping between systems in the external world and their conceptual models of those systems?
- How are feedback loops encoded in the brain? And does this change with increasing experience and expertise?

References cited

Burmester, A. (2017). Working memory: How you keep things "in mind" over the short term. The Conversation. Retrieved from https://theconversation.com/working-memory-how-you-keep-things-in-mind-over-the-short-term-75960.

Byrne, R. M. J. (2007). Precis of The Rational Imagination: How People Create Alternatives to Reality. *Behavioral and Brain Sciences,* 30, 439-480.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educ. Psychol. Rev.,* 26, 197-223.

DiSessa, A. A. (2018). Chapter 5: A friendly introduction to "knowledge in pieces": Modeling types of knowledge and their roles in learning. In G. Kaiser, H. Forgasz, M. Graven, E. Simmt, & B. Xu (Eds.), *Invited lectures from the 13th International Congress on Mathematical Education*. ICME-qe Monographs: Springer, Cham.

Doyle, A. C. (1894). The Adventure of Silver Blaze, *The Memoirs of Sherlock Holmes*, George Newnes.

Dumper, K., W. Jenkins, A Lacombe, M. Lovett, and M. Perlmutter (n.d.). 8.2 Part of the Brain Involved in Memory, PressBooks, at: https://opentext.wsu.edu/psych105/chapter/8-3-parts-of-the-brain-involved-in-memory/.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5), 752-775.

Gentner, D., & Smith, L. (2012). Analogical Reasoning. In *Encyclopedia of Human Behavior (Second Edition)* (pp. 130-136): Elsevier. Online at: https://groups.psych.northwestern.edu/gentner/papers/gentnerSmith_2012.pdf.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1-38.

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. V. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition*. doi:http://dx.doi.org/10.1016/j.cognition.2016.06.011.

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. V. (2019). Does expertise moderate the seductive allure of reductive explanation. *Acta Psychologia, 198*(July 2019).

Hutchins, E. (2001). Cognition, Distributed. In *International Encyclopedia of the Social and Behavioral Sciences* (pp. 2068-2072): Elsevier Science, Ltd.doi:https://doi.org/10.1016/j.actpsy.2019.102890.

Kahneman, Daniel, April 3, 2020, *New Yorker Radio Hour pod cast*
https://www.newyorker.com/podcast/political-scene/why-we-underestimated-covid-19

Kastens, K. A., S. Agrawal, L. S. Liben (2009). How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure, *International Journal of Science Education, special issue on Visual & spatial modes of learning,* J. Ramadas & J. Gilbert (editors), v. 31(3), pp. 365-393.

Kastens, K. A., & Ishikawa, T. (2006). Spatial Thinking in the Geosciences and Cognitive Sciences. In C. Manduca & D. Mogk (Eds.), *Earth and Mind: How Geoscientists Think and Learn about the Complex Earth* (pp. 53-76): Geological Society of America Special Paper 413.

Kastens, K. A., & Manduca, C. A. (2017). Leveraging the power of community of practice to improve teaching and learning about the Earth. *Change: The magazine of higher learning, 49*(5), 14-22.

Kastens, K. A., & Manduca, C. A. (Eds.). (2012). *Earth & Mind II: Synthesis of Research on Thinking & Leaning in the Geosciences*. Boulder, CO: Geological Society of America Special Paper 486, 210 p.

Kastens, K. A., and Shipley, T. F. (2022), An instrument to quantify how well students can recognize feedback loops in narratives and its use for evaluating pedagogical strategies, *Systems Dynamics Society Conference* Work-in-Progress.

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review, 102*(3), 490-502.

Kessel, A. M. (2008). Cognitive methods for information visualization: Linear and cyclic events. (PhD), Stanford University, Palo Alto, CA.

Kimmerer, R. W. (2015). *Braiding Sweetgrass: Indigenous wisdom, scientific knowledge, and the teachings of plants.* Minneapolis, MN: Milkweed Editions.

Kuhn, D. (2001). How do people know? *Psychological Science, 12(1),* 1-8.

Kuhn, D. (2004). Developing Reason. *Thinking and Reasoning*, 10(2), 197-219.

Kuhn, D. (2007). Jumping to conclusions. *Scientific American Mind, Feb/March*, 44-51.

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of Childhood Cognitive Development* (Blackwell) (pp. 497-523): Blackwell.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition & Instructions, 18*(4), 495-523.

Kurtz, K., Miao, C.-H. M., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*(4), 417-446. Gentner, D., & Colhoun, J. (2010). Analogical processes in human thinking and learning. In B. Glatzeder, V. Goel, & A. v. Muller (Eds.), On Thinking: Volume 2. Towards a Theory of Thinking (pp. 35-48). Berlin: Springer-Verlag.

Latour, B. (1986). Visualization and cognition: Drawing things together. In H. Kuklick (Ed.), *Knowledge and Society Studies in the Sociology of Culture Past and Present* (Vol. 6): Jai Press, available online at: http://www.bruno-latour.fr/sites/default/files/21-DRAWING-THINGS-TOGETHER-GB.pdf.

Lexington (2023) The great mystery of American politics: Why is the country so evenly divided? What might change that? The Economist, Jan 5, https://www.economist.com/united-states/2023/01/05/the-great-mystery-of-american-politics.

Noel, A. M., & Tversky, B. (2020 in review). Linear and circular thinking about time. *Journal of Cognitive Neuroscience, special issue.*

Nosofsky, R. M., Sanders, C. A., Zhu, X., & McDaniel, M. A. (2019). Model-guided search for optimal natural-science-category training exemplars: A work in progress. *Psychonomic Bulletin & Review, 26*, 48-76.

Papa, F. J., Oglesby, M. W., Aldrich, D. G., Schaller, F., & Cipher, D. (2007). Improving diagnostic capabilities of medical students via application of cognitive sciences-derived learning principles. *Medical Education, 41*, 419-425.

Resnick, I., Kastens, K., & Shipley, T. F. (2018). How students reason about visualizations from large professionally collected data sets: An eye-tracking study of students approaching the threshold of data proficiency. *Journal of Geoscience Education,* 66, 55-76.

Spellman, B., & Mandel, D. R. (2005). Causal Reasoning, Psychology of. In *Encyclopedia of Cognitive Science*: John Wiley & Sons.

Shipley, T., Tikoff, B., Manduca, C., Ormand, C. J. (2013). Structural Geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology*, 54 (August), 72-84

Stillings, N. (2012). Complex systems in the geosciences and in geoscience learning. In K. A. Kastens & C. Manduca (Eds.), *Earth & Mind II: Synthesis of Research on Thinking and Learning in the Geosciences, Geological Society of America Special Publication 486* (pp. 97-112). Boulder, CO: Geological Society of America.

Tversky, B. (2019). *Mind in motion: How action shapes thought*: Basic Books

Tversky, B., & Jamalian, A. (2019). *Gestures alter thinking about time*. Paper presented at the Cognitive Science. https://www.researchgate.net/publication/255961613_Gestures_Alter_Thinking_About_Time

Wahlheim, C. N., Finn, B., & Jacoby, L. L. (2012). Metacognitive judgments of repetition and variability effects in natural concept learning: evidence for variability neglect. *Memory and Cognition, 40*, 703-716.

Weisberg, D. S., Hopkins, E. J., & Taylor,J. C. V. (2018). People's explanatory preferences for scientific phenomena. *Cognitive Research: Principles and Implications*. doi:10.1186/s41235-018-0135-2.