

## The Plausibility Gap: A Model of Sensemaking

Gary Klein  
MacroCognition, LLC

Mohammadreza Jalaean  
MacroCognition, LLC

Robert R. Hoffman  
Institute for Human and Machine Cognition

Shane T. Mueller  
Michigan Technological University

This material is approved for public release. Distribution is unlimited. This material is based on research sponsored by the Air Force Research Lab (AFRL) under agreement number FA8650-17-2-7711. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

### Cite as:

Klein, G. Jalaean, M., Hoffman, R.R., and Mueller, S.T. (2021) and Hoffman, R.R. (2021). "The Plausibility Gap: A Model of Sensemaking." Technical Report, DARPA Explainable AI Program.



---

### Abstract

When people make plausibility judgments about an assertion, an event, or a piece of evidence, they are gauging whether it makes sense. Therefore, we can treat plausibility judgments as sensemaking activities. In this paper, we review the research literature, presenting the different ways that plausibility has been defined and measured. Then we describe the research program that allowed us to formulate our sensemaking perspective on plausibility. The model is based on an analysis of 23 cases, most of which involved understanding and interacting with information technology. The resulting model describes the user's attempts to construct a narrative as a state transition string, relying on plausibility judgments.

## Introduction

Our thesis is that when people make plausibility judgments about an assertion, an event, or a piece of evidence, they are gauging whether it makes sense, and therefore we can treat plausibility judgments as sensemaking activities. This thesis seems straightforward, at least to us, but to the best of our knowledge it has not appeared previously in the research literature on plausibility.

Plausibility judgments are important because they come into play in a variety of ways such as comprehension, problem solving, and anomaly detection. Connell and Keane (2004) stated that, “Plausibility has the hallmarks of a phenomenon so pervasive and important that, like air, no one notices it. Time and again, many cognitive accounts appeal to the idea of plausibility without specifying its cognitive basis” (p. 185).

We have been guilty of this omission ourselves. In our own work, we have invoked the notion of plausibility judgments without unpacking what that meant. Hoffman, Klein and Miller (2011) reviewed the literature on the criteria for what counts as a cause for an effect and identified three factors: co-variance (the putative cause comes before the effect), mutability (the putative cause was theoretically reversible), and propensity (the putative cause had the potential to bring about the effect). Hoffman et al. relabeled “propensity” as “plausibility” because “propensity” puts the focus on the putative cause whereas “plausibility” puts the focus on the person. The first two, co-variation and reversibility, are fairly easy to determine. The third criterion, plausibility, was left shrouded in mystery.

For another example, consider the Data/Frame model of sensemaking (Klein, Moon and Hoffman, 2006), which posits that the sensemaking process is typically initiated when someone questions an existing frame. This questioning involves tracking anomalies, detecting inconsistencies, judging plausibility, and gauging data quality. And again, we never tried to examine what happens when people judge plausibility.

Now that we have become sensitized to plausibility judgments, we are seeing them more frequently. For example, observations of hospital Emergency Departments (Hose et al., 2018) include examples in which the attending physician will receive some new item of information and will remark, “That doesn’t make sense,” before re-examining the patient. The “That doesn’t make sense” reaction indicates that something happened that was not plausible. Similarly, panel operators in petrochemical plants will comment, “That wasn’t supposed to happen,” when they see a sensor reading that seems anomalous.

## Literature Review

Because of our newfound interest in the nature of plausibility judgments, in this study, we conducted an extensive literature review in order to learn more about how plausibility judgments are made. Articles were systematically retrieved and screened in three stages. Academic Search Complete, Annual Review of Psychology, and Google Scholar database were searched using Boolean phrases such as Plausibility Judgement OR Plausibility Reasoning OR Plausibility Gap. Search settings included expanders such as application of related words and equivalent subjects. Stage 1 yielded 1042 articles for screening. In stage 2, a total of 28 primary, peer-reviewed articles

were selected by looking at the titles and reading the abstracts for further investigation. We ultimately excluded plausibility judgement in the domains of physical science, computer science, and engineering because our primary goal was to look at the plausibility judgment in the domain of social science, specifically psychology. After reading all 28 articles, 16 core articles were selected to be included in this literature review, as listed in Table 1. These fell into a few categories.

**Table 1. Categories of studies on plausibility judgement.**

<b>Categories</b>	<b>Articles</b>
Philosophical Perspective	Leake (1995)
Logical Perspective	Isberner and Isberner (2016); Lombardi, Nussbaum, and Sinatra, (2016a); Lombardi, Danielson, and Young (2016b)
Computational Perspective	Collins (1978); Collins and Michalski (1981)
Educational Perspective: Fake News	Abendroth and Richter (2020); Connell and Keane (2006); Matsuki et al. (2011); Sinatra and Lombardi (2020)
Educational Perspective: Conceptual Change	Lombardi, Nussbaum, and Sinatra (2016a); Nahari, Glicksohn, and Nachson (2010)
Linguistics Perspective	Connell, 2004; Connell and Keane 2002, 2003a, 2003b, 2004

### **Philosophical perspective**

Leake (1995) distinguished several types of plausibility criteria: a minimality criteria consistent with Occam's razor, a proof-based criterion, and criteria that are based on costs and probabilities.

### **Logical Perspective**

Isberner and Isberner (2016) were interested in developing a formal approach to plausibility from the field of knowledge representation and conducted research that manipulated plausibility by presenting verbal material that were logical or that violated logic. They measured the delays in response time as a function of the match or mismatch between the background world knowledge and the verbal statements. Rescher (1976, in Lombardi, Nussbaum, Sinatra, 2016a) defined plausibility as a function of the trustworthiness and quality of the source material. Connell and Keane asserted that a concept or scenario is plausible if it is conceptually consistent with other information. Lombardi, Danielson, Young (2016b) add information is judged as more plausible if it is less complex and requires less conjecture.

### **Computational perspective**

Collins (1978) envisioned a computational approach that treated plausibility as a function of the certainty of the information and the certainty of the inferences. Collins and Michalski (1981) contrasted different types of possible inferences: argument-based, reference-based, descriptor-based, and term-based.

### **Educational Perspective: Fake News**

The issue of plausibility judgments taps into several themes of interest to educational researchers. Abendroth and Richter (2020) presented an information processing account of plausibility. They saw the importance of plausibility judgments in identifying fake news — information on social media that is accidentally or, most commonly, deliberately false. Matsuki et al. (2011) defined plausibility as the acceptability or likelihood of a situation or a sentence describing it.” Connell and Keane (2006) stated it differently: “The degree of fit between a given scenario and prior knowledge.” “What is perceived to be potentially truthful when evaluating explanations,” (Lombardi et al., 2016a). For Abendroth and Richter, plausibility judgments are essential for identifying fake news by relying on individuals’ subjective perception of potential truthfulness of statements. How well does an item of information conform with a reader’s prior knowledge, beliefs, or current understanding of a situation? Sinatra and Lombardi (2020) argued that people detect fake news by assessing the credibility of the source and appraising lines of evidence, along with comparisons to alternatives and probabilistic reasoning.

### **Educational Perspective: Conceptual Change**

Lombardi, Nussbaum and Sinatra (2016a) had a different agenda — facilitating the conceptual change of students, particularly on topics such as climate change and evolution for which large numbers of students appeared to hold views that are discrepant with those of most scientists. Lombardi et al. wanted to get students to change their views. They defined plausibility a judging of the potential truthfulness of statements and concepts. It is related to concepts such as probability, coherence, comprehensibility, credibility, and believability, but it is distinct from each of these. Lombardi et al. drew on the work of Kahneman and Klein (2009) in claiming that plausibility judgments can be automatic (System 1) as well as analytical (System 2). Further, analogies can be useful in gauging plausibility — along this line, Nahari, Glicksohn, and Nachson (2010) suggested that instances are judged more plausible if it is easier to recall similar instances, essentially the availability heuristic identified by Tversky and Kahneman 1974). Lombardi et al. (2016a) suggested that another part of the plausibility equation is having a plausible physical mechanism to account for an effect — essentially, a cause. However, most of their concern is with increasing the credibility of scientists and teachers and increasing the credibility of source validity judgments to promote conceptual change. They also raise the topic of individual differences in need for cognition and openness to conceptual change.

### **Linguistics Perspective**

The issue of plausibility in the linguistics domain was addressed as a unique cognitive process and separate from other central issues in cognitive psychology (Connell, 2004; Connell and Keane 2002, 2003a, 2003b, 2004). To address plausibility independently, they proposed the Knowledge-Fitting Theory (Connell and Keane, 2003a). The knowledge-fitting theory identifies two stages of the plausibility judgement process including a comprehension stage (understanding the scenario) and an assessment stage (examining scenario fit to prior knowledge). To make a plausible judgement, people try to create a mental link between what the scenario describes and the previous knowledge they have about the scenario. As such, the core of the KFT is the strength of relationship between the scenario and prior knowledge.

Connell (2004) found that the way a particular scenario is described influences the amount of prior knowledge activation. If a little background knowledge is activated, it would be difficult to understand the scenario and if a lot of background knowledge is activated, then the comprehension of the scenario would be easy. Thus, two factors affect the plausibility judgement of a scenario: the word-coherence of the description and the concept-coherence of the scenario's elements and events (Connell and Keane, 2004). The word-coherence (scenario word choice) factor, according to Connell and Keane (2004) is associated with the comprehension stage and more specifically the amount of time one would need to make a plausibility judgement. However, the concept-coherence (simple or complex scenarios) factor can be influential in both stages of plausibility judgment including comprehension (time) and assessment (level of plausibility a scenario). Word-coherence eases the understanding of a scenario by activating prior knowledge and Concept-coherence eases the understanding and determines assessment accuracy by the amount of prior knowledge activation.

All of these perspectives are interesting and useful, and we have no disagreement with any of the plausibility criteria suggested by these researchers — logical consistency, credibility of sources, consistency with other information, reduced complexity, perception of truthfulness, alignment with prior beliefs, knowledge and understanding, probability, coherence, comprehensibility, ease of recalling similar instances, physical mechanism, and so forth.

However, we did not see a Naturalistic Decision Making perspective represented in this work, and specifically, we did not see a reflection of research and modeling that involved sensemaking. In all of the papers we have reviewed, only one mentioned causal inference — Lombardi et al. (2016b) raised the issue of physical mechanisms, which seemed to imply causality — but that was only a brief and passing mention. Two other papers (Connell, 2006; Connell and Keane, 2004) alluded to causal factors in their discussion of prior knowledge as they studied the way subjects made plausibility judgments of textual material — sentence pairs. However, we did not see any investigation of plausibility judgments in naturalistic contexts such as making diagnoses in hospitals or petrochemical plants or military operations.

Therefore, we saw an opportunity to add to the existing research and models by conducting a research project examining textual accounts of explanations found in books, magazines, newspapers, and social media — cases in which people created written documents in an attempt to explain events and systems to readers.

### **Understanding the Process of Forming Explanations**

The overarching research project for this work was an investigation of the process of forming explanations — the ways that people react when they encounter unexpected, anomalous, and surprising items of information and how they try to diagnose what happened.

For that larger-scale effort we collected a corpus of 73 cases that allowed us to construct models of local and global explaining. Local explaining involved efforts to understand specific events and global explaining involved efforts to understand how machines, systems and organizations

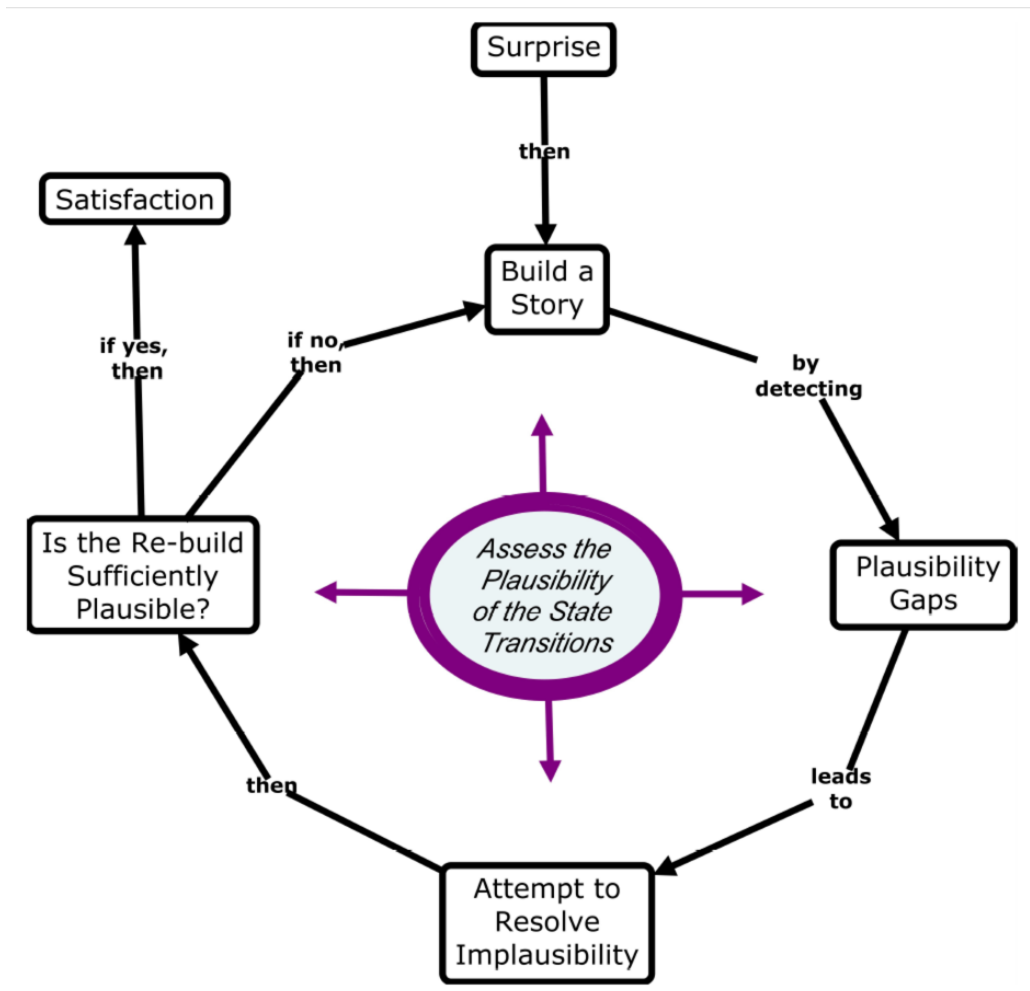
functioned. The 73 cases were a convenience sample. We did not formulate criteria in advance because we wanted to cast a wide net that did not systematically exclude any types of incidents.

For this current project on plausibility, we identified a subset of 23 cases of explanations. The criteria for this subset included having richer details about the explainer or learner's reasoning; having more of a focus on humans interacting with machines or information technology and coming from sources we judged to be more reliable than our initial sample, which included examples from Reddit or from personal accounts. The subset of 23 cases is listed in Appendix A, and includes the Air France 447 crash, the USS Vincennes shootdown, the workings of AlphaGo, and the grounding of the cruise ship *Royal Majesty*.

The method we used was to review the details of each of these selected cases. The reviewer (the senior author) tried to imagine himself as the protagonist in the incident who was seeking to gain an understanding, using the information available at the time. The reviewer then examined the reports of each of the cases and relying on induction and abduction synthesized these accounts to formulate a general model of the sensemaking involved and how it depended on plausibility judgments.

### **Key Finding: The "Plausibility Gap" Model**

What people try to do when explaining an anomaly or surprise is to construct a story about how something came to pass — for example, a story about how a device works. But this notion that we are building a story is banal. What matters the most is the process of story-building that people go through. In building a story, the reasoner works out a causal sequence, going from the beginning state (State-0) to the end state (State-N). In addition, a good story will also contain an insight—the resolution of a surprise (see Figure 1, below).



**Figure 1. The Process of Story-Building.**

If we are going to the trouble of building a story then there must be a reason—something we don't understand, or something that surprised us. So typically, there is an insight lurking within our story-building effort.

Figure 1 presents the model as a cycle (a closed loop), beginning with the trigger that initiates the story-building process. As stated, this is usually a surprise or a desire to overcome ignorance. The loop closes when the plausibility assessment is judged to be sufficient, and the insights gained resolve the triggering conditions. If learners determine that the story isn't adequately plausible, they try to build a different story and start the process again. Figure 2, an elaborated version of Figure 1, emphasizes two functions in story building: *Identification of leverage points* for elaborating a story and *Plausibility judgments* for assessing that story. They essentially are the forces driving the explanatory process.

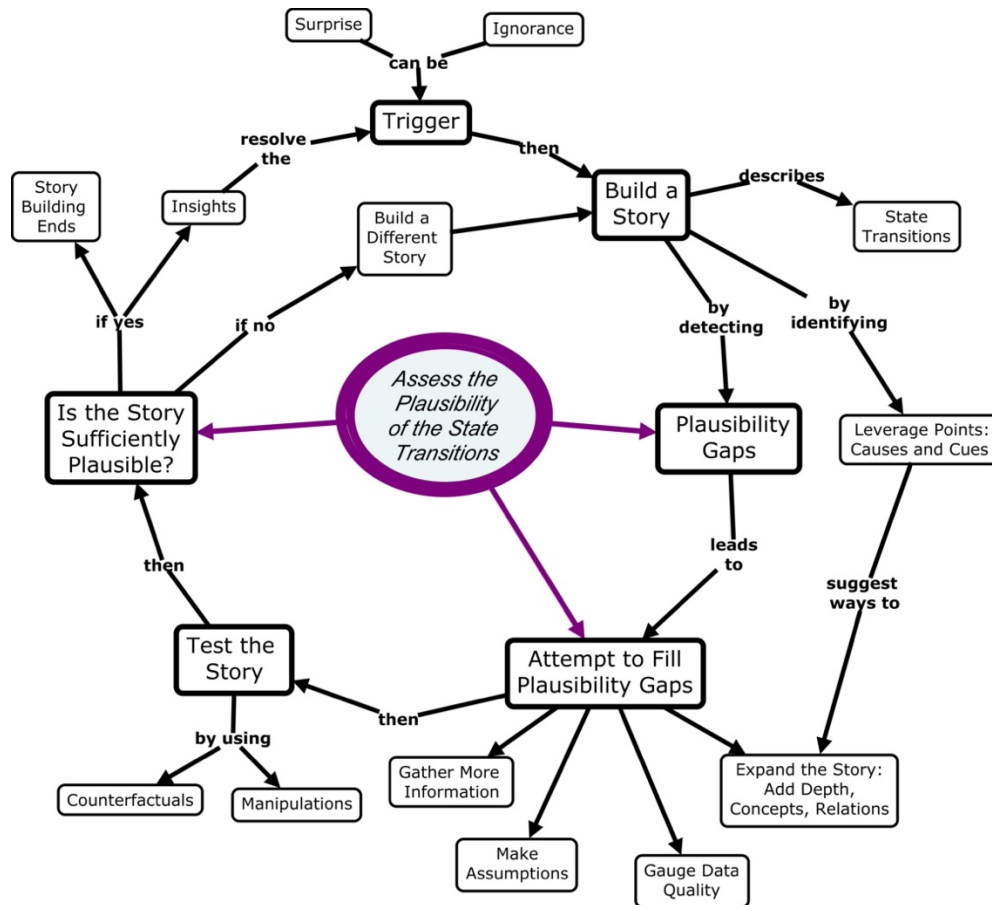


Figure 2. The "Plausibility Gap " Model.

### Plausibility Judgements

One of the primary forces that guides story-building is plausibility. For each state transition, whether from State-0 to State-N or, usually, the intervening transitions, the reasoner gauges the plausibility of that transition. Our notion of plausibility is psychological, not logical, epistemic, or philosophical. Plausibility judgments entail plausibility gaps that need to be filled in. These plausibility gaps stimulate curiosity. The learner searches for additional information that will fill in the gaps.

The learner also fills the gaps by making assumptions—a key part of the explaining process. If the plausibility gaps can't be filled to the learner's satisfaction, confidence in the story is reduced or even lost. The learner may shift into a “snap-back” mode, as described by Cohen, Freeman and Thompson (1997). During an incident, if our initial diagnosis is wrong, the anomalies will persist and increase. So, the effort of explaining away the anomalies and filling the gaps becomes greater and greater, either because there are more and more anomalies to explain away, or because the effort at explaining away is just too great. At that point, we become skeptical of our story and we start searching for an alternative story, or we become more receptive to an alternative story that we had been dismissing. Conversely, when it is easy to build the mental simulation of how the state transitions work, we tend to judge those transitions as plausible. This is discussed in the



judgment and decision-making literature as the simulation heuristic (Kahneman and Tversky, 1984).

To make plausibility judgments for the transitions in a story, we (the researchers) think a learner engages in mental simulation, imagining each state transition as a response to the causes of which the learner is aware. If a state transition flows smoothly, then there is no problem. On the other hand, if the known causes do not account for the transition, then the person experiences this mismatch as a strain, a cognitive strain (Cohen, Freeman and Thompson, 1997). If the strain becomes too great, some sort of mental “tilt” arises, and the learner loses faith in the story. Figure 2 shows how the learner might try to reduce the cognitive strain by gathering more information, assessing the quality of the anomalous data (and hoping to find a reason to disregard the anomalous data), making assumptions, or by expanding the story. However, at its core, this psychological judgment of plausibility hinges on the Learner’s imagination. It is a psychological judgment, not a logical one. Even in imagining a physical process, say a panel operator for a petrochemical plant, the operator will be imagining, “if I were a molecule of ethane in this splitter unit of a petrochemical plant, and I was exposed to this level of heat, how much pressure would I be feeling? Will my passage up the distillation tower be expedited? And if so, how quickly will this reaction take, once a set of control actions are taken?”

### **Leverage Point Identification**

To identify leverage points for building a story a person has to draw on knowledge of the types of causes for events such as those that triggered the explanation process. This causal set is activated just-in-time in response to surprise or to ignorance. In the course of self-explaining, the causal set will be expanded and deepened, and the Learner’s mental model will become richer — the Learner’s overall causal repertoire will be expanded. The leverage points a person identifies (the causes the Learner considers and the cues the Learner notices) will depend on the sophistication of the Learner’s mental model and the kinds of stories the Learner has considered in the past. So, the Learner’s stories determine the leverage points that are identified, and the leverage points the Learner identifies will activate a set of causes and make certain cues more salient. Within the Plausibility Gap model, some leverage points for constructing a story may be provided to a Learner and others will be discovered. These leverage points can be seen as affordances inasmuch as they shape the gist of the story.

These functions of identifying leverage points and judging plausibility have received little attention in the literature, so one of the contributions of the Plausibility Gap Model is to highlight them. The model also describes a stopping rule for ceasing to expand and deepen a story—when the plausibility is experienced as sufficient.

### **The Process of Deepening**

When do learners deepen their story? Note that the causes in a story are their mental model. Rozenblit and Keil (2002) discussed the illusion of explanatory depth, suggesting that people do not deepen enough—that people are satisfied with a shallow understanding. But surely people cannot deepen all the way down because there is no “all the way down.” So, Learners stop deepening when they are satisfied with the plausibility of the story.

Story building continues as Learners introduce leverage points and clues into the causal chain and network until they have an account that satisfies their sense of plausibility: “Yes, this makes sense, it could easily have happened this way.” One of the 23 accounts is the analysis presented by Malcolm Gladwell (2014) about David Koresh and the Waco Texas tragedy. The behavior of the Branch Davidians seemed completely irrational at the start, but by the end of Gladwell’s account, their behavior made a lot of sense. It was plausible.

## Causal Repertoires

What counts as causes — components in these stories and other types of accounts? We generally know the types of things that come to mind as potential causes. These include events, decisions, forces, missing data, erroneous data, and flawed beliefs. But these factors are too general. For the analysis of the 23 cases of story building, we found that we were attending to potential causes even as we were constructing the story. The Plausibility Gap model assumes that people have causal repertoires—Learners’ mental models include a capability for generating potential causes for a given outcome, and different people would have different but overlapping causal repertoires of the types of things they would consider in building a story. The question may arise that what does the Plausibility Gap model look like when applied to actual instances? We use an example, the Air France Flight 447 accident in 2009 to answer this question.

The Airbus 330 of France airlines took off from Rio de Janeiro in Brazil, headed to Paris, carrying 228 passengers and crew members. Several hours into the flight, ice crystals formed on the pitot tubes, preventing the airplane from determining its speed. As a result, the autopilot turned off. The airplane was using the latest in intelligent technology. The manufacturers had led aircrews to believe that it was impossible to stall the plane. The airplane was just too smart and would not allow pilots to engage in unsafe actions that might result in a stall. And that was true as long as the sensors were working. However, with the autopilot disengaged, all bets were off. Now the airplane could enter into a stall. Unfortunately, the pilot flying apparently did not know this (or had never been told it).<sup>1</sup> So he continued to climb steeply, feeling a false sense of invulnerability. The plane was in fact climbing so steeply, and its airspeed was so reduced, that it was on a trajectory to stall.

At some point, the pitot tubes seem to have unfrozen, even though the autopilot didn’t come back on. Now the airplane did sense the airspeed and did identify the near-stall condition. As a result, a stall warning came on. This auditory warning confused the pilot flying who thought the airplane was installable. He continued climbing. And then the stall warning went off. The stall warning ceased due to the slow speed (presumably because you don’t want the stall warning going off while the airplane is taxiing on the ground). The pilot flying must have felt relieved that the stall warning went off and took this as a good sign instead of a very ominous sign.

A more experienced pilot realized the flight configuration was extremely dangerous. He seems to have taken over the controls and he put the nose down to increase the speed. As a result — the stall warnings came back on. This happened because the airspeed had increased over the minimum. Now the pilots were thoroughly befuddled. Putting the nose down (to escape the stall conditions)

---

<sup>1</sup> This interpretation of the pilot’s decision making is based on conversations with Nadine Sarter (personal communication). Other accounts have been offered, such as the pilot’s inadequate training (Palmer, 2013).

was getting them yelled at by the system but continuing to climb was absurd. As they tried to sort out what was happening, the airplane did stall, and it dropped into the ocean. It was several years before the airplane was located and the flight data recorder could be recovered.

No one had imagined that a jetliner could be flying so slowly. As a result, the stall warning, intended to help the pilots avert danger, actually helped to kill them.

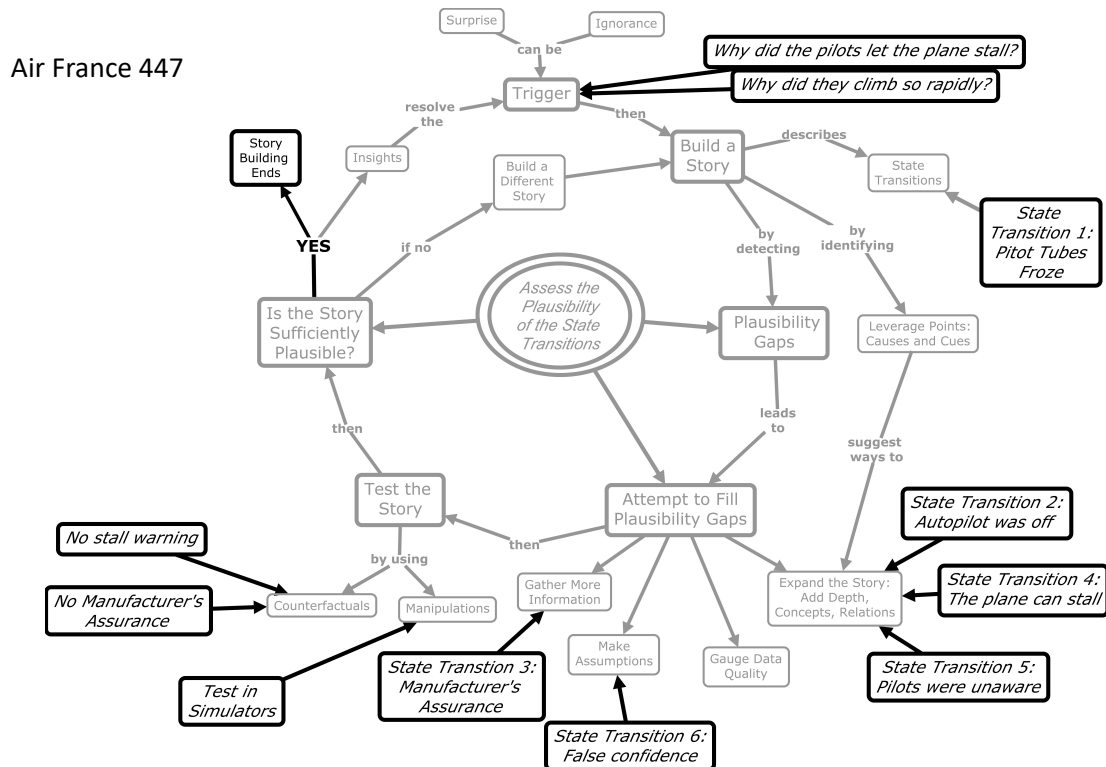


Figure 3. Applying Plausibility Gap Model Judgment in Air France Flight 447 Example.

### Relation of the Plausibility Gap Model to Other Models

The concept of plausibility judgments emerged and deepened for us as we realized that plausibility judgments were so central to the process of explaining, and as we discovered that existing accounts of plausibility judgments were so incomplete: they emphasized logical analyses, scrutiny of the credibility of sources, and so forth and failed to consider plausibility judgments as a form of sensemaking.

Our thesis is that there may be value in taking a sensemaking perspective on plausibility judgments. Gauging that something is plausible or implausible is, at some level, assessing whether or not it makes sense to the individual, and even to the team.

To make plausibility judgments for the transitions in a story, we (the researchers) think a Learner engages in mental simulation, imagining each state transition as a response to the causes of which

the Learner is aware. If a state transition flows smoothly, then there is no problem. On the other hand, if the known causes do not easily account for the transition, then the person experiences this mismatch as a strain, a cognitive strain (Cohen, Freeman, and Thompson, 1997). If the strain becomes too great, some sort of mental “tilt” arises, and the learner loses faith in the story. Figure 2 shows how the Learner might try to reduce the cognitive strain by gathering more information or by making assumptions, or by expanding the story. However, at its core, this psychological judgment of plausibility hinges on the learner’s imagination. It is a psychological judgment, not a logical one.

In our review of the literature, we found that previous accounts of plausibility judgment either ignored the role of prior causal knowledge or gave it little attention. But there was one exception: the Knowledge-Fitting Theory of Plausibility (Connell and Keane, 2004; Connell, 2006). Their work did highlight the role of prior knowledge, especially causal knowledge, in making plausibility judgments. They studied the assessment of textual materials, and how participants in their experiments made what they are told fit what they knew about the world. They were addressing what we refer to as sensemaking.

Our account differs from theirs in a few important ways. First, we were considering natural events, such as the crash of Air France 447, rather than textual stimuli such as “The bottle fell off the shelf. The bottle smashed” vs “The bottle fell off the shelf. The bottle melted.” Second, our account centers around story building and state transitions as opposed to overall plausibility judgments of statements. Third, and most critical, our account hinges on the psychological strain of imagining the state transitions whereas Connell and Keane calculated plausibility in terms of the number of scenario versions that could be recalled or constructed given the way the stimulus material was primed.

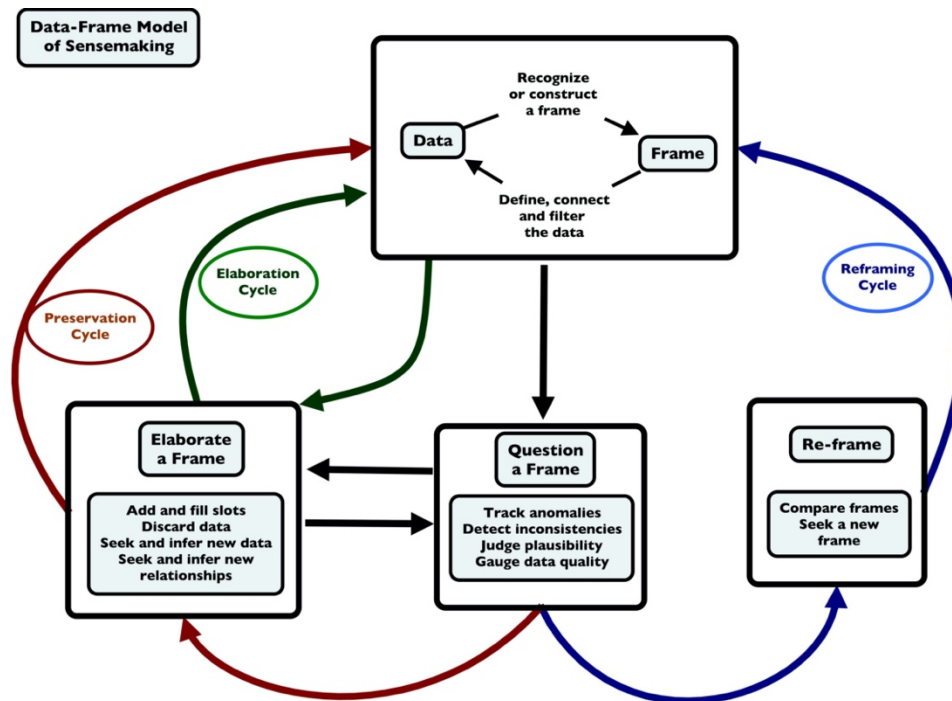
The number of scenario versions is a proxy for the cognitive strain of imagining a transition but seems fundamentally different. For example, one of our examples, the USS Vincennes shootdown of an Iranian airliner in 1988, depended on the Vincennes captain’s efforts to explain away data there were inconsistent with the two stories he was comparing: the object his crew had identified was a commercial airliner or it was a military airplane preparing to attack his ship. There was more cognitive strain to explain away the inconsistencies in the commercial airliner story, and so he rejected that story. This type of plausibility judgment is easy to handle by the Plausibility Gap account of Figures 1 and 2. In contrast, the Connell and Keane account is not really designed to describe judgments of this kind.

This description of plausibility judgments has a number of features:

- Focus on human cognition. This is different than a focus on logical analyses or source credibility.
- Story building process. Our approach views plausibility judgments as the attempt to construct a story, a narrative, to explain the phenomenon of interest.
- Initiating condition. Story-building is initiated when people encounter a surprise or anomaly that “doesn’t make sense.” The response is to build a story to resolve the anomaly.

- Anomaly detection is not simply noticing outliers, although that is how it is often treated. We detect anomalies when our expectancies are violated. Therefore, statistical analysis of outliers is a misleading and insufficient treatment of anomaly detection.
- Limitations of story-building. Stories generally have a sequential form, a chain of causes, which is often a necessary simplification but sometimes an over-simplification that misses explanations involving multiple intersecting causes.
- State transitions. The sequential structure of stories can be seen as state transitions, moving from one state to the next as new events and information are received and as the causal implications are worked out. These state transitions can be considered as a form of mental simulation (Klein and Crandall, 1995).
- Filling gaps. Typically, the state transitions will leave gaps — the causes present in one state do not neatly align with the following state.
- Cognitive strain. These gaps pose problems for assessing the story as plausible versus implausible. The cognitive strain makes it more difficult to accept a story as plausible.
- Some of the primary activities for filling the gap and resolving implausibility are to gather more information, to determine flaws in the data collection process that identified the gap, to make assumptions, and to elaborate the story (adding depth, additional concepts and relationships) in order to explain away the gap,
- Centrality of plausibility judgments to story building. Therefore, the plausibility judgments are embedded in the process of constructing stories to account for surprises and anomalies.
- Ease of imagining the state transitions. This ease of imagining, which is related to the availability heuristic identified by Tversky and Kahneman (1974) is a subjective judgment that the causal factors in the precursor state are sufficient to account for the subsequent state. This judgment depends heavily on the sophistication of the person's mental model of causes.
- Existence of a stopping point. The sensemaking process that utilizes plausibility judgments need not go on forever. The stopping point occurs when the plausibility gaps is reduced, and cognitive strain is minimized. This aspect of the Plausibility Gap model is consistent with the Klein, Hoffman and Mueller (2019) model of the process of explaining. Both of these models call out plausibility of transitions as a part of determining the stopping point.
- Linkage to decision making and sensemaking. Our account takes a different direction than previous research on topics such as conceptual change in education.

Figure 2 (above) illustrates our assertions of the way that plausibility judgments are made. As would be expected, Figure 2 is consistent with established models of sensemaking (Klein, Moon and Hoffman, 2006a, b). This includes the Data/Frame model illustrated in Figure 4 (below). But as we began this investigation, we did not assume the D/F model because we wanted to take a fresh look at plausibility judgments. Once we had iteratively formulated the Plausibility Gap model, we went back to compare it with the D/F model of sensemaking. The Plausibility Gap model includes all the key processes in the D/F model, including framing, questioning of the frame, and elaborating the frame. In essence, the Plausibility Gap model is a possible path through the D/F model, intended to highlight the process of self-explaining as story-building.



**Figure 4. The Data/Frame Model of sensemaking (adapted from Klein, et al., 2007).**

The Plausibility Gap account is particularly relevant to Re-Framing, and may be considered as an elaboration of the Re-Framing activity, specifically, “seeking a new frame.” The story-building activity in the Plausibility Gap is a primary means of seeking a new frame. In other words, the Plausibility Gap model can be thought of as one particular pathway through the Data/Frame sensemaking process.

The Plausibility Gap account also may be useful in clarifying aspects of self-explaining. The topic of Explainable AI has primarily examined ways to provide system users with explanations or at least explanatory materials. However, we suggest that users are typically taking the initiative and are attempting to take any available clues and materials in order to form their own local and global explanations — essentially to build their own stories. We suggest that the Plausibility Gap model might highlight some important tactics users are applying during this self-explaining process.

## Conclusion

The Plausibility Gap model has the potential to achieve several outcomes. It can add the concept of plausibility judgments as sensemaking to the treatment of plausibility in the research literature. It can also inform efforts that depend on plausibility judgments, such as making diagnoses in healthcare and in managing complex units in petrochemical plants.

The Plausibility Gap model may have implications for system design, particularly the design of advanced information technology involving automation, Artificial Intelligence and Machine Learning. Instead of treating these systems as opaque and inscrutable, designers may find it useful to offer system users transparency into the data used to train these systems so that the users can assess data quality in addressing plausibility gaps. Designers can make it easier for system users

to track cognitive strain by flagging the discrepancies that emerge and highlighting the plausibility gaps.

With regard to training, developers can seek to make plausibility judgments more accurate by helping people build stronger mental models. (See the Mental Model Matrix developed by Borders, Klein, and Besuijen, 2019). Developers can try to foster a mindset of curiosity regarding anomalies rather than dismissing anomalies as inconvenient. Developers can explore ways to encourage people to engage in counterfactual reasoning in identifying plausibility gaps and in assessing the quality of attempts to fill those gaps. Developers can seek methods to train perspective-taking skills to facilitate plausibility judgments involving other people. Developers can also try to help trainees gain a better appreciation of the limits of story-building in domains involving multiple intersecting causes.

### Acknowledgement and Disclaimer

This material is approved for public release. Distribution is unlimited. This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA) under agreement number FA8650-17-2-7711. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

### References

- Abendroth, J., and Richter, T. (2020). Mere plausibility enhances comprehension: The role of plausibility in comprehending an unfamiliar scientific debate. *Journal of Educational Psychology*.
- Borders, J., Klein, G., and Besuijen, R. (2019). An operational account of mental models: A pilot study. *International Conference on Naturalistic Decision Making*, San Francisco, CA.
- Cohen, M., Freeman, J.T., and Thompson, B. (1997). Training the naturalistic decision maker. In C. E. Zsombok and G. A. Klein (Eds.), *Naturalistic Decision Making* (pp. 257-68). Mahwah, NJ: Erlbaum.
- Collins, A. (1978). Fragments of a theory of human plausible reasoning. *American Journal of Computational Linguistics*, 13-20.
- Collins, A., and Michalski, R. S. (1981). Toward a Formal Theory of Human Plausible Reasoning.
- Connell, L. (2004). Making the implausible plausible. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 26, No. 26).
- Connell, L., and Keane, M. T. (2006). A model of plausibility. *Cognitive Science*, 30(1), 95-120.
- Connell, L., and Keane, M. T. (2003). PAM: A cognitive model of plausibility. In *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 25, No. 25).
- Connell, L., and Keane, M. T. (2003). The knowledge-fitting theory of plausibility. *Manuscript in preparation*.

- Connell, L., and Keane, M. T. (2002). The roots of plausibility: The role of coherence and distributional knowledge in plausibility judgements. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 24, No. 24).
- Connell, L., and Keane, M. T. (2004). What plausibly affects plausibility? Concept coherence and distributional word coherence as factors influencing plausibility judgments. *Memory and cognition*, 32(2), 185-197.
- Gladwell, M. (2014, March)). Sacred and profane: How not to negotiate with believers. *The New Yorker*. [<https://www.newyorker.com>].
- Hoffman, R.R., Klein, G., and Miller, J.E. (2011). Naturalistic investigations and models of reasoning about complex indeterminate causation. *Information and Knowledge Systems Management*, 10, 397-425.
- Hose, B. Z., Carayon, P., Hoonakker, P., Wooldridge, A., Brazelton, T., Dean, S., and Gurses, A. (2018, August). Challenges of disposition decision making for pediatric trauma patients in the emergency department. In *Congress of the International Ergonomics Association* (pp. 339-345). Springer, Cham.
- Isberner, M. B., and Kern-Isberner, G. (2016, March). A formal model of plausibility monitoring in language comprehension. In *The Twenty-Ninth International Flairs Conference*.
- Kahneman, D., and Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64, 515–526. <http://dx.doi.org/10.1037/a0016755>.
- Kahneman, D., and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39, 341-350.
- Klein, G., Moon, B., and Hoffman, R.R. (2006a, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-26.
- Klein, G., Moon, B. and Hoffman, R.R. (2006b, November/December). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, pp. 88-92.
- Klein, G., Phillips, J.K., Rall, E.L., and Peluso, D.A. (2007). A data-rame theory of sensemaking. In R.R.Hoffman (Ed.), *Expertise out of context* (pp. 113-155). New York: Lawrence Erlbaum Associates.
- Leake, D. B. (1995). Abduction, experience, and goals: A model of everyday abductive explanation. *Journal of Experimental and Theoretical Artificial Intelligence*, 7(4), 407-428.
- Lombardi, D., Danielson, R. W., and Young, N. (2016b). A plausible connection: Models examining the relations between evaluation, plausibility, and the refutation text effect. *Learning and Instruction*, 44, 74-86.
- Lombardi, D., Nussbaum, E. M., and Sinatra, G. M. (2016a). Plausibility judgments in conceptual change and epistemic cognition. *Educational Psychologist*, 51(1), 35-56.
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., and McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4), 913.
- Nahari, G., Glicksohn, J., and Nachson, I. (2010). Credibility judgments of narratives: Language, plausibility, and absorption. *American Journal of Psychology*, 123(3), 319-335.



- Palmer, Bill (2013). *Understanding Air France 447*. William Palmer, Jr.
- Rozenblit, L., and Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521-562.
- Rescher, N. (1976). *Plausible reasoning: An introduction to the theory and practice of plausibilistic inference*. Amsterdam, the Netherlands: Van Gorcum.
- Sinatra, G. M., and Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist*, 55(3), 120-131.
- Tversky, A., and Kahneman, D. (1974). Judgment and uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.

## **Appendix A**

### **Listing of the Cases**

These cases are ones discovered in literatures or the media that were particularly informative of the Plausibility Gap Model. Links to the sources can be provided on request.

#### **Cases Involving Machines or Information Technology**

1. Why did Watson give the answer "Toronto" in Jeopardy?
2. Why are there maggots in my dead refrigerator?
3. Why did Air France #447 crash?
4. How does AlphaGo work?
5. Distillation tower: why did operators miss the upset in Scenario 3?
6. Why did my GPS take me down an absurd route?
7. Why did CPT Rogers of Vincennes shoot down a commercial airliner in 1988?
8. Why did KAL 007 get shot down?
9. Why did the cruise ship Royal Majesty get grounded?
10. Why did the airplane crash after the pilot failed to arm the spoilers?
11. Why did the automatic blood pressure machine fool the surgical team?
12. Why do autopilots sometimes quit working with no warning?
13. In Desert Storm, why did our Patriot missile system shoot two friendly airplanes?
14. How do trains negotiate curved tracks?
15. How did the firefighter know to order his crew out of the burning building?
16. Why did the British naval officer order the shutdown of a new track?
17. What caused the mysterious outages on the Singapore subway system in July 2017?

#### **Cases Involving Other Complex States of Affairs, Events, or Circumstances**

18. Why did Walter Reed believe that mosquitoes were not involved in spreading Yellow Fever when he went to Cuba to study the disease?
19. Why did the police officer shoot the innocent African American shopper at the WalMart in Beavercreek, OH?
20. Why did the Department of Justice confrontation with Koresh end in disaster?

21. Why do Westerners and Arabs baffle each other in the way they think?
22. Why was the German Blitzkrieg successful against the French in WW2, even though France had a very strong military?
23. Why did the US Navy ship John S. McCain collide with another ship near Singapore in 2017?