Why Some Surprises are More Surprising than Others: Surprise as a Metacognitive Sense of Explanatory Difficulty

Meadhbh I. Foster[1] and Mark T. Keane[1]

[1]School of Computer Science and Informatics,

University College Dublin, Ireland

Running head: EXPLAINING SURPRISE

Correspondence concerning this article should be addressed to:

Meadhbh I. Foster

School of Computer Science and Informatics

University College Dublin

Belfield, Dublin 4

Ireland

E-mail: meadhbh.foster@ucdconnect.ie

Telephone: 00353 (0)1 7162907

Abstract

Early theories of surprise, including Darwin's, argued that it was predominantly a basic emotion. Recently, theories have taken a more cognitive view of surprise, casting it as a process of "making sense of surprising events". The current paper advances the view that the essence of this sense-making process is *explanation*; specifically, that people's perception of surprise is a metacognitive estimate of the cognitive work involved in explaining an abnormal event. So, some surprises are more surprising because they are harder to explain. This proposal is tested in eight experiments that explore how (i) the contents of memory can influence surprise, (ii) different classes of scenarios can retrieve more/less relevant knowledge from memory to explain surprising outcomes, (iii) how partial explanations constrain the explanation process reducing surprise, (iv) how, overall, any factor that acts to increase the cognitive work in explaining a surprising event, results in higher levels of surprise (e.g., task demands to find several rather than just one explanation). Across the present studies, using a many different materials, paradigms and measures, it is consistently and repeatedly found that the difficulty of explaining a surprising outcome is the best predictor for people's perceptions of the surprisingness of events. Alternative accounts of these results are considered, as are future directions for this research.

Keywords: surprise judgments, comprehension, explanation, difficulty

1. Introduction

Why are some surprises more surprising than others? Consider our surprise at the deaths of famous people: we were less surprised when we heard that Margaret Thatcher or Ronald Regan had died, but we were very surprised at the deaths of Michael Jackson or Lady Diana. This shock of surprise we experience is often accompanied by the question "How could that happen?" As such, we argue that surprise is a graded experience that depends fundamentally on explanation.

Surprise seems to necessarily be about explanation; an event is not surprising if it can be explained easily, and becomes more surprising the harder it is to explain. Normally, our comprehension of the world proceeds smoothly, establishing the coherence of events either predictively or retrospectively, making sense of what we encounter. However, when that coherence breaks down, a more-complex-than-usual comprehension step is required, which we call *explanation.* Thatcher's death was not surprising because we immediately understand that she was old and in declining health, whereas Jackson's death *was* surprising because it was so hard to explain. Michael Jackson was not that old, (certainly) not overweight and should have been fit from dancing; but had we known about his history of injury and self-medication, we clearly would have been less surprised by his demise.

Surprise is a graded experience; it is not all-or-nothing, we experience different degrees of surprise depending on the situation. For example, our perceptions of surprise differ by degrees for Thatcher's, Jackson's and Lady Diana's deaths (possibly in this increasing order). So, any adequate theory of surprise needs to account for how relative differences in experienced surprise arise.

In summary, the present theory argues that people's perception of surprise hinges on a metacognitive sense of the amount of cognitive work necessary to explain some target surprising event. In the remainder of the paper, we outline this *metacognitive explanation-*

based (MEB) theory of surprise, and report eight experiments that explore the factors impacting the cognitive work carried out in explaining surprising outcomes[1].

In the following sub-sections, we first sketch previous theories of surprise, before collating the evidence for the centrality of surprise in diverse cognitive phenomena (e.g., counterfactual thinking, hindsight bias, and learning). After briefly reviewing this literature, the current theory is outlined along with its key predictions, before reporting eight experiments testing these predictions. The paper concludes with a discussion of how these findings relate to different theoretical options and directions for future research.

1.1 A cognitive emotion: Theoretical perspectives on surprise

The phenomenon of surprise has been intensively researched since Darwin's time, perhaps because it involves an interesting mixture of emotion and cognition. Though surprise clearly involves an emotional reaction (often accompanied by a startle response), it also seems to serve a strategic, cognitive goal, as it directs attention to explain why the surprising event occurred and to learn for the future (e.g., Macedo, 2010; Maguire, Maguire & Keane, 2011; Ranganath & Rainer, 2003). Originally conceived of as a "basic emotion" (e.g., Darwin, 1872; Ekman & Friesen, 1971; Izard, 1977; Plutchik, 1991; Tomkins, 1962), more recently surprise has been re-appraised as a cognitive state because, unlike most emotions, it can be either positively or negatively valenced (Ortony & Turner, 1990; see also Kahneman & Miller, 1986; Maguire et al., 2011). Indeed, nowadays, cognitive aspects of surprise are routinely mentioned in the affective literature; for instance, Wilson and Gilbert (2008) explicitly discuss explanatory aspects of surprise in affective adaptation.

In general, cognitive theories of surprise divide into two identifiable camps, the "probability" and "sense-making" approaches. Probability theories focus on the properties of

[1] We use the term "surprising outcome" in this paper to refer to target surprising events because the terminology used in previous studies is too theory-laden; for instance, "unexpected event" suggests one necessarily has expectations about the event when this is not always the case, and "abnormal event" presupposes some unspecified, normative standard.

surprising outcomes, characterising them as low-probability events, disconfirmed expectations, schema-discrepant events or events of contrasting probabilities (e.g., Meyer, Reisenzein & Schützwohl, 1997; Reisenzein & Studtmann, 2007; Schützwohl & Reisenzein, 1999; Teigen & Keren, 2002, 2003). Sense-making theories stress the importance of understanding and integrating the surprising event; the cognitive goal elicited by surprising events is to establish the surprising outcome's coherence with respect to previous events, a task that is often carried out retrospectively rather than predictively (e.g., Kahneman & Miller, 1986; Maguire & Keane, 2006; Maguire et al., 2011; Pezzo, 2003; Thagard, 2000).

We lean more towards the sense-making view than the probability one. We argue that probability theories do not suggest the present experiments, though they can often be extended *ad hoc* to account for some of the results found here. Within the sense-making perspective, the main novelty of the present approach is its emphasis on the metacognitive, explanatory aspects of the sense-making process. As we shall see in the next section, these different theoretical accounts have often arisen as an aside to examining other cognitive processes.

2. The centrality of surprise: From basic emotions to hindsight bias

Beyond the literature that focuses directly on surprise, there is a substantial literature that identifies surprise as a "variable of interest" in exploring other diverse cognitive phenomena; from counterfactual thinking, to hindsight bias, to learning. This literature testifies to the centrality of surprise in cognition, as well as providing some evidential constraints on the core phenomenon.

2.1 Surprise, norms and counterfactual thinking

Kahneman and Miller's (1986) "norm theory" focuses primarily on the role of norms in counterfactual thinking, rather than on surprise *per se*. Although the literature on counterfactuals often uses unexpected events in its paradigms, surprise has not been studied to the same extent (see, e.g., Byrne, 2005). Kahneman and Miller argue that norms are used to generate contrasting counterfactuals to surprising events. They propose that events are perceived as abnormal (and surprising) if other outcomes are normatively, highly available, rather than because some expectation about the outcome has failed. The availability of these alternative outcomes may affect causal ratings of preceding events (e.g., McCloy & Byrne, 2002; Roese & Olson, 1996). If these alternative events are highly available, the outcome appears very abnormal, and hence surprising. In norm theory, these counterfactual, alternative outcomes are seen as being retrieved from memory or constructed after-the-fact (see Roese, 1997, for a review). Furthermore, it is the fact that these counterfactuals contrast with the outcome, that makes the outcome surprising (see also Teigen & Keren, 2003). There is also some evidence that counterfactuals may guide the search for the necessary causes of unexpected outcomes (e.g., Dehghani, Iliev & Kaufmann, 2012; Khaneman & Tversky, 1982; Roese & Olson, 1996; Wells & Gavanski, 1989), and so, potentially, surprising outcomes. For example, a serious accident will provoke an examination of its causes using generated counterfactuals (cast here as a "search for explanations"). However, in a series of experiments, Mandel (2003) found no indication that reasoning using counterfactuals had a stronger impact on causal judgments than factual reasoning, suggesting that counterfactual reasoning may not always affect surprise judgments. Furthermore, counterfactual thoughts and causal explanations may differ in their specificity (cf. McEleney & Byrne, 2006). In Experiment 4, we examine whether the elicitation of counterfactual knowledge impacts surprise.

2.2 Surprise and hindsight bias

In the hindsight bias literature several studies have examined the effects of surprise on hindsight bias. Hindsight bias, or the "knew it all along" effect, is the tendency to see events that have already occurred as being more foreseeable than they actually were. In this literature several different views on the relationship between surprise and hindsight bias have been advanced. Some propose that highly-surprising outcomes should reduce hindsight bias, as the effortful search to account for the surprising event produces an awareness that the outcome is very different from what was already known about the event (e.g., Müller & Stahlberg, 2007; Ofir & Mazursky, 1997). In contrast, others have proposed that surprising outcomes should increase hindsight bias (e.g., Schkade & Kilbourne, 1991), while some even argue that surprise has no effect at all (e.g., Carli, 1999; Wasserman, Lempert & Hastie, 1991). However, recent work reflects an emerging consensus that surprise can reduce hindsight bias, under specific circumstances.

Pezzo (2003; Pezzo & Pezzo, 2007) proposed a model of hindsight bias, in which unexpected outcomes trigger a sense-making process. Pezzo's initial model predicts that if this sense-making is successful then hindsight bias will occur, but if sense-making is unsuccessful then no hindsight bias will occur, and the surprise at the outcome will remain high (his later model includes defensive processing and retroactive pessimism). Similarly, Nestler and Egloff (2009) have reported that when participants perceived an outcome to be highly surprising, yet explainable, they saw it as more inevitable; that is, if causal antecedents for the outcome could be identified to explain why it occurred, the outcome was more likely to be one judged as having "to turn out like that" (see also Ash, 2009). Finally, Roese and Vohs' (2012) recent review concludes that surprising outcomes only increase hindsight bias if a coherent explanation for the event is found, one that successfully resolves the surprise. That is, it is the act of explaining the past outcome that makes it appear inevitable. Though this research does not often directly describe how explanation operates in surprise (e.g., Pezzo does not measure sense-making activity or its success/failure directly; but for an exception see Munnich, Milazzo, Stannard & Rainford's, 2014, recent work), these accounts are entirely consistent with the present approach. Indeed, in Experiment 5, we adapt a well-known manipulation from the hindsight bias literature to test for metacognitive aspects of surprise (i.e., eliciting different numbers of explanations for an outcome).

2.3 Explanation, surprise and learning

There is a long-standing view that surprise plays a key role in learning, and can increase the retention of information (e.g., Munnich, Ranney & Song, 2007), perhaps because surprise makes the event more interesting and likeable (Loewenstein & Heath, 2009). People learn about their environment by explaining it (see Lombrozo, 2012, for a review), whether it be in child development (Piaget, 1952) or education (Adler, 2008), and this explanation process may be triggered by surprise (Ramscar, Dye, Gustafson & Klein, 2013; Tsang, 2013) or inconsistencies (Johnson-Laird, Girotto, & Legrenzi, 2004). These views are echoed in Artificial Intelligence (AI), where surprise has been identified as a cognitive mechanism for identifying learning events in robotic, agent architectures (Bae & Young, 2008, 2009; Macedo & Cardoso, 2001; Macedo, Reisenzein & Cardoso, 2004; Macedo, Cardoso, Reisenzein, Lorini, & Castelfranchi, 2009). Traditionally, explanation is seen as playing a role in building causal models or predictive schemas to deal with future events (Heider, 1958; Lombrozo & Carey, 2006). However, apart from having a predictive role when a new situation is initially encountered, explanation may also serve to help people decide how information should be weighted or how attention should be allocated, as events occur (Keil, 2006).

In the education literature, Adler (2008) proposes that surprises give rise to a need for explanation and, as such, are of great value to learning. When students encounter a surprising piece of information their attention is aroused, provoking more intensive processing of the to-be-learned material (i.e., there is a call to explain, to correct and better understand the material). Self-explaining and self-explanation training is also known to improve text comprehension and learning (e.g., Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, De Leeuw, Chiu & Lavancher, 1994; Durkin, 2011; Roy & Chi, 2005), particularly for low-knowledge readers (McNamara, 2001; McNamara & Scott, 1999). Self-explanation appears to have a greater impact if there are reliable patterns and consistencies in the material uncovered by the explanations (Williams, Lombrozo & Rehder, 2010); although, conversely, it can be detrimental to learning in some cases if it leads to overgeneralisation (Williams, Lombrozo & Rehder, 2013). Adler also suggests that one can be surprised after being provided with an explanation for "why you should be surprised", even though no surprise was initially experienced. He gives an example of two friends watching a (US) football game, one of whom knows little and the other who knows a lot; when one player kicks the ball, a punt on third down, the friend who knows little is unsurprised, until the experienced friend *explains* what has happened and the novice is now surprised at what has occurred. While the present study does not explicitly address learning, it aims to elucidate the specific dependencies between surprising outcomes and explanation; as such, it should inform a deeper understanding of the role of surprise in education.

2.4 Summary on the cognitive centrality of surprise

This quick review of the literature identifies surprise as a "variable of interest" in other diverse cognitive phenomena and testifies to the centrality of surprise in many areas of cognition. There is a constant theme throughout this literature that surprise involves explanation, though these works seldom tackle surprise "head on" or, indeed, systematically explore how this explanation process might change from one surprising situation to the next. Here, we explicitly address surprise, examining some of the key factors that impact the explanation process and, in turn, the perception of surprise. However, before reporting these experiments, we outline the theory from which they arise.

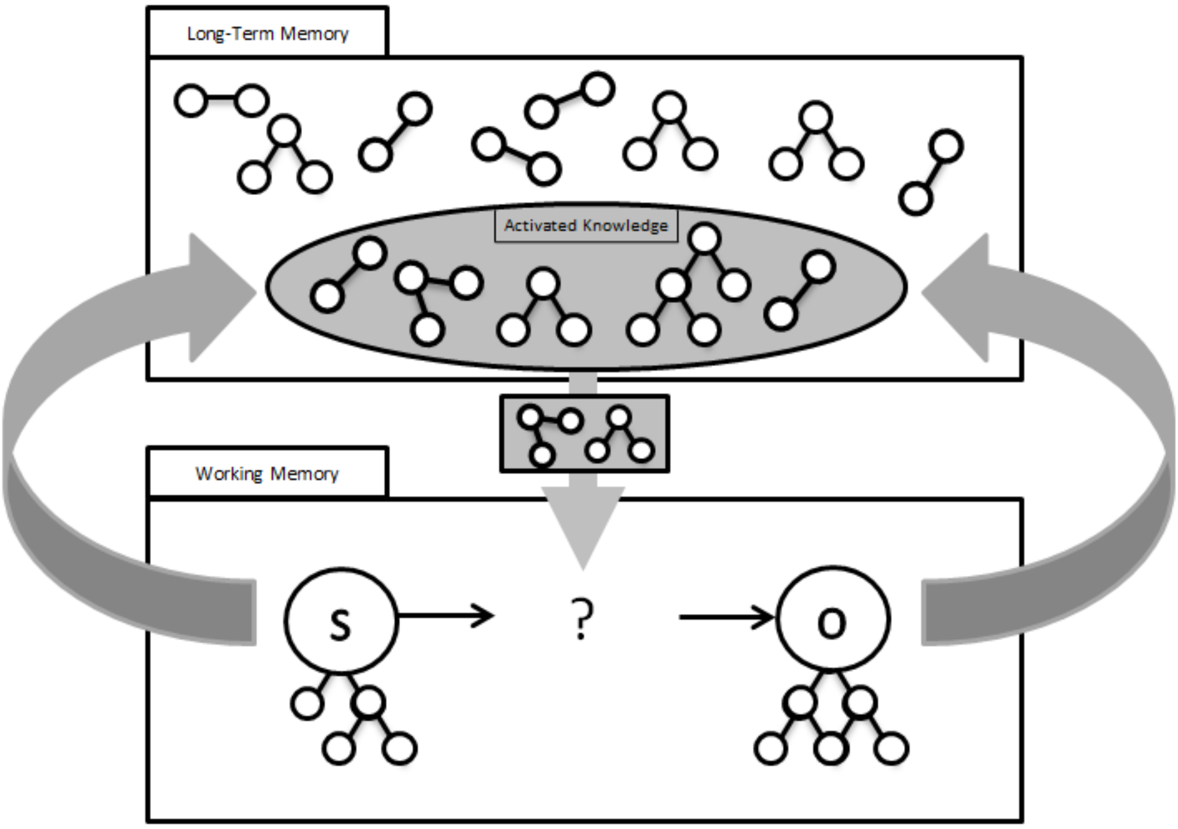


Figure 1. The MEB Theory of Surprise: In understanding a surprising event, the information provided in the setting (S) and outcome (O) activates knowledge in memory (the greyed elipse) that can be retrieved (shown by the downward arrow) to build explanations (the greyed rectangle) to link the setting (S) to the outcome (O).

3. Surprise as a metacognitive sense of explanatory difficulty

Anecdotally, people's responses to surprising events can vary in their cognitive complexity, intensity and longevity. Some surprises are shortlived, transient upsets that are quickly resolved; "Where are my keys, they are not in my pocket! Oh, yes, I put them in my bag". Other responses appear to be prolonged, intense attempts to make sense of the surprising event; survivors of disasters often suffer post-traumatic stress in which they repeatedly attempt to understand the experience over months or even years (e.g., survivors of the Concordia or Zeebrugge ferry disasters).

These scenarios raise questions about the role of surprise in human cognition. What is the point of surprise? Why do sentient beings, like ourselves, get surprised? Does it have some evolutionary value? Or, more precisely, "what is the cognitive goal of surprise?" and "what is the computational problem being solved by surprise?"

3.1 The role of surprise: The computational level

We have seen that surprise plays a role in many diverse cognitive phenomena. This fact, combined with its ubiquity in everyday life, suggests that it plays a significant role in human cognition. Our view is that surprise acts to direct current attention to discrepant information in the environment, assigning cognitive resources to resolve that discrepancy (i.e., explain it), subsequently marking this resolved-scenario for future retrieval. The emotional aspect of the surprise helps mark the memory for subsequent retrieval, such that if, for example, a similar surprising-event occurs in the future, it will cue the resolved-scenario (perhaps even prompting a generalisation from both events). Furthermore, the degree of surprise experienced, which is a side-effect of the explanation process, is an index of the seriousness of the discrepancy occurring; that is, the extent of the divergence between the surprising event and prior knowledge. So, the perception of the suprisingness acts to signal the importance of the resolved-scenario for future use (e.g., in learning or thinking).

Theoretically, at the computational level, a theory of surprise needs to specify the "goal of the computation" and "what needs to be computed?" (see Jones & Love, 2011; Keane, Ledgeway & Duff, 1994). Following on from the above proposals, we see the *cognitive goal* of surprise as being "to explain the discrepancy that arises when a surprising outcome occurs". This requires the computation of two things: (i) an explanation that links the setting and outcome of the scenario in a coherent way (i.e., one that "makes sense"), and (ii) a metacognitive assessment of the amount of cognitive work done in this explanation step, which is a proxy measure for the divergence between the surprising event and prior knowledge. So, in the lost-keys example, the keys-are-in-my-bag explanation is retrieved from memory, resolving the discrepancy of "the keys not being in my pocket"; the emotional shock is minimal, as is the cognitive work done. This surprise experience does not prompt a major reassessment of one's knowledge. But, in the ferry-disaster case, following the significant shock experienced, an extended attempt to explain the the disaster could occur, involving significant cognitive work. Indeed, in time, this surprise experience could prompt a deep reassessment of one's views about the safety of ferries.

Stated succinctly, the MEB theory of surprise posits that: *Experienced surprise is a metacognitive assessment of the cognitive work carried out to explain an outcome*. Very surprising events are those that are difficult to explain, while less surprising events are those which are easier to explain. Importantly, this high-level, computational account of surprise is supplemented by algorithmic proposals that generate the key empirical predictions tested in this paper (see Jones & Love, 2011; Keane et al., 1994).

3.2 The course of surprise: The algorithmic level

At the *algorithmic level*, a theory of surprise needs to detail "*how* surprise is computed", specifying the various cognitive structures and processes that affect the performance of these computations. In this respect, we adopt several widely accepted ideas that are shared by most cognitive architectures and theories of comprehension (e.g., Anderson, 1983, 1993; Graesser, Millis, & Zwaan, 1997; Graesser, Singer & Trabasso, 1994; Kintsch, 1998; Laird, 2012; Laird, Newell & Rosenbloom, 1987). During comprehension, people construct bridging inferences between observed events and activated memory representations, developing complex relational and causal structures (e.g., Gernsbacher, 1990, 1991, 1997). So, the comprehension of a scenario is assumed to involve building a *situation model* in working memory (e.g., Graesser & McNamara, 2011; O'Brien, Rizzella, Albrecht & Halleran, 1998; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). Broadly speaking, the surprising scenario divides into a representation of a *setting* (i.e., the initial state(s) of the scenario, expressed by its various concepts and relations) and the *surprising outcome* (the target state capturing the concepts and relations of the discrepant event). Figure 1 depicts these components graphically, where the various pieces of knowledge are shown as graph-like structures in long-term memory and working memory. During normal comprehension the setting and outcome are continuously cueing relevant knowledge in long-term memory, to form elaborative and bridging inferences. As such, we assume that this creates a *region of activated knowledge*, based on spreading activation in long-term memory. This region supplies the knowledge used to build explanations that resolve the surprising outcome; to conceptually link the scenario's setting to the outcome that occurs (see Figure 1).

Finally, the perception of surprisingness is based on a metacognitive assessment of the effort-to-explain, the amount of cognitive work carried out to explain the outcome. Explanation finding will involve different amounts of cognitive work in retrieval, inference and integration processes. The cognitive system roughly tracks the amount of work done in explaining the surprising outcome and this assessment underlies the perceived surprisingness of the target event. In low-work cases, explanation finding may involve the straight retrieval of pre-formed explanations (e.g., "the last time I lost my keys they were in my bag"). In high-work cases, the explanation finding may involve a succession of steps, including several retrievals, complex inferencing and the integration of diverse pieces of knowledge from long-term memory (e.g., new object- or actor-concepts may need to be introduced into the scenario to explain it, such as "someone must have stolen my keys").

This model of surprise[2] generates many novel, testable predictions. As we shall see in the next section, most of our experiments explore how these algorithmic-level factors affect surprise.

Table 1

Examples of the materials used in Experiment 1 (adapted from Maguire et al., 2011): The known (Louise) and less-known (Bob) scenarios are divided into setting and outcome parts (not explicitly labelled in the presented materials).

	Known (Louise Scenario)	**Less-Known (Bob Scenario)**
Setting	Louise is going shopping. She takes out €200 from the ATM and puts it in her wallet. She gets the bus into town and stops at her favourite clothes shop.	Bob has wanted to quit his job for months. Today is the final straw. He has been working overtime every day and has been getting no credit at all. Bob marches into his boss's office in a fit of rage.
Outcome	She had lost her wallet.	He gave his boss a hug.

4. Predictions from MEB Theory

A number of implications for the phenomenon of surprise arise from the present theory, that give rise to the predictions tested here. They are that:

(i) *Memory contents are critical in surprise*: the contents of long-term memory will play a key role in resolving surprising events, they are "raw material" for explanation; if there is extensive domain knowledge about a scenario then the

[2] Obviously, these processes could be modelled in more detail; for example, explanation building could be modelled used case-base reasoning techniques (Haton, Keane & Manago, 1995; Leake, 1991; Schank, Kass & Riesbeck, 1994), or, perhaps, as parallel-constraint satisfaction (Thagard, 2000). However, the current model is of sufficient granularity to generate the predictions tested in the current experiments.

surprising event is more likely to be resolved easily, than if there is very little domain knowledge about the scenario[4].

(ii) *Scenarios are cues*: The scenario (i.e., its setting and outcome) acts as a cue to relevant knowledge in memory, so the resolution of a surprising event is critically influenced by the given information in the scenario. It follows that there may be distinct classes of scenario or parts-of-scenarios that systematically differ in how well they tap knowledge in memory. For instance, we explore two broad classes of scenario, termed "known" versus "less-known" scenarios that differentially cue knowledge. Furthermore, parts of a scenario could also miscue the explanation process by accessing knowledge that inhibits the resolution of the surprising outcome (i.e., when less-relevant knowledge is cued).

(iii) *Partial explanations will reduce surprise*: Obviously, providing additional information that partly furnishes an explanation (e.g., an enabling sentence or a partial explanation) should predictably reduce surprise in systematic ways.

(iv) *Task demands can affect surprise*: Finally, it follows from the theory that any task demand that reduces/increases the cognitive load in explaining the surprising outcome, should systematically lower/raise perceptions of surprise respectively (e.g., instructions to explain it).

In the following sub-sections, we elaborate the theory's predictions, map them to the experiments carried out, and indicate supporting evidence for them in the existing literature.

[4] Although, there may be a limiting case in which there is absolutely no knowledge about the scenario, that may have its own distinct, unique properties

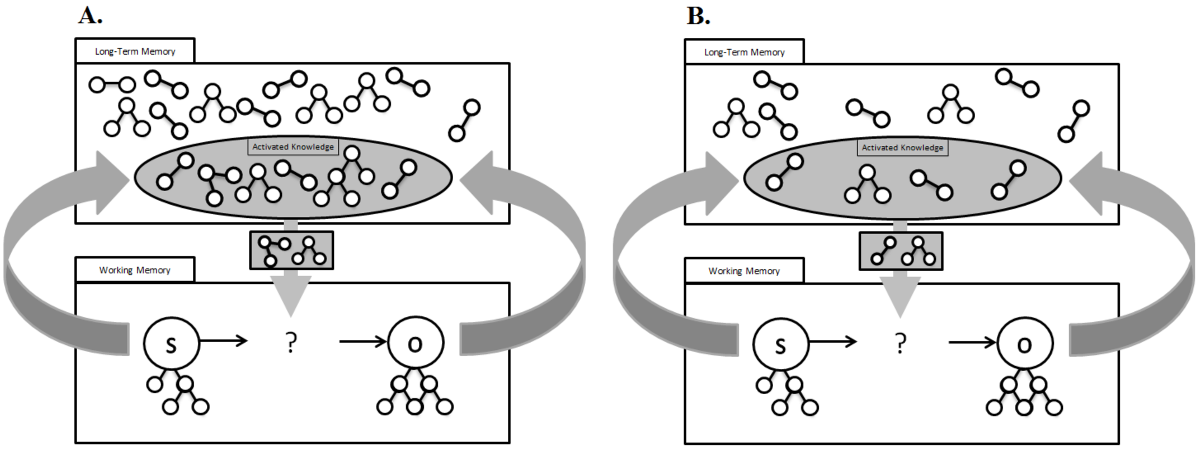


Figure 2. The contents of activated knowledge in memory for A) known scenarios and B) less-known scenarios. Known scenarios have a lot of associated relevant knowledge that is activated in memory, cued by the setting (S) and outcome (O), allowing for easier explanation (and less surprise) relative to less-known outcomes, which activate less knowledge incurring a harder explanation process (hence, more surprise)

4.1 Memory contents are critical in surprise

Clearly, other things being equal, the sheer amount of domain knowledge about a scenario in memory will have an impact on the levels of surprise people experience. If the region of activated knowledge in memory contains extensive domain knowledge about the scenario, then the "raw material" for explaining surprising outcomes will be richer, explanations will be more readily produced and, other things being equal, preceived surprise will be lower[6]. Figure 2 graphically represents this state of affairs, showing how the region of activated knowledge may be densely (Figure 2A) or sparsely (Figure 2B) populated for explanation purposes. Empirically, it is hard to measure these differences in memory contents as they are filtered by how the scenario cues that knowledge. However, in some of our experiments, we provide evidence to show that explanations have systematic regularities,

[6] Anecdotally, this might account for why experts in a domain tend to be less surprised by abnormal events; they have "seen it all before", that is, they have extensive available knowledge to explain discrepant events, in quantities that novices lack.

indicative of differences in the knowledge used to form them (see Experiments 2, 3, 4 and 6a).

4.1.1 Previous research on memory contents in surprise

We know of no prior empirical work that explicitly addresses this contents-of-memory issue, though the proposal echoes previous ideas. As we saw earlier, Kahneman and Miller (1986) stressed the role of norms in bringing certain parts of prior knowledge into focus when a surprising event occurs. In terms of the current theory, this amounts to a proposal that memory relies more on retrieving normatively-packaged knowledge for explaining surprising outcomes than on, for example, unique, precedent events (unless a norm has been computed from this single exemplar, cf. Kahneman & Miller, 1986).

4.2 Scenarios are cues

Scenarios cue relevant knowledge for use in explanation; if the scenario cues a lot of relevant knowledge for explaining the surprising event, then explanation will be facilitated and surprise will be low, whereas if the scenario does not cue relevant knowledge, then explanation will be inhibited and surprise will be high. Obviously, this cueing role of the scenario interacts with the amount of knowledge in memory; though, we really only see this knowledge through the filter of what the scenario cues (see Figure 2). So, in theory, even if memory is filled with potentially-relevant knowledge, should the given concepts in the scenario not properly cue this knowledge, then explanation may be difficult or, even, fail.

MEB theory explicitly partitions scenarios into (i) setting information and (ii) outcome information. In any scenario, the setting information establishes the context for what is occurring - it identifies the main actors, relevant background knowledge and the events that are unfolding, whereas the outcome tells us about the target surprising event that has occurred. So, in an account of Lady Diana's death, the setting describes the main actors and the events leading up to the accident. Minor details given in the scenario could cue or miscue one explanation over another. If I was telling you the story of Lady Diana's death and I said "a *pale* Lady Di left her Paris hotel before her death" with no futher facts about how she died, you would probably develop a health-explanation for her death rather than an accident-explanation[7]. So, what is given in the scenario can strongly influence the ease and nature of the explanations produced to resolve the surprising outcome. In terms of our graphical depiction of the theory, the given scenario defines a region of activated knowledge shown (see Figures 1 and 2); so, different scenarios will set up different regions of activated knowledge, which may contain more or less relevant knowledge depending, in part, on the contents of memory. In the current set of experiments, three distinct aspects of the cueing role of scenarios are tested; specifically, whether there are:

(i) distinct classes of scenarios that cue differentially (so-called known versus less-known scenarios),

(ii) distinct classes of outcomes in scenarios, that cue differentially (known versus less-known outcomes),

(iii) parts of a setting that can miscue knowledge (using cueing versus miscueing keywords)

4.2.1 Classes of scenario (known versus less-known)

Adopting the view that scenarios cue different collections of knowledge in memory raises the possibility that there are different classes of scenario; some scenarios may cue large amounts of relevant knowledge whereas others may not. Imagine a surprising scenario in which you discover your wallet is missing from your trouser-pocket. This missing-wallet scenario seems likely to cue a lot of knowledge about missing items being lost or stolen.

[7] This is like the "rifle hanging on the wall" rule in drama (known as "Chekhov's gun"). If a scene in a play has a rifle on the wall, then inevitably it will be used at some later part of the drama, i.e., only elements that are irreplaceable and necessary to the story should be described.

Scenarios like this one, that cue a lot of relevant knowledge, could be called “known” surprising scenarios. In contrast, imagine a scenario where a friend describes hugging his boss, after planning to quit his job in anger (see Table 1). This boss-hugging scenario seems, intuitively, to cue less knowledge to help resolve the surprising outcome. As such, these types of scenarios could be called “less-known” surprising scenarios. Though both of these classes of scenario are surprising in their own way, the former seems less surprising than the latter, because of the relevant knowledge brought to bear to resolve the surprising outcome. Experiments 1 and 2 specifically test these different classes of surprising scenario, as the variable of Scenario-Type (known versus less-known).

4.2.2 Classes of Outcome (known versus less-known)

Just as there may be different classes of scenario (known and less-known), there may also be different classes of outcome (known and less-known), that differentially impact perceived surprise, keeping the setting information constant. Indeed, some surprising events appear to cue a set of “ready-made” explanations (see Schank, 1986). Imagine, one day you are walking home and discover that your wallet is missing from the pocket of your jeans. You would be surprised, but also have some ready explanations for what might have occurred (e.g., “Is it in the shop I just left?”, “Could it have been robbed?”, “Might I have dropped it?”). Now, imagine you are walking home and discover that your belt is missing from the waist of your jeans. Again, you would be surprised, but few explanations present themselves (the only one we could think of, eventually, was leaving one’s belt in the security area at the airport). So, here, a small change in the object mentioned in the outcome (i.e., wallet or belt) subtly changes the activated knowledge, altering the ease with which the surprise is resolved. Here, we would predict that people would find wallet-losing a lot less surprising than belt-losing, other things being equal[8]. This prediction is tested using a number of different measures in Experiments 3-6.

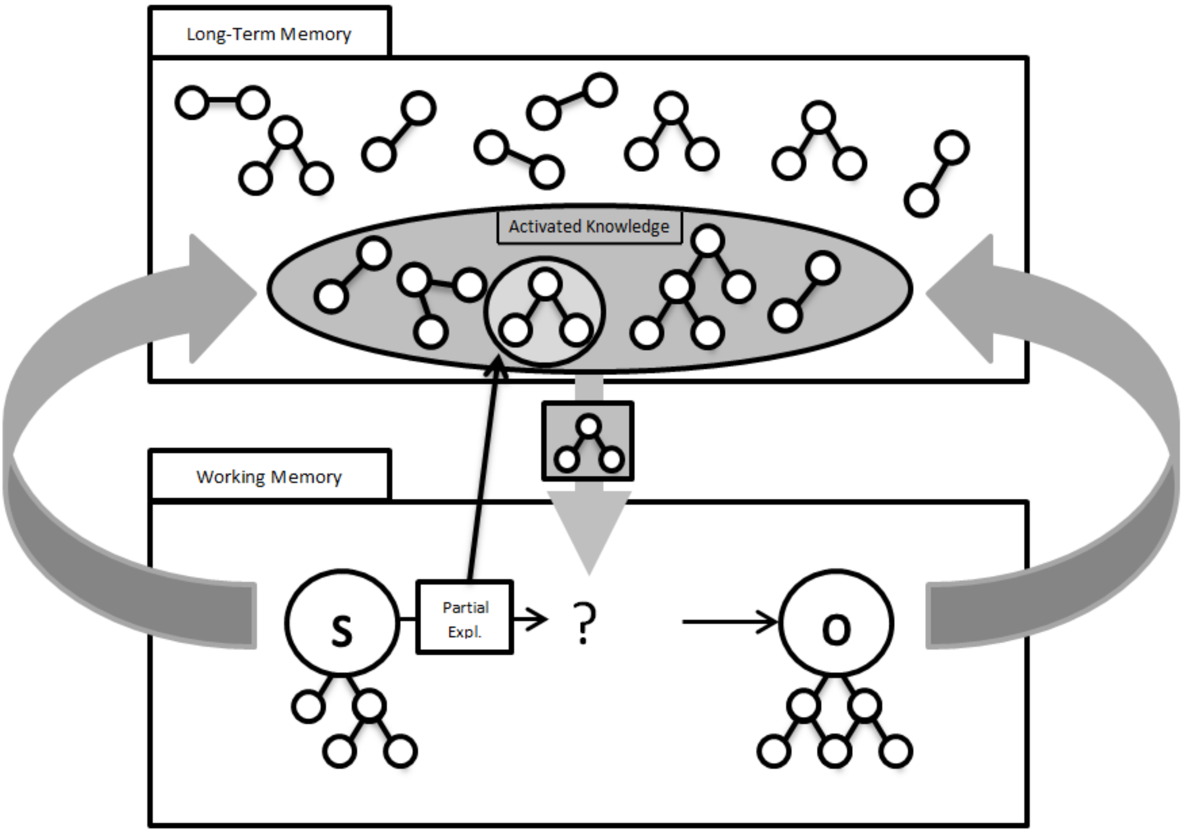


Figure 3. The retrieval of concepts from memory when a partial explanation (enabling information) is provided. The setting (S) and outcome (O) cue knowledge in memory, creating a region of activated knowledge. If a partial explanation is provided then the knowledge that is activated is further constrained and becomes the most likely candidate (greyed square) for explaining the outcome.

4.2.3 Parts of the setting can miscue knowledge

The idea that the scenario is a cue to critical knowledge in long-term memory, carries with it the corollary that specific parts of a scenario could also miscue knowledge; that is,

[8] Indeed, we would also predict that if the setting mentioned "airport security" or "airport" then the experienced surprise would decrease markedly for the belt-losing outcome, as the retrieval process would be more directed by this given information towards concepts that could be used to build the airport-explanation.

some contents of a setting could cue knowledge that is really *not* useful in explaining the outcome that has occurred (i.e., people could be "sent the wrong way"). As we saw earlier, if I began my story of Lady Diana's death by mentioning that she was "pale" on the night in question and stopped there, you would probably find yourself, with difficulty, trying to construct health-explanations for her death. This idea is tested in the final series of experiments in this paper (Experiments 6a-6c) using a novel paradigm, in which keywords are introduced with the setting of the scenario, that either cue or miscue explanations for the outcome.

4.2.4 Previous research on scenarios as cues

By definition, previous surprise research has often examined apects of how scenarios cue relevant knowledge (e.g., Choi & Nisbett, 2000; Gendolla, 1997; Gendolla & Koller, 2001; Maguire et al., 2011; Teigen & Keren, 2002, 2003), though it has not been framed in this way. For example, Maguire et al. (2011; Maguire & Keane, 2006) presented people with short stories, in which they first presented the setting information, followed by a surprising outcome, and asked them to judge the surprisingness of the outcome (see materials in Table 1). Much of their research used scenarios where the outcome confirmed or disconfirmed expectations, and they showed that the ease of integrating an outcome into an existing representation (presumably created from the preceding setting) affected perceived surprise. From our perspective, this ease of integration arises from cueing of information that allows for easier explanation of the outcome of that scenario.

However, no previous experiments have specifically explored the idea that (i) there are different classes of scenario that cue memory differentially (known versus less-known), (ii) there are different classes of outcome that cue differentially (known versus less-known), or (iii) that miscueing can occur.

4.3 Partial explanations may reduce surprise

When considering the information given in the setting, there is one special case of cueing that deserves separate treatment on it own; this is the case where a partial explanation is given in the setting that is relevant to explaining the outcome. For instance, if I provided you with additional information about a surprising outcome, for example – that Michael Jackson took 50 sleeping pills before he died – then this extra information would reduce one's surprise at the event, as an explanation can now be built, with minimal knowledge retrieval (e.g., only a single causal inference is required, "that the pills *caused* his death"). Here, the setting has been extended to provide key, enabling information for one explanation over others; when people avail of this information, the explanation process is eased and surprise should decrease (see Figure 3). Interestingly, this partial explanation should also act to constrain the production of alternative explanations, impeding the consideration of such other explanations because they do not fit the given information (a prediction we provide evidence for later, in Experiment 2). In the current paper, we attempt to replicate and elaborate previous findings on this, though we cross it with other manipulations (see Experiment 2).

4.3.1 Previous research on partial explanations in surprise

Maguire et al. (2011) manipulated such enabling information (or partial explanations) in the settings of different story scenarios; they found that providing enabling conditions for explaining an outcome reduces the perceived surprisingness of these outcomes. For example, in one of their story-scenarios, a woman withdraws money from an ATM, putting it into her wallet (the setting), only to find out later that the wallet is missing from her handbag (the surprising outcome). When additional information was added to the setting – "her handbag was open" – judgments of the surprisingness of the outcome decreased. According to the present theory, this enabling information either makes explanation very easy (e.g., just a single inference needs to be made), or it directly cues relevant knowledge for an explanation (e.g., that “the wallet was stolen” or that “the wallet fell out of the handbag”).

4.4 Task demands may reduce surprise

A third and final set of predictions from the MEB theory is that various task demands should impact surprise; any task demand that reduces/increases the cognitive load in explaining the surprising outcome should tend to reduce/increase perceptions of surprise. So, for instance, if the task demands require explaining an outcome before judging its surprisingness, the instruction to produce an explanation should focus and ease processing, reducing the perceived surprisingness of the outcome (see Experiments 3 and 4). However, this effect should be partially reversible if participants are asked to produce several explanations, as opposed to one, as this manipulation increases the cognitive load of the task (see Experiment 5).

4.4.1 Previous research on task demands in surprise

No previous work known to us has examined the task demands around instructions to explain the outcome affecting surprise. However, one other task demand involving “time-of-processing” has been studied. Maguire and Keane (2006) found that given enabling information reduced surprise judgments more if it was presented *in advance* of the outcome sentence (in the setting of the discourse), as opposed to *simultaneously* with the outcome sentence. This “time-of-processing” effect, for us, hints at the dynamic nature of the way knowledge is recruited to explain an outcome and reflects the metacognitive aspect of surprise; namely, that even though the actual information provided was the same in both scenarios, giving people more time to consider it reduced the cognitive load during judgment, resulting in lower surprise ratings.

5. Experiment 1: Effects of Scenario-Type on surprise judgments

A key prediction of the present theory is that there are distinct classes of scenario, that cue knowledge differentially to help/hinder the explanation process, with attendant effects on perceived surprise. So, "known" scenarios (e.g., Louise losing her wallet) should ease the explanation process, reducing perceptions of surprise, and "less-known" scenarios (e.g., Bob hugging his boss) should inhibit the explanation process, increasing perceptions of surprise. Anecdotally, when we re-examined the story scenarios used by Maguire et al. (2011) they appeared to divide into these two classes. This intuition was confirmed by a pre-test that involving sorting the stories into based on an assessment of their outcomes (see *Materials* section).

Using the two identified classes of scenarios, in the present experiment, we repeated Maguire et al.'s earlier study to see whether the scenarios (known versus less-known) differentially affect people's surprise ratings. The current theory predicts that "known" scenarios should cue more relevant knowledge that eases explanation-building for the surprising outcome, thus reducing perceived surprise, whereas "less-known" scnearios should furnish less relevant knowledge, making the explanation process more difficult and, thus, relatively increasing perceived surprise.

5.1 Method

5.1.1 Participants and design

Forty University College Dublin (UCD) students (18 male, 22 female) with a mean age of 21.88 years (*SD* = 3.62, range = 18-33) voluntarily took part in this experiment. Informed consent was obtained prior to the experiment. Originally, a task-variable was also examined in this experiment (participants were asked to either reflect on the scenario or were distracted by counting backwards, before the outcome sentence was presented) but this variable proved to have no effect on subsequent surprise judgments and, as such, is not reported in the present analyses. For the record, this negative finding indicates that an interspersed "irrelevant" task is clearly not counted as cognitive work towards or against explaining the outcome and, therefore, does not contribute to perceived surprise.

5.1.2 Materials

In order to identify the two classes of scenario, as known or less-known, an separate group of people (*N* = 5) we asked to sort the 16 scenarios used by Maguire et al. into two groups (see Table 1 for an example). These sorters were asked to determine if a given scenario had an outcome that "*falls within the range of reasonable outcomes to the scenario*" (i.e., known surprising outcomes) or whether it "*falls less within the range of reasonable outcomes to the scenario*" (i.e., less-known surprising outcomes). Of the 16 stories, the raters consistently deemed 6 of the stories to have known outcomes and 10 to have less-known outcomes (Cronbach's alpha showed inter-rater reliability to be high, α = .79).

5.1.3 Procedure and scoring

Participants were tested individually, asked to read 16 short stories, and judge the surprisingness of their outcomes. The order of presentation of these stories was randomised anew for each participant. Stories were presented sentence by sentence on a laptop computer screen as participants pressed the spacebar, with each sentence replacing the preceding one on the screen. Twenty seconds after the presentation of each scenario, the outcome was presented after the question "*How surprised would you be if*", and participants were asked to indicate on a 7-point scale their surprise judgment (1: not surprising, to 7: very surprising). Participants' ratings for the 6 known scenarios and the 10 less-known scenarios were averaged to compare the two groups statistically.

5.2 Results and discussion

A paired t-test confirmed that participants judged *known* scenarios to be less surprising ($M = 3.74$, $SD = 0.90$) than *less-known* scenarios ($M = 5.19$, $SD = 0.85$), $t(39) = -11.86$, $p < 0.001$; a by-materials independent t-test was also statistically reliable. So, though all of these stories were considered to have surprising outcomes, some were less surprising than others.

5.2.1 The MEB account.

These results follow directly from the MEB theory's proposals about how scenarios cue different knowledge for explaining surprising events. The type of scenario (known versus less-known) was found to have a significant effect on participants' surprise judgments. With the benefit of hindsight, this finding may seem quite obvious; that there are, of course, some scenarios that are less surprising than others, because of the amount of relevant knowledge cued by their given contents. However, notably, these distinctions have not previously been systematically examined in the literature. Indeed, Maguire et al. (2011) used these story-materials as a homogenous group in a whole series of experiments, without recognising that they involved two distinct sub-classes of scenario. The distinction only became clear in the context of the current theory.

5.2.2 Formulating a probabilistic account

So, how would probability theory deal with these results? It could be argued that probability theories are not strongly motivated to identify different classes of scenario, but they could see them as emerging if scenarios were grouped together on the basis of them having the same or similar probabilities. Accepting this caveat, these accounts would presumably argue that known scenarios and less-known scenarios are suprising because they both involve low-probability outcomes, but that, relatively, the former have *higher* low-probability outcomes than the *lower* low-probability outcomes of the latter group. It is, however, difficult to find good evidence for this account.

Using the exact same story scenarios, Maguire et al. (2011, Experiment 4) asked different groups of particpants to either (i) judge the surprisingness of presented outcomes, (ii) generate the most likely outcome and estimate its probability, or (iii) estimate the probability of the surprising outcomes (without judging surprise). They then performed correlations and regression analyses to determine whether the subjective probabilities found predicted the pattern of surprise judgments. First, they found that the probability of the likely outcome did not predict surprise ratings (r = -.34, p > .05). Second, they found that people's probability estimates of the presented, surprising, outcome *did* predict surprise (r = -.52, p =.003). On the face of it, this appears to support a simple probability account, but it is a weak support, as the "prediction" only works after-the-fact. First, since the surprising outcomes for these scenarios are rarely (if ever) generated, it is difficult to estimate their probability before-the-fact. Second, it is cognitively implausible to suppose that people generate expectations for every possible outcome in a scenario before-the-fact (c.f., Kahneman & Miller, 1986). Third, the finding that after-the-fact probability estimates correlate with surprise ratings is less adequate because these estimates may simply reflect how the outcome's probability has been changed by having judged its surprise.

However, perhaps there are more sophisticated probability accounts that could do a better predictive job. Some probability theorists argue that it is the contrasting probabilities between prior and posterior belief distributions, alternative causal models, or the expected outcome, that should predict surprise (e.g., Baldi & Itti, 2010; Griffiths & Tenenbaum, 2007; Teigen & Keren, 2003). Maguire et al. checked this contrasting-probability account and found that it did not predict surprise ratings (r =.30, p > .05). So, while a probability account seems initially plausible it is hard to show that it is strongly predictive of surprise judgments for these scenarios. We will revisit these issues throughout the paper, as new evidence is found.

Table 2

Examples of the materials used in Experiment 2. The known (Louise) and less-known (Bob) scenarios are shown divided up into setting and outcome parts. The enabling/non-enabling sentences are italicised to emphasise the key differences between story versions (N.B. italics and labels were not used in the presented materials, except for the outcome label).

		Scenario-Type	
		Known (Louise Scenario)	**Less-Known (Bob Scenario)**
Setting		Louise is going shopping. She takes out €200 from the ATM and puts it in her wallet. She gets the bus into town and stops at her favourite clothes shop.	Bob has wanted to quit his job for months. Today is the final straw. He has been working overtime every day and has been getting no credit at all. Bob marches into his boss's office in a fit of rage.
Partial Explanation	Present	*Her handbag was open.*	*He was offered a promotion.*
	Absent	*She looked in her handbag.*	*He approached the large wooden desk.*
		How surprised would you be if:	How surprised would you be if:
Outcome		She had lost her wallet.	He gave his boss a hug.

6. Experiment 2: Effects of Scenario-Type on surprise ratings, response times and explanation productivity scores

Experiment 1 provided evidence for the novel prediction that there exist different classes of surprising scenario; that known scenarios can be distinguished from less-known scenarios. However, the results of Experiment 1 do not provide direct evidence that people explain these scenarios differently or, indeed, whether the time-course for resolving known scenarios differs from less-known scenarios. So, in this experiment we selected the four best examples of known and less-known scenarios (based on the rating pre-test in Experiment 1). To gather evidence on the knowledge activated by each scenario, after the surprise-rating task participants were asked to generate explanations for the outcome of each scenario (i.e., the explanation productivity measure). Finally, response times was also recorded in this experiment. On this point, one would expect the known scenarios to be resolved and judged faster than the less-known scenarios, because of the quantity of relevant knowledge cued by the former over the latter.

Experiment 2 also manipulates part of the *setting* of the scenario by proving a partial explanation, to examine how supporting setting information might impact perceptions of surprise. Previously, Maguire et al. presented people with enabling sentences (e.g., "Her handbag was open") that were matched with non-enabling control sentences (e.g., "She looked in her handbag", see Table 2). They found that the provision of such enabling conditions reduced surprise ratings (see Johnson-Laird et al., 2004, for a discussion of the differences between enabling and causal conditions); but they did not examine explanation productivity or response time. In the context of the present theory, this additional information is seen as providing partial explanations that ease the cognitive load by either (i) directing the retrieval process to more relevant parts of memory, or (ii) directly providing a significant part of the explanation, thus minimising retrieval and inference, or (iii) both (see Figure 3, for a graphical depiction).

We did productivity & Maguire did not

So, Experiment 2 had a 2 x 2 design with Partial-Explanation (present vs. absent) as a between-subjects variable and Scenario-Type (known vs. less-known) as a within-subjects variable. The MEB theory predicts that these manipulations will affect the explanation process, influencing surprise judgments. More specifically, that (i) known scenarios known should be easier to explain, lowering surprise judgments relative to less-known scenarios, (ii) partial explanations present in the settings should ease the explanation process, and reduce surprise judgments relative to settings lacking such partial explanations, and (iii) these two variables should act together in an additive fashion so that the lowest perceived surprise should be for known scenarios when partial explanations are present.

6.1 Method

6.1.1 Participants and design

Forty UCD students (14 male, 26 female) with a mean age of 22.6 years (*SD* = 4.46, range = 18-39) took part voluntarily in this study. Informed consent was obtained prior to the experiment. Participants were randomly assigned to one of two conditions in a 2 (between-subjects; Partial-Explanation: present, absent) x 2 (within-subjects; Scenario-Type: known, less-known) mixed-measures design.

6.1.2 Materials

Eight story scenarios were used in this experiment, based on the best examples of these two classes of scenario (known and less-known) from the set used by Maguire et al. (2011). For this Scenario-Type variable, four stories with known scenarios and four stories with less-known scenarios were used. For the Partial-Explanation variable, the partial-explanation-present stories contained sentences with enabling information that were presented in the story setting before each outcome sentence, while in the partial-explanation-absent stories an irrelevant sentence of equal length was inserted at the same point in the story (see Table 2). The order of presentation of these stories was randomised anew for each participant.

6.1.3 Procedure and measures

The experiment had two tasks: a surprise judgment task and a generation task (always presented in that order). The *surprise judgment task* used a version of the surprise rating task from Experiment 1; participants were tested individually on a laptop computer, asked to read the eight stories, and to judge the surprisingness of their outcomes. Additionally, for this experiment each individual's response time to carry out each surprise judgment was recorded. In the *generation task*, participants were given a booklet containing the same eight scenarios and were asked to produce "*as many different explanations as they could for why the final event in the story might have occurred*" (in the scenarios with enabling sentences, they were specifically asked for alternatives to the partial explanation). Maguire et al. (2011, Experiment 3) asked participants to generate one explanation for each story scenario, but they did not analyse these explanations apart from selecting the most dominant explanation to present to another group of participants.

Three measures were recorded: (i) the surprise judgment rated on a 7-point scale, (ii) the response time to make this surprise judgment, measured from the point of presentation of the outcome sentence to the participant's indication of their judgment, and (iii) the total number of different explanations produced by each participant to each given scenario (the *explanation productivity* measure).

6.2 Results and discussion

Overall, the results replicated the findings of Experiment 1, that the type of scenario influences perceived surprise. The role of partial explanations in the setting also replicated previous findings. Additionally, the experiment revealed novel findings on (i) the interaction between Scenario-Type and Partial-Explanation variable in surprise judgments, (ii) the impact of Scenario-Type on the time-course of surprise judgments, and (iii) people's productivity in explaining different outcomes (varied by Scenario-Type and Partial-Explanation).

6.2.1 Surprise judgments

A two-way ANOVA revealed a main effect of Scenario-Type where participants judged stories with known scenaros to be less surprising (*M*= 2.81, *SD* = 1.17) than less-known scenarios (*M* = 4.50, *SD* = 1.45), *F*(1, 38) = 112.29, $p < .001$, $\eta_p^2 = .75$. There was also a significant main effect of Partial-Explanation, *F*(1,38) = 25.12, $p < .001$, $\eta_p^2 = 0.40$, replicating the findings of Maguire et al.; participants judged the settings without partial explanations as more surprising (*M* = 4.41, *SD* = 1.04) than those with partial explanations (*M* = 2.90, *SD* = 0.85).

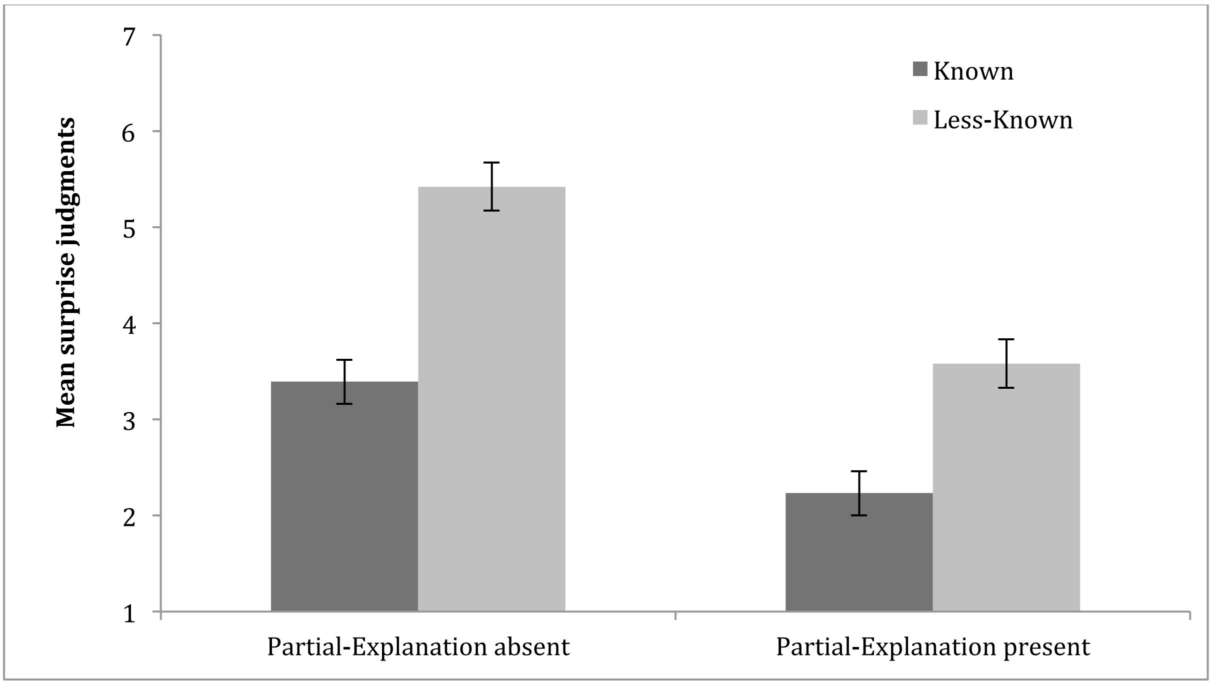


Figure 4. Mean surprise judgments in Experiment 2 for both levels of Scenario-Type (known vs. less-known) and Partial-Explanation (absent vs. present) with standard errors (*N* = 40).

Furthermore, the two variables combine additively to impact surprise judgments, and there is a small but reliable interaction between Scenario-Type and Partial-Explanation, showing that providing partial explanations lowered surprise more for less-known scenarios than for known scenarios, $F(1, 38) = 4.63$, $p = .03$, $\eta_p^2 = .11$ (see Figure 4). The condition with known scenarios and partial explanations yields the lowest surprise judgments (M = 2.23, SD = 0.76) whereas the condition with less-known scenarios and no partial explanation yields the highest surprise judgments (M = 5.42, SD = 1.11), with the other conditions falling in between; the condition with known scenarios and partial explanations absent (M = 3.39, SD = 1.23) and that with less-known scenarios and partial explanations present (M = 3.58, SD = 1.14). The order of these conditions suggests that Scenario-Type has a marginally higher impact on perceived surprise in this interaction than the provision of partial explanations in the setting. So, in terms of the MEB theory of surprise, known scenarios ease the resolution of the surprising outcome and, consequently, lower surprise judgments. In contrast, the less-known scenarios elicit a more difficult explanation process, consequently, leading to high ratings of surprise. Known scenarios with partial explanations in the setting furnish explanations more readily for the outcome, whereas less-known scenarios lacking partial explanations leave more cognitive work to be done in retrieval, inference and explanation-building for the outcome, which leads to higher ratings of surprise.

6.2.2 Response times

The response times taken to perform surprise judgments provide further converging evidence for the predictions made (see Figure 5). A two-way ANOVA of response times revealed a main effect of Scenario-Type, $F(1,38) = 11.61$, $p = .002$, $\eta_p^2 = .23$, showing that participants took longer to make surprise judgments for less-known scenarios (M = 6351.55ms, SD = 2806.65ms) than for known scenarios (M = 5395.69ms, SD = 2331.23ms). However, there was no main effect of Partial-Explanation, $F(1,38) = 1.104$, $p = .30$, $\eta_p^2 = .03$, and no reliable interaction between Scenario-Type and Partial-Explanation, *F*(1,38) = 0.60, *p* = .44, η_p^2 = .02; although, on average, participants took longer to make surprise judgments when partial explanations were absent (*M* = 6275.76ms, *SD* = 2605.08ms) than when they were present (*M* = 5471.47ms, *SD* = 2221.73ms).

This result suggests that the time-course of surprise judgments is most influenced by Scenario-Type; "missing-wallet" type scenarios appear to cue relevant knowledge that is easily built into explanations compared to "boss-hugging" scenarios, speeding the judgement of surprise. In contrast, partial explanations in the setting appear to ease the explanation process and impact surprise judgments, but they do not significantly reduce the time taken to judge surprise (at least, relative to the impact of scenario types).

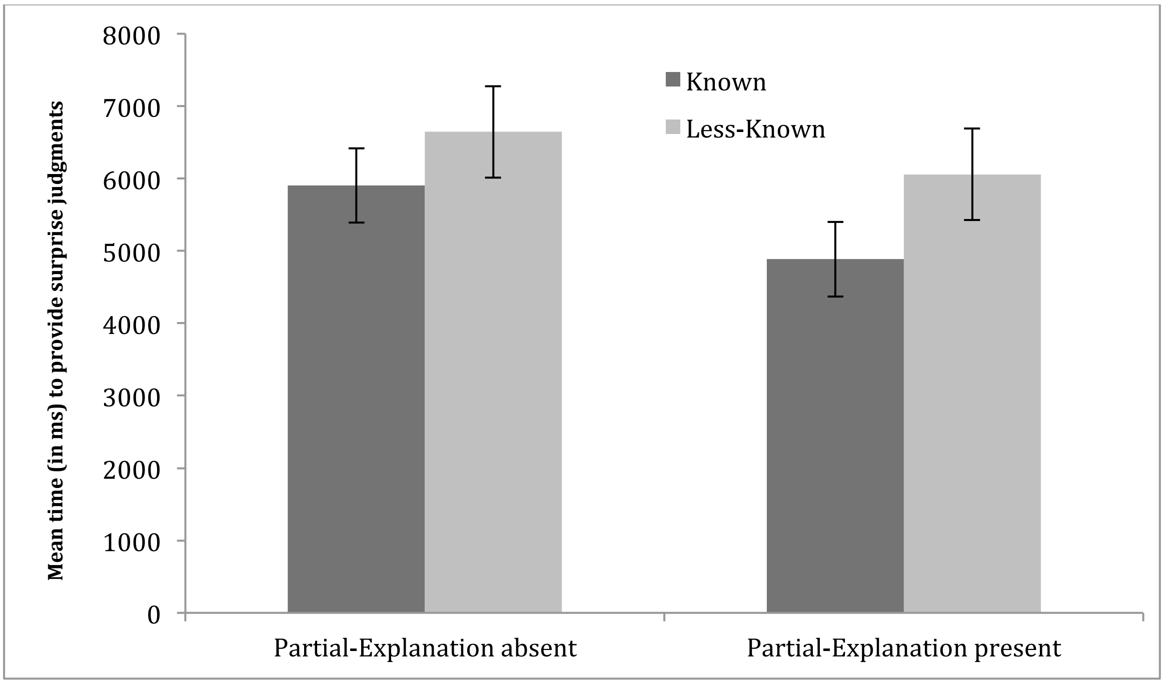


Figure 5. Mean response times in Experiment 2 for both levels of Outcome-Type (known vs. less-known) and Partial-Explanation (absent vs. present) with standard errors (*N* = 40).

6.2.3 Explanation productivity

MEB theory predicts that explanation productivity should increase for the Scenario-Type variable, with more explanations produced for known scenarios relative to less-known scenarios; known scenarios should cue larger amounts of relevant knowledge to support many different explanations, relative to less-known scenarios (see graphic in Figure 2, though this productivity is obviously mediated by success in explanation building). In contrast, the theory predicts productivity decreases for the Partial-Explanation variable, with lower productivity for scenarios where the partial explanation was present as opposed to absent. This effect occurs because the partial explanation further constrains the space of possible explanations, perhaps ruling out possible explanations that conflict with it. So, this promoted explanation should, pragmatically, block the production of other alternative explanations (even when people are instructed to produce such alternatives).

Prior to this data analysis, three participants (7.5% of the data) were discarded because they failed to follow the instructions given for the generation task. A two-way ANOVA on the productivity scores (i.e., number of different explanations generated by each participant) showed a main effect of Scenario-Type, $F(1, 35) = 4.00$, $p = .05$, $\eta_p^2 = .10$, as more explanations were produced for known scenarios ($M = 2.01$, $SD = .96$) than less-known scenarios ($M = 1.83$, $SD = .72$). A main effect of Partial-Explanation was also found, $F(1,35) = 6.33$, $p = .017$, $\eta_p^2 = .15$, showing that scenarios with partial explanations yielded fewer explanations ($M = 1.58$, $SD = .53$) than those without partial explanations ($M = 2.21$, $SD = .90$). There was no reliable interaction between these variables, $F(1,35) = 0.63$, $p = .44$, $\eta_p^2 = .02$. So, known scenarios increase the productivity of explanations produced, whereas the presence of partial explanation constrain productivity (see Figure 6).

Indeed, a more detailed analysis of the explanations produced provided some interesting evidence on the constraining role of partial explanations. In the partial-explanation-absent version of the Louise story, 35% of people generate the explanation that “she dropped the wallet before putting it into her handbag”, yet, in the partial-explanation-present version (which includes the sentence “her handbag was open”), 0% of people generate this “early-dropping” explanation, presumably because the partial explanation promotes the presupposition that the wallet was in the handbag. Hence, in this condition people suggest explanations involving the wallet being stolen from the handbag or involving the later-dropping of the wallet from the handbag.

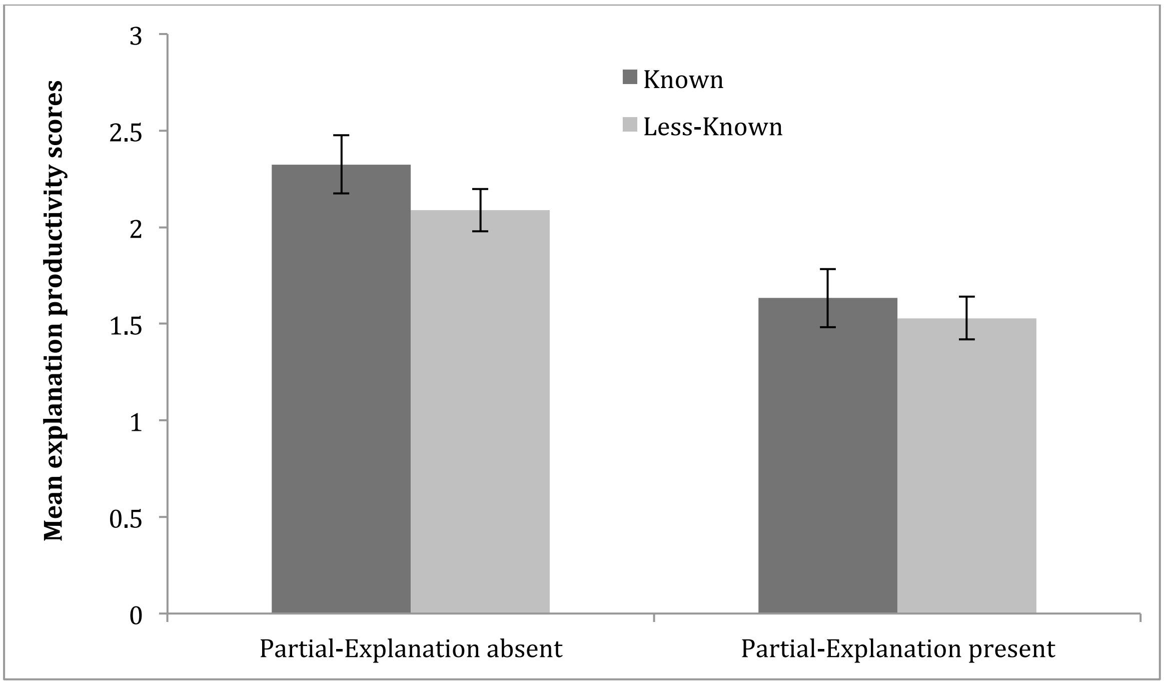


Figure 6. Mean productivity scores in Experiment 2 for both levels of Outcome-Type (known vs. less-known) and Partial-explanation (absent vs. present) with standard errors (*N* = 37).

6.3 Probability accounts of these results

We have seen how the MEB theory led to the predictions confirmed in this experiment for these variables, but how would probability accounts deal with these findings? Arguably, most probability theories can account for the effect of partial explanations on surprise judgments. If surprising events are low-probability events, then supplying partial explanations must raise the probability of the outcome, making it less surprising, as it is no longer such a low-probability event. Indeed, Maguire et al. (2011) confirmed that additional enabling conditions altered the subjective probabilities of outcomes in a way that predicted surprise ratings.

Similarly, taking a Bayesian approach, it could be posited that when an enabling condition is added to the setting (as a partial explanation) the prior probabilities of the previous setting information and background knowledge are updated, leading to new probability estimates for the outcomes in these scenarios. This extension of the probability approach looks like a plausible way to handle the partial explanation finding on surprise judgments. However, this finding is only one of several found in this experiment. The other findings are somewhat harder to accommodate.

The conclusions made at the end of Experiment 1, still apply to the Scenario-Type effects found here; namely, that a probability account is not as predictively adequate as it should be. Similarly, because probability accounts tend not to be performance theories (Jones & Love, 2011), they are silent on the findings about the specific time-course of processing surprise judgments or, indeed, the productivity of explanations. So, it is hard to escape the conclusion that the best probability accounts envisaged for these results provide limited coverage of the results found.

Table 3

Sample scenario used in Experiment 3 (only the outcome label was shown in presented materials).

Setting	Rebecca is on the beach. She goes for a swim in the water	
Outcome	**Known** After she dries herself off she notices that her skin has turned red.	**Less-Known** After she dries herself off she notices that her skin has turned turquoise

7. Experiment 3: Effects of Outcome-Type and Task on surprise judgments and explanations

In Experiments 1 and 2 we found evidence for the Scenario-Type factor on surprise judgments, the timing of judgments and post-judgment explanation productivity. In this experiment, first, we examined a finer-grained manipulation of the scenarios by holding the setting information constant and only changing the outcome. This Outcome-Type manipulation is akin to our intuitive example in which "losing one's wallet" appears to be a known outcome, whereas "losing one's belt" appears to be a less-known outcome, in terms of the knowledge they elicit. Obviously, the first prediction from MEB theory is that known-outcomes will be judged to be less surprising than their matched less-known outcomes. Second, rather than asking for explanations *after* the surprise judgments (as in Experiment 2), half of the participants were asked to explain each outcome as it was presented *before* making their surprise judgment. This Task variable involved asking one group to explain the outcome and another group to answer some comprehension questions about the scenario. It was expected that the instructions to explain the outcome would facilitaate the resolution of the outcome because it encourages explanation building, relative to the group that simply answered some comprehension questions about the scenario; so the former would be less surprised at the outcomes than the latter.

So, the experiment involved a 2 x 2 design with Task (explanation vs. comprehension) as a between-subjects variable and Outcome-Type (known vs. less-known) as a within-subjects variable. First, it was predicted that scenarios involving known surprising outcomes would be rated as less surprising than those with less-known surprising outcomes, for the reasons outlined in previous experiments. Second, it was predicted that the task demand to find an explanation would result in lower surprise judgments, relative to the task demand of answering comprehension questions on the same stories. An interaction between these two variables was not predicted, as the instructional manipulation should affect both outcome-type in the same way.

This experiment also developed a new, more objective way to operationally define known/less-known materials, and introduces a new agreement measure for generated explanations.

7.1 Method

7.1.1 Participants and design

Forty UCD students (12 male, 28 female) with a mean age of 21.2 years (*SD* = 2.07, range = 19- 29) took part voluntarily in this study. Informed consent was obtained prior to the experiment. Participants were randomly assigned to one of two conditions in a 2 (between-subjects; Task: explanation versus comprehension) x 2 (within-subjects; Outcome-Type: known versus less-known) mixed-measures design.

7.1.2 Materials

A material set was created consisting of simple story scenarios with outcomes that were either known or less-known (see Table 3). The type of outcome was operationally defined using (i) a pre-test sorting task by an independent group of participants and (ii) Latent Semantic Analysis (LSA) scores of coherence. For the sorting task, 20 story scenarios were presented in a pre-test to independent sorters (*N* = 10). The sorters were assigned to two groups: one group received half the scenarios with known surprising outcomes and the other half of the scenarios with less-known surprising outcomes, and the second group received the opposite. Each sorter saw only one outcome for each given scenario. They were asked to determine if a given scenario has an outcome that "*falls within the range of reasonable outcomes to the scenario*" (i.e., known surprising outcome) or whether it "*falls less within the range of reasonable outcomes to the scenario*" (i.e., less-known surprising outcome). Of the 20 stories, the raters consistently deemed 9 stories to have separable known and less-known surprising outcomes (Fleiss' kappa showed substantial agreement, κ = .68, Landis & Koch, 1977).

Given the closeness between known and less-known outcomes in these materials, we also expected to be able to operationalize them using a more objective operational definition by measuring the coherence of the overall scenario; that is, the known outcomes should be ones in which it is easier to establish conceptual coherence between the setting information and the outcome than for the less-known outcomes. So, we also checked the coherence of the known and less-known variants of these 9 stories by operationalizing coherence using LSA scores. In discourse research (cf. Graesser & McNamara, 2011), the explanatory coherence of texts is often operationalized by using LSA, where higher LSA scores indicate that one text is more coherent than another (Landauer & Dumais, 1996, 1997). Using LSA's term-to-term pairwise comparison on the text that differed between the selected 9 stories used in the experiment, the scenarios with the known outcomes scored higher (*M* = .62, *SD* = .2) than their matched counterparts with less-known outcomes (*M* = .53, *SD* = .21), a difference that was statistically reliable, *t*(8) = 3.08, *p* = .015.

Four material sets were created from these 9 materials (each with two levels, a known and less-known outcome; 18 variants of scenario in total). Each of these material sets comprised nine scenarios, with either four scenarios with known surprising outcomes and five with less-known surprising outcomes, or five scenarios with known surprising outcomes and four with less-known surprising outcomes. This Material-Set variable proved to have no effect on subsequent surprise judgments (*p* > .05), so these results are not reported in the following analyses.

Finally, the order of presentation of these stories was randomised for each participant. Stories were presented on separate pages of a booklet, with the scenario setting on the top of the page, followed by the outcome (known/less-known), the statement of the task (comprehension/explanation), and a 7-point scale on which to rate the surprisingness of the outcome (1: not surprising, to 7: very surprising).

7.1.3 Procedure and scoring

Participants were asked to read nine stories and to judge the surprisingness of their outcomes (see Table 3). For the Task variable, the participants in the explanation condition were asked to produce the first explanation they could think of for why the outcome may have occurred, before rating it for surprise; in the comprehension condition the participants were asked to answer two simple comprehension questions about the scenario before rating it for surprise. For each story, the first question in the comprehension condition was about the story setting, and the second question was about the outcome.

Prior to the experiment, we conducted a pre-test (*N* = 4) to verify that, on average, it did not take significantly longer to produce an explanation compared to answering the two short comprehension questions; time taken to do one task or the other was not reliably different ($t(2) = -1.414$, $p = .29$, explanation $M = 6.5$ minutes; comprehension $M = 7.5$ minutes).

Two measures were recorded: (i) the 7-point scale judgment of surprise, and (ii) the explanations produced by participants for each scenario in the explanation group. Prior to data analysis one participant (2.5% of the data) was discarded because that person failed to follow the instructions given.

7.2 Results and discussion

Overall, the results confirmed the predictions that Outcome-Type and Task both impact people's perceptions of surprise. The intuition that known outcomes are less surprising than less-known outcomes when the setting is held constant was confirmed, as was the prediction that instructions to explain the outcome would reduce the overall perception of surprise. So, for example, though both outcomes were deemed to be surprising, the lost-wallet type of scenario was found to be less surprising than the lost-belt type of scenario. No reliable interaction was found between the two variables.

7.2.1 Surprise judgments

A two-way ANOVA confirmed that participants judged stories with known outcomes ($M = 3.92$, $SD = 1.18$) to be less surprising than those with less-known outcomes ($M = 5.73$, $SD = 0.95$), $F(1,37) = 128.82$, $p < .001$, $\eta_p^2 = .78$ (see Figure 7). This Outcome-Type effect occurs because known outcomes direct the retrieval of more relevant concepts that are easily built into explanations for the outcome, lowering surprise judgments. In contrast, stories with less-known outcomes less relevant knowledge to be retrieved, knowledge that may be more difficult to adapt to build explanations; so this less-known outcome is harder to explain, resulting in relatively higher surprise judgments.

There was also a significant main effect of Task, $F(1, 37) = 10.18$, $p = .003$, $\eta_p^2 = .22$, indicating that the explanation group judged the outcomes to be less surprising ($M = 4.40$, $SD = 1.03$) than the comprehension group ($M = 5.27$, $SD = 0.62$). This effect presumably occurs because participants in the explanation group have less cognitive work to do than the comprehension group; the former have produced an explanation before the surprise judgment, but the latter have answered comprehension questions *and* attempted to construct an explanation before the surprise judgment. No interaction between the two variables was found, $F(1, 37) = 0.00$, $p = .99$, $\eta_p^2 < .001$.

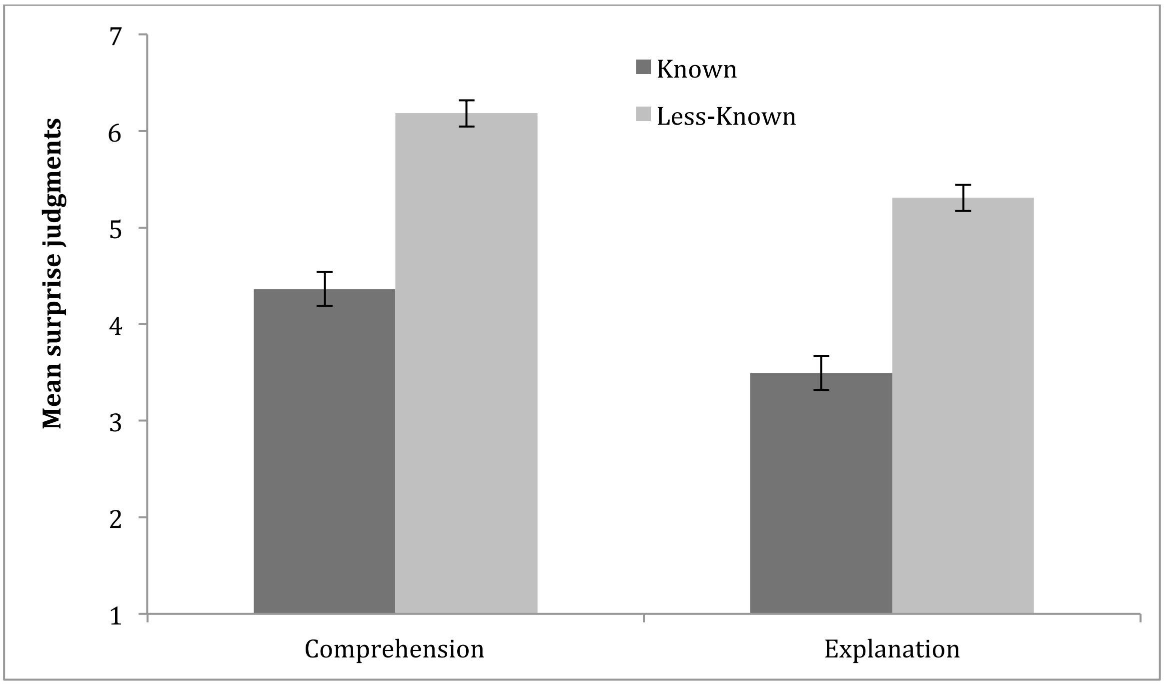


Figure 7. Mean surprise judgments in Experiment 3 for both levels of Outcome-Type (known vs. less-known) and Task (comprehension vs. explanation) with standard errors ($N = 59$).

7.2.2 Explanations

MEB theory proposes that known scenarios cue more relevant knowledge, in part because there is simply more domain knowledge about these scenarios. Additionally, resolving surprise should be largely based on normatively-packaged knowledge known to groups of people (if we follow Kahneman & Miller's (1986) suggestions). Taken together, these proposals suggest a new measure for people's explanations; namely, *proportion-of-agreement*. Proportion-of-agreement was determined by classifying all the explanations produced by a group and noting the proportion of times a given explanation was produced within the total set of explanations for each scenario. For example, in the Louise-handbag scenario, if the "robbery" explanation was produced by 10 of 20 participants in the experiment then it would be assigned a 0.50 proportion-of-agreement score. Having scored the explanations produced in the explanation conditions in this way, we conducted a paired t-test on the Outcome-Type variable, using these proportion-of-agreement scores as the dependent measure.

This t-test revealed a significant difference between the two levels of Outcome-Type, showing that there was greater agreement in the explanations produced to known outcomes ($M = .37$, $SD = .11$) than to less-known outcomes ($M = .23$, $SD = .06$), $t(19) = 4.51$, $p < 0.001$. Known outcomes appear to cue a shared set of "ready-made" concepts that are known to most people, whereas less-known outcomes elicit a more diverse set of non-standard explanations. This result thus provides indirect evidence for the differential nature of the knowledge brought to bear for known versus less-known scenarios.

7.3 Probability accounts of these results

How would probability accounts deal with these results? The Outcome-Type effects found here are a more fine-grained version of the Scenario-Type effects found in previous experiments. So, one could argue that these results are equally problematic for probability theories for the same reasons.

However, with respect to the Task effects, perhaps probability accounts could be extended to predict the lowering of surprise for participants in the explanation condition. The act of explaining could increase likelihood ratings, as has been shown to occur for explaining hypothetical outcomes (e.g., Koehler, 1991), and explaining an outcome has also been found to make people less likely to detect inconsistencies (Khemlani & Johnson-Laird, 2012). So, explanation could be argued to raise the subjective probability of a low-probability event, making it less surprising. From a Bayesian perspective, explanation could somehow change the posterior belief distributions to make them closer to prior belief distributions, perhaps because of the identification of new information that seems to increase the probability of the outcome occurring. However, even accepting that these theoretical extensions are *ad hoc* and not currently part of these probability theories, there are a few other problems in adopting them.

First, there is no evidence to suggest that our participants treat the outcome as a hypothetical possibility in either group, as it is presented as what actually occurs next in the scenario. Second, the outcome was presented before the participants were asked to explain it, so they should have been able to note possible inconsistencies in the outcome before coming up with an explanation. Additionally, as probability accounts tend not to be performance theories, they are silent on the findings about agreement between explanations. So, again, it is hard to escape the conclusion that probability accounts envisaged for these results fail to meet much of the evidence found.

Table 4

Examples of materials used in Experiment 4 (only the outcome label was shown in presented materials).

	None	Usual	Exceptional
Sentence 1	Lorna is in an ethnic restaurant.	Lorna is in her favourite ethnic restaurant that she has often gone to before.	Lorna is in a new ethnic restaurant that she has never gone to before.
Sentence 2	She has ordered her food and, after a while, the waiter brings it to her.		
Outcome	**Known** When she asks for a knife she is told that they have none.		**Less-Known** When she asks for a knife she is brought a banana.

8. Experiment 4: Effects of Outcome-Type, Task and Counterfactual-Hint on surprise judgments, response times and explanations

This experiment attempted to replicate the effects found for Outcome-Type and Task in Experiment 3, as well as introducing a novel manipulation to the setting information, designed to elicit counterfactuals. Recall, Tversky and Kahneman (1973; Kahneman & Miller, 1986) proposed that "abnormal events" (our "surprising outcomes") will seem more abnormal if contrasting counterfactual alternatives are highly available; that is, the abnormal event (i.e., losing your wallet) will appear more abnormal if the contrasting counterfactual is highly available (e.g., the normal event of "having your wallet"). Kahneman and Miller also proposed that "highly available alternatives are attributed greater causal effectiveness than equally potent but less mutable factors" (p. 144). That is, the availability of a normal event (the counterfactual) can provide an explanation for the abnormal event (the factual one), as people often use the difference between the two events to find an explanation (e.g., "if the wallet isn't in the pocket of a new pair of jeans, maybe it's in my usual trousers"). So, the elicitation of such counterfactuals could reduce the perceived surprisingness of an outcome, as it could provide a "quick and easy" explanation of the surprising outcome. To use an example commonly given in the counterfactual literature, if you are told "Jack had a car crash when he did not take his usual route home", people naturally draw on the counterfactual scenario of Jack taking his usual route home to find an explanation (e.g., "the crash happened because he took the different route"). However, this prediction assumes that the counterfactual-inspired explanation is always used, which may not be a given; for instance, it may have to compete with other explanations that come to mind by non-counterfactual means.

The literature on counterfactuals largely confirms that they tend to be elicited for actions rather than inactions, controllable events rather than events that are outside the actor's control, and for non-routine, exceptional events rather than usual events (e.g., Byrne, 2002; Wells & Gavanski, 1989; Kahneman & Miller, 1986; Kahneman & Tversky, 1982). Here we focus on the possible influence of non-routine events. Returning to our "wallet-losing" example in more detail, we feel that a scenario in which "John arrived at the office, wearing his brand new suit, and found that his wallet was not in the pocket" may prompt an explanation that the wallet is in his old suit, in a way that "John arrived at the office wearing his usual suit, and found that his wallet was not in the pocket" does not. However, although unexpected events have often been used in research on counterfactuals (e.g., McEleney & Byrne, 2006; Roese & Olson, 1996), surprising events have not been specifically studied. In addition to this, recent research suggests that Kahneman and Miller's original proposals may be somewhat more nuanced (see Byrne, 2005, 2007; Dixon & Byrne, 2011; Mandel, 2003; see also Results & Discussion). In this experiment, the scenarios used in Experiment 3 were modified to elicit counterfactuals by changing the setting to stress that the setting of the event was either routine (usual) or non-routine (exceptional) for the actor involved (see Table 4).

So, the design for this experiment manipulated Task (comprehension versus explanation), Outcome-Type (known versus less-known) and Counterfactual-Hint (none, usual or exceptional). Again, it was predicted that scenarios involving known surprising outcomes would be rated as less surprising than those with less-known surprising outcomes. Second, it was also predicted that the task demand to find an explanation would result in lower surprise judgments, relative to the task demand of answering comprehension questions on the same stories. For the Counterfactual-Hint variable, following Kahneman and Miller, if the elicitation of counterfactuals by stressing the exceptionality of the setting eases the explanation process, then surprise would decrease for these scenarios.

8.1 Method

8.1.1 Participants and design

Sixty UCD students (27 male, 33 female) with a mean age of 20.95 years (*SD* = 4.228, range = 18-44) took part voluntarily in this study. Informed consent was obtained prior to the experiment. Participants were randomly assigned to one of two conditions in a 2 (between-subjects; Task: comprehension versus explanation) x 2 (within-subjects; Outcome-Type: known versus less-known) x 3 (within-subjects; Counterfactual-Hint: none, usual, exceptional) mixed-measures design.

8.1.2 Procedure, materials and scoring

As in Experiment 3, participants were asked to read nine stories and to judge the surprisingness of their outcomes. Rather than asking participants how surprised they would be if this event occurred (as they were in Experiments 1- 3), they were asked instead to judge how surprised they would be by the event "if they were the character described". This change to the framing of the judgement task was introduced to see if it elicited different rating behaviour (we expected no change in responses from the previous instructions and, indeed, none were found in the final results).

For the Counterfactual-Hint variable, the event in the story setting either (i) gave no hint as to its routine nature (none), (ii) gave the explicit hint that the scenario event was regular or routine (usual), or (iii) gave the explicit hint that the scenario event was non-usual or non-routine (exceptional). For the Outcome-Type variable, the participants saw either a known or less-known surprising outcome for each story; only one outcome and one setting was seen by each participant for each story (see Table 4 for an example of the materials used). The LSA scores for the two new variants of the setting, usual and exceptional, showed no main effect of this Counterfactual-Hint variable ($p > .59$) when compared to the original scenarios containing no hints (i.e., the none version of the scenarios). Six material sets were created. Each of these comprised all 9 scenarios, with three variants of each setting type (none, usual, exceptional). Of these, either four scenarios were presented with known surprising outcomes and five with less-known surprising outcomes, or five scenarios with known surprising outcomes and four with less-known surprising outcomes. As this Material-Set variable proved to have no effect on subsequent surprise judgments ($p > 0.05$), it is not reported in the following analyses.

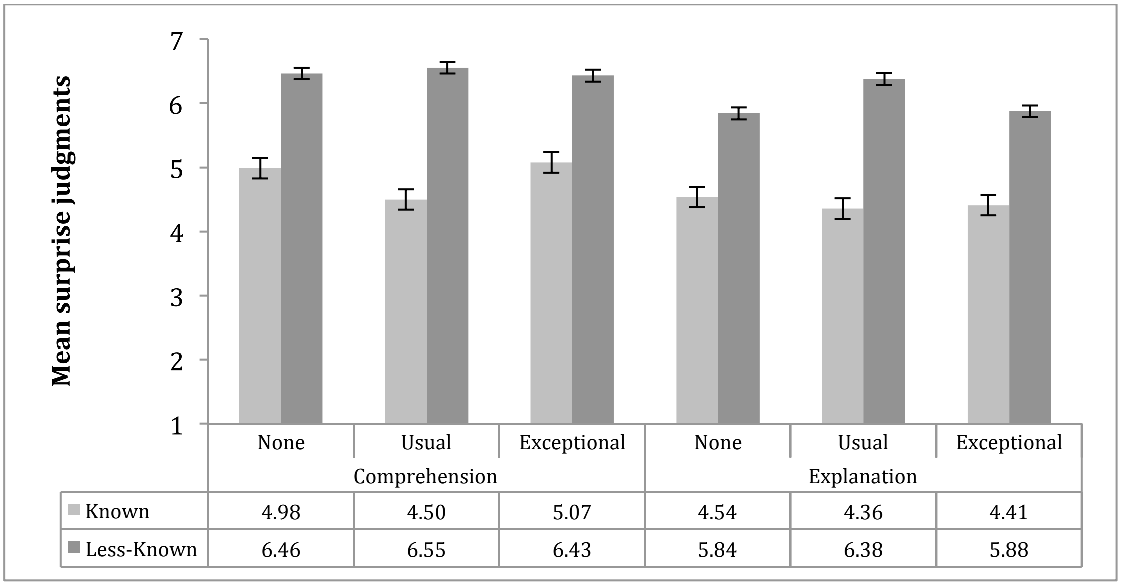
The order of presentation of these stories was randomised anew for each participant. Stories were presented sentence-by-sentence on a desktop computer-screen as participants pressed the spacebar, with each sentence appearing below the preceding one on the screen, until the outcome was presented. At this point, the participants in the explanation condition were instructed to "*type in the first explanation you can think of for why this outcome may have occurred*", while the participants in the comprehension condition sequentially saw and typed in the answer to two simple comprehension questions about the story. One of these questions was about the information provided in the setting, and the other was about information provided in the outcome. Neither of these questions drew the participants' attention to the Counterfactual-Hint variable, *per se*. Initially, the participants in this condition saw the first question and, after providing an answer, they pressed the return key, this first question disappeared and the second question appeared. After this explanation/comprehension step, all participants pressed the return key and the question "*If you were [character's name], how surprised would you be by this outcome*" appeared on the screen. On presentation of this question, participants indicated on a 7-point scale their surprise judgment (1: not surprising, to 7: very surprising). Three measures were recorded: (i) the 7-point judgment of surprise, (ii) the response time, and (iii) the explanations produced by each participant for each scenario.

8.2 Results and discussion

Overall, the results again confirmed the predictions that known surprising outcomes and the task demand of producing an explanation decrease the perception of surprise, however, there was scant evidence for a counterfactual effect. Prior to data analysis 4 participants (6.7% of the data) were discarded from further analysis because they failed to follow the instructions given (e.g., failing to provide explanations in the explanation condition or providing explanations in their answers in the comprehension condition).

8.2.1 Surprise judgments

A three-way ANOVA confirmed that participants judged known outcomes to be less surprising ($M = 4.51$, $SD = 1.11$) than less-known outcomes ($M = 6.21$, $SD = .75$), showing a main effect of Outcome-Type, $F(1, 54) = 92.46$, $p < .001$, $\eta_p^2 = .63$. Again, as in Experiment 3, a significant main effect of Task, $F\ (1, 54) = 4.65$, $p = .036$, $\eta_p^2 = .08$, was found, indicating that participants judged the outcomes of scenarios to be less surprising when they had provided explanations for them ($M = 5.09$, $SD = .85$), as opposed to answering comprehension questions ($M = 5.56$, $SD = .63$; see Figure 8). However, there was no main effect of Counterfactual-Hint, $F(2, 108) = .002$, $p > .05$, $\eta_p^2 < 0.001$, no reliable interaction between Outcome-Type and Counterfactual-Hint, *F*(2, 108) = 2.78, $p > .05$, $\eta_p^2 = .05$, and no other reliable interactions were found between the variables (all other *F*s < 1).



	None	Usual	Exceptional	None	Usual	Exceptional
	Comprehension			Explanation		
■ Known	4.98	4.50	5.07	4.54	4.36	4.41
■ Less-Known	6.46	6.55	6.43	5.84	6.38	5.88

Figure 8. Mean surprise judgments in Experiment 4 for all levels of Outcome-Type (known vs. less-known), Task (explanation vs. comprehension) and Counterfactual-Hint (none, usual, exceptional) with standard errors (*N* = 56).

8.2.2 Response times

The response times taken to provide a surprise judgment for each scenario (measured from when participants finished the explanation/comprehension task) were also analysed. A 2 (Outcome-Type) x 2 (Counterfactual-Hint; only the usual and exceptional conditions were included in this analysis as these scenarios were equally matched in length) x 2 (Task) mixed-measures ANOVA was performed on the response times. This analysis revealed no main effect of Outcome-Type, *F* (1, 54) = 2.968, $p = .09$, $\eta_p^2 = .05$, and no main effect of Counterfactual-Hint, *F* (1, 54) = .014, $p > .05$, $\eta_p^2 < .001$, or Task, *F* (1, 54) = 1.62, $p > .05$, $\eta_p^2 = .029$, and no interactions between the variables (all *F*s < 1), see Figure 9. So, an equivalent Outcome-Type effect on response time was not found here, akin to the Scenario-Type effect found in Experiment 2 (though this effect did approach significance at p = .09). This difference is probably due to the subtlety of the Outcome-Type manipulation relative to the Scenario-Type manipulation; the latter uses different sets of scenarios, whereas the former uses closely-matched scenarios that only different in one aspect of the outcome. So, the present result is perhaps not surprising, especially when one considers that these response times are in the order of seconds.

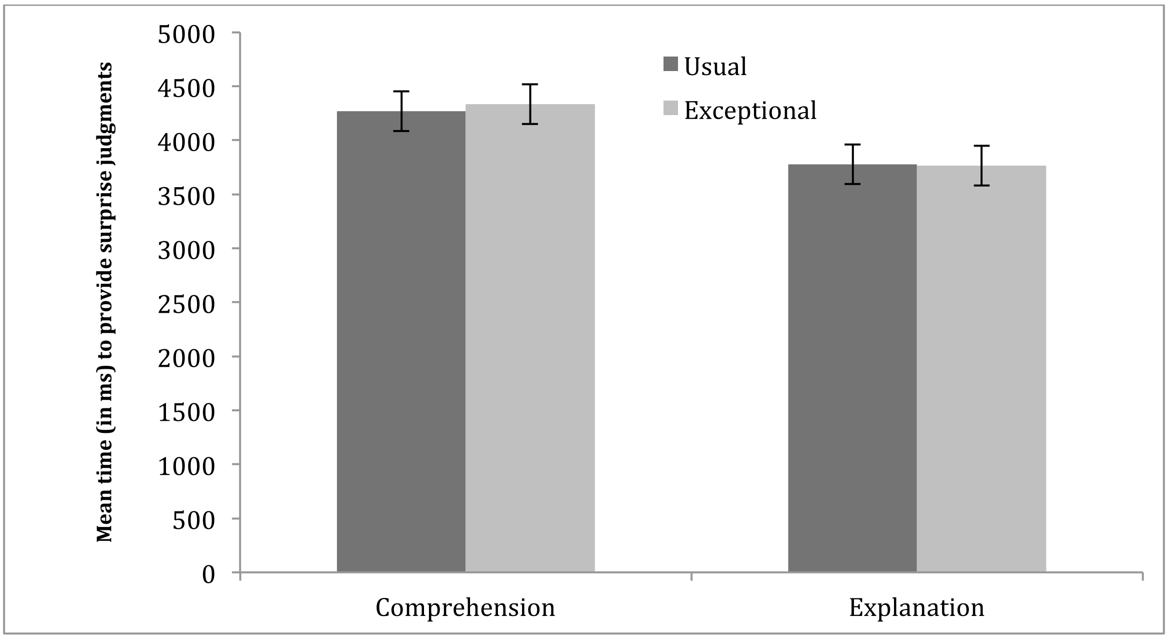


Figure 9. Mean response times for surprise judgments in Experiment 4 for both levels of Counterfactual-Hint (usual vs. exceptional) and Task (comprehension vs. explanation) with standard errors (N = 56).

8.2.3 Explanations

Participants' explanations in the explanation condition were recorded for each scenario and scored using the proportion-of-agreement measure described previously. These scored explanations were analysed in a 2 (Outcome-Type) x 3 (Counterfactual-Hint) repeated measures ANOVA, and revealed a main effect of Outcome-Type, showing greater agreement in the explanations produced for known outcomes (*M* = .38, *SD* = .08) over less-known outcomes (*M* = .23, *SD* = .11), *F*(1, 27) = 118.96, *p* < .001, η_p^2 = .82. There was no main effect of Counterfactual-Hint, and no interaction between these variables was found (both *F*s < 1). So, again, participants seem to have a greater degree of shared knowledge in the concepts retrieved for the explanation of known outcomes than they do for less-known outcomes, but the elicitation of counterfactuals does not seem to assist in directing the retrieval of concepts to be used for explaining the outcome.

8.3 Accounting for the role of counterfactuals in surprise

On the face of it, this experiment does not support the prediction that eliciting a counterfactual response, by indicating the normality/exceptionality of the setting, affects surprise. There are many possible reasons for why this prediction was not confirmed. Recent research on counterfactuals suggests that the optimality of the alternative action is important. Usually, exceptional actions lead to exceptional outcomes, and normal actions lead to normal outcomes; exceptional actions normally elicit counterfactual alternatives changing these to normal ones. However, if an exceptional setting leads to a better outcome, then the setting can be changed to an exceptional setting rather than to the “normal” one (cf. Dixon & Byrne, 2011). To use the car crash scenario as an example, rather than Jack taking his usual route home, to avoid the crash he could have left at a different time; altering the scenario to have a different exceptional setting rather than changing it to the “normal” one. So, it may be perceived, metacognitively, as just as difficult to work out the explanation from this scenario as to produce an explanation without a counterfactual hint, leading to similar levels of surprise. Additionally, Khemlani, Sussman, and Oppenheimer (2011) have found that people prefer explanations that have narrow latent scope; that is, explanations that account for only the phenomena that are observed. It is possible that, though explanations were prompted by the elicited counterfactuals, they were too broad and general to be considered "good" explanations. In fact, even if counterfactual reasoning was elicited and used, it may not prove to have had an effect on the surprise ratings, as factual reasoning may have an equal impact on judgments (e.g., Mandel, 2003) although no significantly greater agreement in explanations was found for the exceptional settings, so these findings do not suggest this conclusion.

8.5 Probability accounts of these results

As in the account of the previous experiment, there is a way that probability accounts could be extended to predict the lowering of surprise for participants in the explanation condition; the act of explaining could increase likelihood ratings for the outcome, changing posterior belief distributions somehow. Again, however, this only seems to be feasible if the participants in the comprehension group treat the outcome as a hypothetical possibility, which seems unlikely. The Outcome-Type effects found here prove to be problematic for probability theories for the reasons outlined earlier, and again, as probability accounts tend not to be performance theories, they are silent on the findings about agreement between explanations.

9. Experiment 5: Effects of explanation task on surprise judgments

Experiment 4 replicated the explanation effect first found in Experiment 3, where it was shown that people explaining an outcome, rather than answering comprehension questions, tended to perceive outcomes as less surprising. MEB explains this effect as being due to people's metacognitive sense of the cognitive work done to explain the surprising outcome; when that cognitive work is eased by accomplishing the explanatory work in advance, their perception of surprise decreases. In this experiment, instructions to produce multiple explanations are tests. It may be that case that producing multiple explanations is not be the same as producing a single explanation.

Indeed, the metacognitive aspect of MEB theory leads to a somewhat counterintuitive prediction when one considers this issue; namely, that if we make the explanation task more difficult by raising its cognitive load, it should be possible to reverse people's reduced surprise ratings, even though they are still being asked to explain the outcome. Hence, if we ask participants to produce either one or three explanations for a (known) surprising outcome, their metacognitive sense should be that more cognitive work is required for the latter over the former and, consequently, even though they are still explaining the occurrence of the outcome, they will perceive the latter outcomes to be more surprising than the former (this manipulation is analogous to one used in the hindsight bias literature; see, e.g., Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka & Simons, 1991). Furthermore, it should be stressed that this "number-of-explanations" effect should only hold for known surprising outcomes, not for less-known outcomes, because known outcomes have a ready supply of possible explanations, whereas less-known outcomes do not. For less-known outcomes the task will be quite difficult whether they are trying to produce one or three explanations.

So, this experiment involved a 2 x 2 design with Task (one explanation vs. three explanations) as a between-subjects variable and Outcome-Type (known vs. less-known) as a within-subjects variable. First, as before, we predict a main effect of Outcome-Type, as the known outcomes should be rated as less surprising than the less-known outcomes. Second, we predict an interaction between Task and Outcome-Type, as the requirement to produce three explanations should be perceived as more difficult and elevate surprise ratings for the known outcomes, while not markedly affecting the surprise ratings for the already difficult to explain less-known outcomes.

9.1 Method

9.1.1 Participants and design

Forty UCD students (18 male, 22 female) with a mean age of 20.9 years (*SD* = 1.92, range = 18-25) took part voluntarily in this study. Informed consent was obtained prior to the experiment. Participants were randomly assigned to one of two conditions in a 2 (between-subjects; Task: one explanation, three explanations) x 2 (within-subjects; Outcome-Type: known, less-known) mixed-measures design.

9.1.2 Materials, procedure and scoring

Eight scenarios were used with two variants of each (known and less-known) in the two different Task conditions (i.e., one material was dropped from the sets used in Experiments 3 and 4). Four material sets were created for this experiment. Each of these comprised all eight scenarios, each with four scenarios with known surprising outcomes and four with less-known surprising outcomes. As this Material-Set variable proved to have no effect on subsequent surprise judgments ($p > .05$), it is not reported in the following analyses.

As in previous studies, participants were tested individually, asked to read eight stories and to judge the surprisingness of their outcomes. The order of presentation of these stories was randomised for each participant. Stories were presented on separate pages of a booklet, which began with the instructions to generate explanations and rate each scenario for surprise. Each story was presented on a separate page with the scenario setting on the top of the page, followed by the outcome (known/less-known), the instructions to produce either one or three explanations for why this outcome occurred, and the scale on which to provide their surprise judgment. One measure was recorded: the 7-point scale judgment of surprise.

9.2 Results and discussion

Overall, the results again confirmed that Outcome-Type impacts people's perception of surprise. Although there was no reliable main effect of Task, there was a reliable interaction between Task and Outcome-Type as predicted; specifically, when participants were asked to complete the more difficult task of producing three explanations versus only one for the known outcomes, their surprise increased, although this was not the case for the already highly-surprising, less-known outcomes.

9.2.1 Explanation manipulation check

Firstly, a paired t-test showed that there was no significant difference in the number of explanations produced by participants in the three-explanation condition, between known (M=2.69, SD=.58) and less-known (M=2.65, SD=.55), $t(19)=.68$, $p = .51$. Most people were able to produce three explanations for both known and less-known outcomes.

9.2.2 Surprise judgments

A two-way ANOVA confirmed that participants judged stories with less-known outcomes to be more surprising ($M = 5.74$, $SD = .78$) than those stories that had known outcomes ($M = 3.46$, $SD = .78$), $F(1, 38) = 341.93$, $p < .001$, $\eta_p^2 = .9$, see Figure 10. Again, this effect appears to be because the known outcomes cue more relevant knowledge for easier explanation-building than less-known outcomes, and so are less surprising. The less-known outcomes cue less relevant knowledge, hence, are more difficult to explain, and so are rated as more surprising.

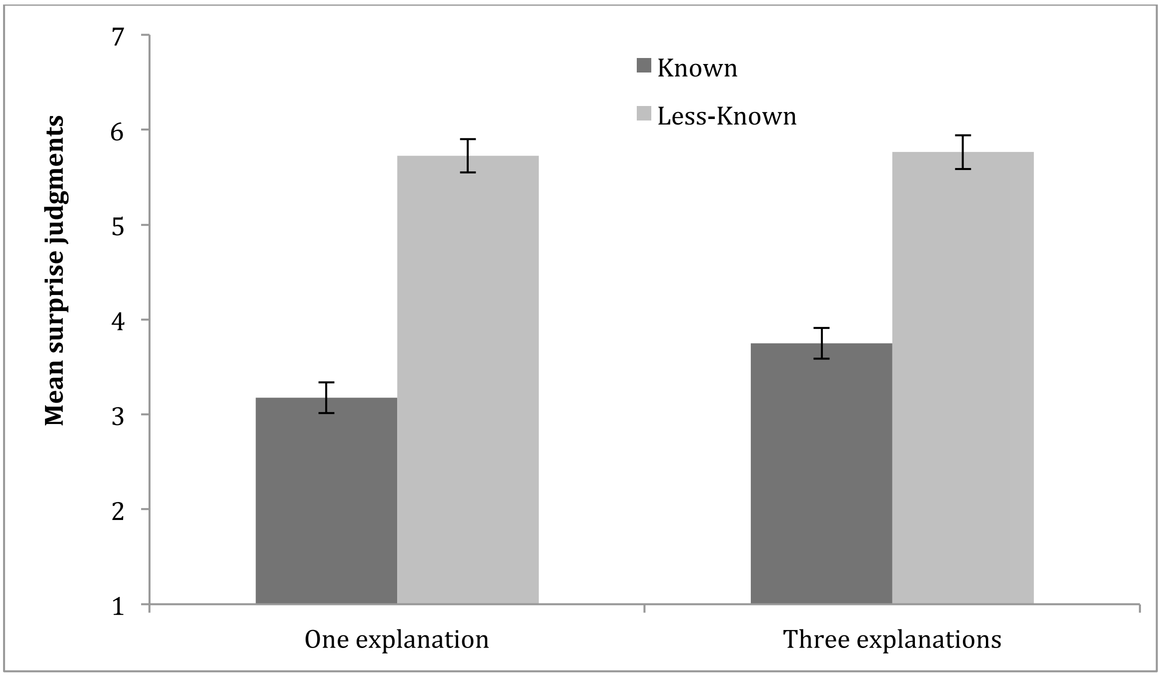


Figure 10. Mean surprise ratings in Experiment 5 for both levels of Outcome-Type (known vs. less-known) and Task (one vs. three explanations) with standard errors (*N* = 40).

There was no main effect of Task, *F*(1, 38) = 2.18, *p* = .15, η_p^2 = .05. However, the predicted interaction was found between Task and Outcome-Type, *F*(1, 38) = 4.75, *p* = .036, η_p^2 = .11. For known surprising outcomes, when participants were asked to produce three explanations their surprise ratings were higher (*M* = 3.75, *SD* = .91) than when they had to produce only one explanation (*M* = 3.17, *SD* = .5); post-hoc tests showed that this difference was statistically significant, *p* = .019. But, for less-known surprising outcomes, there was no difference in surprise ratings when participants were asked to produce three explanations (*M* = 5.76, *SD* = .8) or one explanation (*M* = 5.72, *SD* = .78), *p* >.05. The latter pattern has all the hallmarks of a ceiling effect for a metacognitive sense of the cognitive work done (recall, in previous experiments, less-known outcomes have not risen higher than M=6.21 out of 7, *SD* = 0.75). Less-known outcomes are already perceived as difficult to explain (whether it be one or three explanations) and are rated as highly surprising in both conditions. Known outcomes are initially easier to explain; producing more explanations increases the perceived difficulty of finding explanations, which is then reflected in surprise judgments.

9.2.3 Probability accounts of these results

The number of explanations effect presents some difficulties for a probabilistic perspective, as there seems to be nothing in a probability account that would suggest coming up with more explanations for an event should change the likelihoods in some systematic way, and, in particular, nothing that would suggest the interaction found here. The idea described above that explaining could affect likelihood ratings, would seem to suggest either (i) that no effect would be seen for the "number-of-explanations" task for both type of outcome, or (ii) that the same effect should be seen for both, as each group of participants are explaining the outcome.

10. Experiment 6a: Effect of cueing/miscueing on explanation productivity

Most of the previous experiments have examined how scenarios can act to cue relevant knowledge, easing explanation and reducing surprise (see Experiments 1-5). At the outset, MEB theory also proposed that knowledge could be *miscued* from the given information, but this proposal has not been directly tested. In miscueing, parts of the scenario could direct the retrieval and explanation processes to a "less-productive" region of activated knowledge, impeding the explanation of the outcome and increasing surprise. Such miscues should impact both high-knowledge (known) and low-knowledge (less-known) cases, though perhaps their impact on low-knowledge cases should be slightly less, as *a priori* there is less knowledge to be miscued anyway (i.e., something like a floor effect could occur). In the next three experiments (6a, 6b and 6c), this miscueing effect is tested in a novel paradigm, where the same materials are used but the measures are varied from explanation productivity

(Experiment 6a) to surprise judgments (Experiment 6b) to subjective probability estimates (Experiment 6c). So, taken together, the three experiments are designed to provide converging evidence on the factors affecting surprise when cues/miscues keywords are present. Specifically, Experiment 6c permits us to address directly the adequacy of probability accounts of surprise.

The novel paradigm deployed in this series of experiments made use of keywords presented with the settings of the scenarios. As in previous experiments, the outcomes were varied (known or less-known) but the settings were also augmented with added keywords, in the form of a supportive keyword (cue) or unsupportive keyword (miscue). The selection of supportive/unsupportive keywords was based on a content analysis of the explanations people had previously produced to the surprising outcomes. Table 5 shows one sample scenario from the experiment. For this known outcome scenario, the cueing keyword ("copycat") supports explanation of the outcome ("meeting a neighbour on holidays") using the idea that the neighbour was someone who was copying Gary's holiday choice. In contrast, for this scenario, the miscueing keyword ("circus") does not support obvious explanations of the outcome. For the less-known outcome for this scenario, the cueing keyword ("circus") supports explanation of the outcome ("seeing a rhino in a cafe") using the idea that the animal must have escaped from the circus. In contrast, for this scenario, the miscueing keyword ("copycat") does not support obvious explanations of the outcome.

In this first experiment, in the series of three, the measure used was explanation productivity; that is, whether the manipulations affected the number of explanations people produced to the scenario. The predictions were that (i) scenarios with known outcomes will direct the retrieval process towards relevant concepts in memory to support explanation-building relative to scenarios with less-known outcomes (as seen in Experiments 3-5), (ii) cueing keywords will help direct this retrieval process, increasing the number of explanations produced relative to when miscueing keywords are presented, and (iii) an interaction between the variables was possible, in the form of a greater impact of cueing on the known-scenarios than the less-known (where knowledge is low anyway).

Table 5

Sample scenario used in Experiments 6a, 6b and 6c (only the outcome label was shown in presented materials).

		Known	**Less-Known**
Setting		Gary is on holidays in a small village in West France. He is sitting in a café near his hotel.	
Keyword	***Cue***	copycat	circus
	Miscue	circus	copycat
Outcome		He looks up and sees his next-door neighbour walk by.	He looks up and sees a rhino charging through the cafe.

10.1 Method

10.1.1 Participants and design

Thirty-two UCD students (15 male, 17 female) with a mean age of 22.03 years (*SD* = 3.12, range = 19-31) took part voluntarily in this study. Informed consent was obtained prior to the experiment. A 2 (within-subjects; Keyword: cue, miscue) x 2 (within-subjects; Outcome-Type: known, less-known) repeated-measures design was used.

10.1.2 Materials, procedure and scoring

Participants saw the same eight scenarios used previously, with two variants of each outcome (known and less-known) and two variants of each keyword (the cueing keyword used for each known outcome-type was the miscueing one used for the less-known outcome, and vice versa). Four material sets were created for this experiment. Each of these comprised all eight scenarios; four scenarios with known surprising outcomes, two with cueing keywords and two with miscueing keywords, and four with less-known surprising outcomes, again two with cueing keywords and two with miscueing keywords. As this Material-Set variable proved to have no effect on subsequent explanation productivity ($p >$ .05), it is not reported in the following analyses.

Participants were each given a booklet containing eight stories that they had to read. For every scenario they were asked to "*write down as many explanations as you can, using the given keyword, for why this outcome may have occurred*". The order of presentation of these stories was randomised for each participant. Stories were presented on separate pages of the booklet, which began with the instructions to generate explanations using the given key words. Each story was presented on a separate page with the scenario setting on the top of the page, followed by the keyword and then the outcome (known/less-known). One measure was recorded: the number of explanations produced for each scenario.

10.2 Results & discussion

A repeated measures ANOVA showed a main effect of Outcome-Type, $F(1, 31) = 7.702$, $p = .009$, $\eta_p^2 = .199$, where participants produced more explanations to known outcomes ($M = 1.73$, $SD = .599$) than to less-known outcomes ($M = 1.48$, $SD = .49$), and a main effect of Keyword, F (1, 31) = 4.237, $p = .048$, $\eta_p^2 = .120$, where participants produced more explanations when provided with cueing keywords ($M = 1.73$, $SD = .715$) than miscueing keywords (*M* = 1.48, *SD* = .45), see Figure 11. There was no reliable interaction between the two variables, *F*(1, 31) = 1.947, *p* = .17, η_p^2 = .059. Participants produced the most explanations for known outcomes with cueing keywords (*M* = 1.92, *SD* = .92), and the least to less-known outcomes with miscueing keywords (*M* = 1.41, *SD* = .53), with the other two conditions falling in-between (known, miscue: *M* = 1.55, *SD* = .5, less-known, cue: *M* = 1.55, *SD* = .73).

These results re-confirm that Outcome-Type affects explanation productivity (as we saw in Experiment x) and, in addition, shows that the Keyword manipulation also affects productivity; with cueing keywords resulting higher explanation productivity than miscuing keywords. This productivity difference should, of course, be reflected in more easeful processing of the cued scenarios, involving less cognitive work, and decreases in percieved surprise. OK?

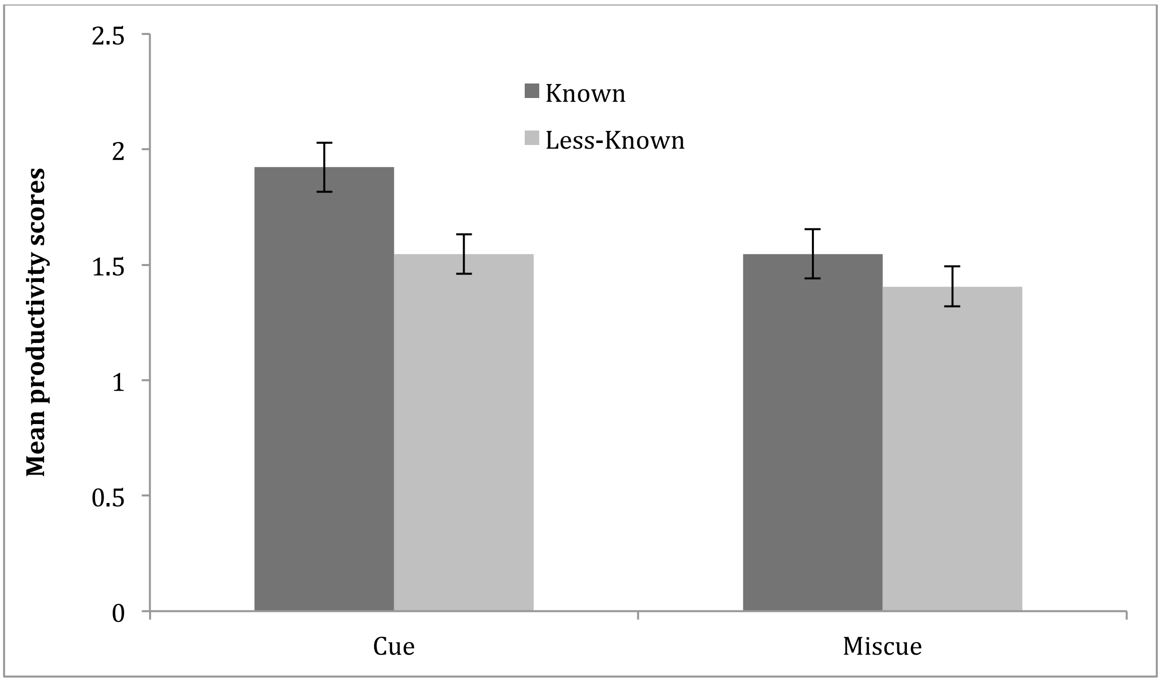


Figure 11. Mean productivity scores in Experiment 6a for both levels of Outcome-Type (known vs. less-known) and Keyword (cue vs. miscue) with standard errors (*N* = 32).

11. Experiment 6b: Effect of cueing/miscueing on surprise judgments

This experiment uses the same design and materials as Experiment 6a, though the measure was varied to be surprise ratings and the response times for these ratings (N.B., this experiment was run in a computerised form so that the response time measure could be collected). The MEB theory predicts that (i) scenarios with known outcomes will result in lower surprise ratings than less-known outcomes because the resolution of the surprising outcome is eased, and (ii) outcomes with cueing keywords should be rated as less surprising than those with miscueing keywords, as the former help direct the retrieval process to a "better" region of activated knowledge for explanation purposes than the latter (as evidenced in Experiment 6a), and (iii) there will be an interaction between these two variables, with the greatest impact of cueing/miscueing on known scenarios than less-known ones, as the latter already suffer from reduced amounts of available knowledge. It was also expected that the response time measure for surprise ratings should provide convergent evidence of this negative impact of miscueing on the known-outcome scenarios.

11.1 Method

11.1.1 Participants and design

Thirty-two UCD students (10 male, 22 female) with a mean age of 19.59 years (*SD* = 1.6, range = 18-27) took part voluntarily in this study. Informed consent was obtained prior to the experiment. A 2 (within-subjects; Keyword: cue, miscue) x 2 (within-subjects; Outcome-Type: known, less-known) repeated-measures design was used.

11.1.2 Materials, procedure and scoring

The materials used were the same as in Experiment 6a (see Table 5 for an example scenario). Participants saw eight scenarios, four scenarios with known surprising outcomes, two with cueing keywords and two with miscueing keywords, and four with less-known surprising outcomes, again two with cueing keywords and two with miscueing keywords. As this Material-Set variable proved to have no effect on subsequent surprise ratings or response times ($p > .05$), it is not reported in the following analyses.

Participants were tested individually on a laptop computer, and asked to read the eight stories, which appeared sentence by sentence as they pressed the spacebar. The order of presentation of these stories was randomised for each participant. Each story was presented on a separate screen with the scenario setting on the top of the page, followed by the key word and then the outcome (known/less-known). Importantly, participants were not asked to explain this outcome, but instead the presence of the keyword was accounted for in the instructions provided at the start as follows: "*Between the setting of the story and the final outcome sentence, a key word will be presented. This word may help you understand why the outcome occurred.*" For each scenario, participants rated the surprisingness of the outcome by pressing number keys on the keyboard, using a scale from 1 (not surprising) to 7 (very surprising). In addition to this surprise rating, response times taken to provide surprise judgments from the presentation of each outcome sentence were recorded.

11.2 Results and discussion

As predicted by MEB theory, scenarios with known outcomes were given lower surprise ratings than less-known outcomes, and outcomes with cueing keywords were rated as less surprising than those with miscueing keywords. Additionally, there was an interaction between these two variables, where miscueing increased surprise ratings for known scenarios but not for less-known scenarios. This effect on known scenarios was also reflected in response times, which were longer for known scenarios when the miscueing keyword was provided, than the cueing keyword.

11.2.1 Surprise ratings

A repeated measures ANOVA showed a main effect of Outcome-Type, $F(1, 31) = 88.44$, $p < 0.001$, $\eta_p^2 = .74$, where participants rated known outcomes as less surprising ($M = 3.84$, $SD = .97$) than less-known outcomes ($M = 6.05$, $SD = .93$), and a main effect of Keyword, $F(1, 31) = 16.89$, $p < .001$, $\eta_p^2 = .35$, in which participants gave lower surprise ratings to scenarios with cueing keywords ($M = 4.52$, $SD = .96$) as opposed to miscueing keywords ($M = 5.37$, $SD = .83$). There was also a reliable interaction between these variables, $F(1, 31) = 4.716$, $p = .038$, $\eta_p^2 = .13$, as predicted; there was a larger effect of Keyword on the known outcome-types (cue; $M = 3.2$, $SD = 1.36$, miscue; $M = 4.47$, $SD = 1.47$) than for the less-known outcome-types (cue; $M = 5.83$, $SD = 1.13$, miscue; $M = 6.28$, $SD = .91$), see Figure 12. These results re-confirm the earlier findings that known outcomes are less surprising than less-known outcomes; in addition, it shows that miscueing keywords have their greatest impact on known scenarios, as they inhibit the normally easy explanation process, raising perceived surprise, presumbly by “sending people the wrong way”. The impact of miscueing on less-known scenarios is attenuated by the fact that they are hard to explain anyway.

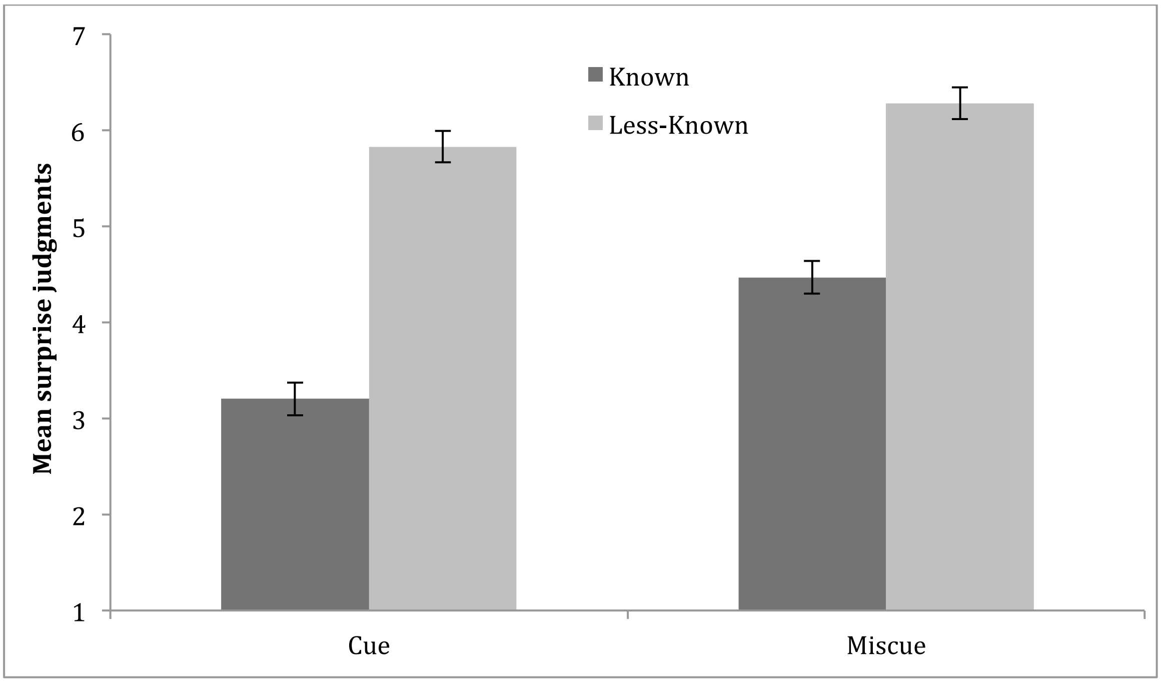


Figure 12. Mean surprise judgments in Experiment 6b for both levels of Outcome-Type (known vs. less-known) and Keyword (cue vs. miscue) with standard errors (N = 32).

11.2.2 Response times

A repeated measures ANOVA on the response times (using the time from when the outcome sentence was presented to when the participants provided a response) showed no main effect of Outcome-Type, $F(1, 31) = 2.8$, $p > .05$, $\eta_p^2 = .08$, or Keyword, $F(1, 31) = 1.94$, $p > .05$, $\eta_p^2 = .06$, but a marginally significant interaction between the two variables was found, $F(1, 31) = 4.1$, $p = .05$, $\eta_p^2 = .12$. This interaction reflects the pattern found for surprise ratings, with Keyword having a greater impact for known outcomes (cueing; M = 5089.62ms, SD = 2108.67ms, miscueing; M = 6081.48ms, SD = 3294.21ms) than for less-known outcome types (cueing; M = 4970ms, SD = 1711.5, miscueing; M = 4949.17ms, SD = 1996.73), see Figure 13.

We have already seen that response time is a fairly crude measure of the explanation process in resolving surprise (Experiment 2 found positive evidence, while analysis of response times in Experiment 4 showed that the main effect only approached significance). The present findings are consistent with what was found before; Outcome-Type impacts surprise judgments but this impact is not completely reflected in response times, perhaps because of the small changes made in materials to make this manipulation (see Table 5). However, other variables like Scenario-Type (where the whole story is changed), and Keyword when interacting with Outcome-Type, (in which participants are misdirected to less relevant knowledge) impact both surprise ratings and response times. This pattern of findings is consistent with the MEB theory, though the course-grained nature of the response-time measure should be recognised.

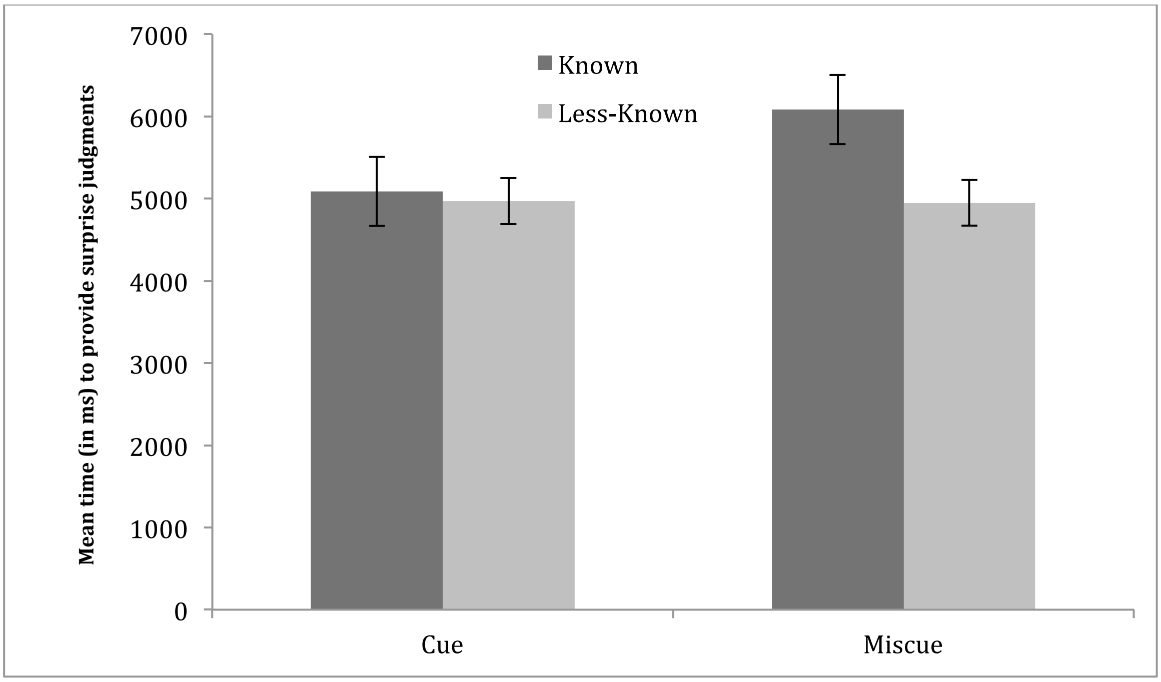


Figure 13. Mean response times for outcomes in Experiment 6b for both levels of Outcome-Type (known vs. less-known) and Keyword (cue vs. miscue) with standard errors (*N* = 32).

11.2.3 Comparison of findings from Experiment 6a and 6b

Even when the participants were not asked explicitly to explain why the scenarios occurred, easing the explanation process by providing cueing keywords decreased surprise relative to providing miscueing keywords. Also, as predicted above, increasing the difficulty of the explanation process by miscueing had a greater effect on surprise judgments for the normally-easier-to-explain known outcomes, than the already-difficult-to-explain less-known outcomes, for metacognitive reasons. THIS NEEDS TO BE BETTER; SAY THAT THE PRODUCTIVITY MEASURE SHOWS SAME PATTERN AS SURPRISE RATINGS. THAT CUES IMPACT BOTH IN THE SAME WAY.

12. Experiment 6c: Effect of cueing/miscueing on probability ratings

This experiment uses the same design and materials as Experiments 6a and 6b though the measure was varied to be probability ratings (rather than surprise judgments or explanation productivity). Recall, that in Experiment 1, we cast some doubt on the validity of probability estimates as predictors of surprise ratings for the story scenarios used. In regression analyses, Maguire et al. (2011) found that probability estimates of likely outcomes or contrasting probabilities did not predict for patterns of surprise ratings. However, they did find evidence of a weaker predictor; namely, that after-the-fact probability estimates of surprising-outcomes were correlated with the pattern of surprise ratings. This is a weaker predictor because these estimates may just reflect how the outcome is assessed after the surprise is resolved, rather than any prior estimate of the outcome's probability. However, as this measure has been shown to provide the best predictive validity for a probabilistic account, it was used in the present experiment. The key question to be answered by this experiment is whether this probabilistic measure retains any predictive validity for the present scenarios.

12.1 Method

12.1.1 Participants and design

Thirty-two UCD students (13 male, 19 female) with a mean age of 24.69 years (*SD* = 3.257, range = 21-36) took part voluntarily in this study. Informed consent was obtained prior to the experiment. A 2 (within-subjects; Keyword: cue, miscue) x 2 (within-subjects; Outcome-Type: known, less-known) repeated-measures design was used.

12.1.2 Materials, procedure and scoring

The materials used were the same as in Experiment 6a and 6b (see Table 5 for an example scenario). Participants saw all eight scenarios, four scenarios with known surprising outcomes, two with cueing keywords and two with miscueing keywords, and four with less-known surprising outcomes, again two with cueing keywords and two with miscueing keywords. As the Material-Set variable (the same four material sets were used as in Experiment 6a and 6b) proved to have a main effect on probability ratings ($p < .05$), it is reported in the following analyses.

Participants were each given a booklet containing eight stories that they had to read. Again, participants were not asked to explain the outcome; the presence of the keyword was accounted for in the instructions as in Experiment 6b. For every scenario they were asked to "*provide a number from 0% (no chance) to 100% (certain) for what you think the probability of this outcome occurring is*". The order of presentation of these stories was randomised for each participant. Stories were presented on separate pages of the booklet, which began with the instructions to provide probability ratings for the outcome of each story. Each story was presented on a separate page with the scenario setting on the top of the page, followed by the keyword (cue/miscue) and then the outcome (known/less-known). One measure was recorded: probability ratings (0-100).

12.2 Results and discussion

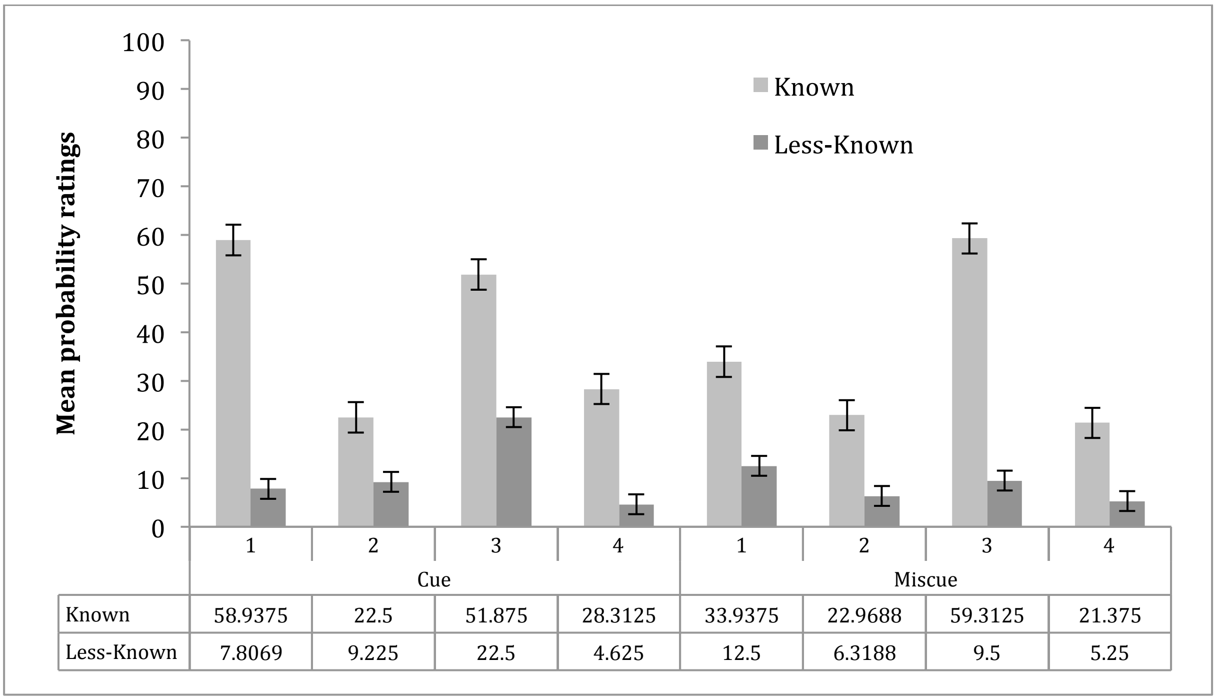
Overall, a different pattern of findings were seen in the probability ratings than for the surprise judgments seen in Experiment 6b, or for the explanation productivity in Experiment 6a. More specifically, although a main effect of Outcome-Type was found, there were reliable interactions between Outcome-Type and Material-Set, and a three-way interaction between Material-Set, Outcome-Type, and Keyword. There was no main effect of Keyword.

12.2.1 Probability ratings

A mixed measures ANOVA showed a main effect of Material-Set, $F(3, 28) = 6.37$, $p = .002$, $\eta_p^2 = .41$, and a main effect of Outcome-Type, $F(1, 28) = 66.96$, $p < .001$, $\eta_p^2 = .705$, where participants rated known outcomes as more probable ($M = 37.40$, $SD = 22.02$) than less-known outcomes ($M = 9.72$, $SD = .11.81$). There was no main effect of Keyword, and no reliable two-way interaction between Outcome-Type and Keyword, or Material-Set and Keyword (*p*’s > .05). There was a significant two-way interaction between Material-Set and Outcome-Type, $F(3, 28) = 3.19$, $p = .039$, $\eta_p^2 = .255$, and a three-way interaction between Material-Set, Outcome-Type, and Keyword, $F(3, 28) = 3.49$, $p = .029$, $\eta_p^2 = .272$, see Figure 14.

Exploring this three-way interaction, post-hoc pairwise comparisons (using Bonferroni correction for multiple comparisons) showed that, in Material Set 1, there was a significant difference between known outcomes with cues and less-known outcomes with cues, and known outcomes with cues and less-known outcomes with miscues, all other comparisons for Material Set 1 did not reach significance. In Material Set 3, there was a significant difference between known outcomes with miscues and less-known outcomes with miscues, and known outcomes with cues and less-known outcomes with miscues, but no other significant differences. Post-hoc pairwise comparisons for Material Set 2 and 4 showed no significant differences.

To explore the pattern of results in this experiment and Experiment 6b (on surprise judgments) correlations and regression analyses were performed. Although a by-participants correlation could not be run as different participants completed each experiment, a by-materials Pearson product-moment correlation coefficient and regression analyses were computed to assess the relationship between probability ratings and surprise judgments. For this, we set aside the Material-Set variable, to simplify the findings and present them in the most positive light; to find any possible relationship between the probability ratings and surprise judgments. Looking at each subset of the key interaction between the Outcome-Type and Keyword variables, a negative correlation was found between probability ratings and surprise judgments for known outcomes with miscues ($r(6) = -.83$, $p = .011$), but no relationship was found for less-known outcomes with cues or less-known outcomes with miscues, or known outcomes with cues. A linear regression was conducted to determine if these probability ratings could be used to predict surprise judgments; the probability ratings for known outcomes with miscues significantly predicted surprise judgments for known outcomes with miscues, β= -.831, $t(6) = -3.66$, $p = .011$. Probability ratings for known outcomes with miscues also explained a significant proportion of variance in surprise ratings for known outcomes with miscues, R^2= .69, ($F(1, 6) = 13.43$, $p = .011$); however, probability ratings were not found to predict surprise judgments for less-known outcomes (with either cues or miscues), or for known outcomes with cues. So, these further analyses suggest that probability ratings are not a good predictor of surprise, except in limited cases.



	Cue				Miscue			
	1	2	3	4	1	2	3	4
Known	58.9375	22.5	51.875	28.3125	33.9375	22.9688	59.3125	21.375
Less-Known	7.8069	9.225	22.5	4.625	12.5	6.3188	9.5	5.25

Figure 14. Mean probability ratings for outcomes in Experiment 6b for all levels of Outcome-Type (known vs. less-known), Material-Set (1-4) and Keyword (cue vs. miscue) with standard errors (*N* = 32).

12.2.2 Comparison of findings from Experiment 6b and 6c

Any comparison of the results of Experiment 6b on surprise judgments and Experiment 6b on probability estimates would have to conclude that the pattern of probability ratings does not correspond to surprise judgments. At best, one could argue that (i) some material-sets (Material Sets 1 and 3) correspond to some of the surprise judgments found, and (ii) some particular manipulations (e.g., known outcomes with miscues) can be predicted by subjective probability estimates, but most cannot.

However, even this argument is not hugely convincing for two reasons. First, recall that this type of probability estimate is already known to be somewhat weak. It is an after-the-fact assessment that could be criticised as just reflecting the changes that have occurred to the outcome after the surprisingness of the outcome has been resolved. So, in fact, it may just reflect some aspect of surprise. Second, even overlooking this weakness, there is a troubling instability in these probability ratings that is not evidenced in surprise ratings; effects of Material-Set were not found for surprise judgments in any of the other experiments reported here. They only occur with respect to these probability estimates, indicating that whatever probability estimates reflect, that cognitive process is one that is modified by material-contexts, in a way that surprise is not.

12.3 Probability accounts of these results

So, how do probability accounts now stand with respect to the different results found in this series of experiments (i.e., Experiments 6a-6c)? First, a probabilistic account makes no predictions about the number-of-explanations effect found in Experiment 6a because probability theory is silent on such process aspects; probability theory does not address memory's impact. Second, we have already seen that the correspondence between probability estimates and surprise ratings is less than adequate. Third, the probability account has no ready account of the cue/miscue effects; cues could encourage hypothetical thinking and somehow change likelihoods, but *a priori* it is not clear why they should vary in systematic ways. Indeed, the evidence from Experiment 6c, shows that cues/miscues do not impact the probabilities of outcomes and, therefore, if based on these probabilities, should not affect surprise judgments, as they did in Experiment 6b.

13. General discussion

This paper has presented MEB, a new cognitive theory of surprise, and a series of eight experiments testing its novel predictions on the role that memory contents, scenarios as cues, partial explanations, and task demands play in surprise. In general, we have seen that surprises increases or decreases as various factors act to ease or inhibit our understanding of events; factors such as the density of relevant knowledge in memory, the specific given information in the setting and outcome, and one's instructions to the task (e.g., being oriented to explain). On the whole the present findings provide significant converging evidence for the theory's key proposals:

(i) *Memory contents are critical:* MEB suggested that the available knowledge in long-term memory plays a key role in resolving the surprise; if there is a lot of deep knowledge about the scenario's domain then surprise is more likely to be resolved easily. Although, empirically, it is hard to measure these differences in memory contents, as they are filtered by how the scenario cues that knowledge, we have provided evidence to show that explanations have systematic regularities, indicative of differences in the knowledge used to form them (see Experiments 2, 3, 4 and 6a). More explanations are produced for high-knowledge scenarios than low-knowledge scenarios (Experiment 2 and 6a), and there is a tendency to produce common explanations for high-knowledge scenarios, reflected in the proportion-of-agreement measure (Experiments 3 and 4).

(ii) *Scenarios as cues:* The scenario (setting plus outcome) acts as a cue to revelant knowledge in memory, so the resolution of a surprising event is critically influenced by the given information in the scenario. We found that scenarios systematically differ in how well they access knowledge in memory. Experiments 1-6 provided repeated demonstrations, using a variety of different material sets, that surprising scenarios are differentiated by having lots of associated relevant knowledge to be used for explaining (including, perhaps, associated "ready-made" explanations that need minimal adaptation to be used), or not (known versus less-known scenarios and outcomes), with the former being consequently perceived as less surprising than the latter. The experiments also found supporting evidence on the time course of surprise judgments; for known scenarios and outcomes compared to less-known scenarios and outcomes (in Experiment 2 and Experiment 4, but not replicated in Experiment 6b, although miscueing keywords were found to increase the time taken to provide surprise judgments for known outcomes more than less-known outcomes). MEB also predicted that when additional information is added to scenarios, this influences surprise in predictable ways; cueing relevant information decreases surprise, while miscueing toward irrelevant information increases surprise (Experiments 6a and 6b). Experiments 6a and 6b together showed that miscues seemed to make explanation more difficult, lowering explanation productivity and increasing surprise.

(iii) *Partial explanations will reduce surprise*: We predicted that adding partial explanations to scenarios would decrease surprise and constrain explanation productivity; Experiment 2 replicated the previous finding that given information in the form of partial explanations (enabling sentences) in the setting of the scenario reduces the perceived surprise of an outcome, because the explanation process is eased. This experiment also provided evidence on the time course of this, suggesting that partial explanations act to constrain concepts retrieved from memory, reducing the range of possible explanations.

(iv) *Task demands:* The eight experiments together provide cumulative support for the metacognitive proposal that as more cognitive work is done perceptions of surprise increase. Being asked to produce explanations as opposed to answering comprehension questions has the effect of reducing the perceived surprise of

> surprising outcomes (Experiments 3 and 4), while easing explanation reduced surprise ratings even when participants were not explicitly asked to produce explanations before providing surprise judgments (Experiments 6a and 6b). Experiment 5, perhaps, provides the most direct evidence in showing that increasing the cognitive load in being asked to explain can reverse people's decreasing ratings of surprise for known outcomes when three explanations rather than one are required in the task, while Experiment 7 shows that making explanation more difficult by providing miscues can also increase surprise.

In the remainder of this section, we outline some outstanding empirical issues, the relationship of the current MEB theory of surprise to previous proposals, and some concluding comments on everyday feelings of surprise.

13.1 Empirical issues

In the introduction we traced how surprise has moved from being considered as a basic emotion to being treated as a more cognitive phenomenon. In this paper, we have focused on the cognitive aspects of surprise. This is not to deny that there is an important affective aspect to surprise that been explored in many studies (e.g., Meyer et al., 1997; Reisenzein, 2000; Silvia, 2009). Clearly, when people encounter surprising events in the world they often immediately experience the "shock of the surprise" and a certain level of arousal; a shock that sets in train the explanatory processes described here. We acknowledge that participants in the present study were unlikely to be similarly shocked, as they were asked to appraise the surprising events encountered by other actors. There is some consensus in the literature that asking people to appraise surprising events is a reasonable and ethically more-acceptable paradigm for exploring surprise (see, e.g., Gendolla & Koller, 2001;

Maguire et al., 2011; Teigen & Keren, 2003). However, having said this, an important empirical question for future work is to confirm that the resolution of surprise is the same for directly and vicariously experienced events. There are no indications from the current literature that such differences exist, we know of no studies that have explicitly compared both contexts for surprise at once.

A more specific empirical query concerns whether different instructions in this appraisal-paradigm could radically alter people's surprise judgments on the key variables examined here. Such a possibility is not supported by the evidence. In Experiment 4 we changed the question participants were posed for the surprise judgment from "How surprised would *you* be" to "How surprised would [*character's name*] be". Although this instruction changes the task from assessing one's own surprise to assessing another's surprise in this context, it is notable that the pattern of findings on the key variables remained the same (a similar non-effect for such instructional changes was also found by Teigen & Keren, 2002, Experiment 5).

13.2 Relation to existing theories

In the introduction to this paper, we pointed out that cognitive theories of surprise can be divided into "probability" and "sense-making" camps; probability theories focus on the properties of surprising outcomes, characterising them as low-probability events, disconfirmed expectations, schema-discrepant events or events of contrasting probabilities (e.g., Meyer et al., 1997; Reisenzein & Studtmann, 2007; Schützwohl & Reisenzein, 1999; Teigen & Keren, 2003), whereas sense-making theories stress the importance of understanding and integrating the surprising event (Kahneman & Miller, 1986; Maguire & Keane, 2006; Maguire et al., 2011; Pezzo, 2003; Thagard, 2000). In this sub-section, we consider the relationship of the current explanation-based theory to previous theories from both probability and sense-making approaches.

13.2.1 Assessing probability theories of surprise

Intuitively, a simple probability account is very attractive, it appears to capture an ineluctable truth, namely that surprising events are, almost by definition, low probability events. So, the classic probability theory maintains that people develop expectations about events unfolding in the world and experience surprise when deviations occur, when schematic predictions fail, or are disconfirmed. The standard evidence for this theory comes from findings of an inverse relationship between judgments/experiences of surprise and estimates of the probability of outcomes (see e.g., Frank, 2009; Lorini & Castelfranchi, 2006; Schützwohl, 1998; Stiensmeier-Pelster, Martini & Reisenzein, 1995). One immediate problem with this account, raised by Kahneman and Miller (1986), is that people cannot always rely on prediction, that often people retrospectively make sense of surprises. In any given everyday scenario, there are a very large number of possible outcomes and it is implausible to assume that people elaborate all of these outcomes, assigning them appropriate probabilities. This fact also seems to limit the applicability of using a Bayesian approach that focuses on prior and posterior belief distributions to predict surprise[9]. For example, it is implausible to assume that every time Michael Jackson was mentioned in the media, people implicitly/explicitly generated an expectation that he had not died, with an associated probability.

In our experience, many cognitive psychologists accept the classic theory, almost implicitly, and would regard the factors examined here as ones that can be easily accounted for as manipulations that change the probabilities of outcomes, and hence perceived surprise.

[9] As we shall see later, this is the crux of the problem for any probabilistic account; many surprising events cannot be deteremined in advance, so that probabilities can be assigned to them.

Throughout this paper, after each experiment, we have made our best efforts to determine how probability theories might handle the results found. On the whole, we have seen that variants of probability theory either do not account for the evidence or, only do so, with ad hoc assumptions that are not obvious parts of probability theory. Specifically, we see three major problems in developing a probability account to account for the present results: (i) empirical demonstrations in the literature disconfirm the classical probability theory, (ii) saving the theory hinges on ad-hoc elaborations of the theory, and (iii) saving the theory requires several considerable, non-trivial extensions to it (indeed, extensions that turn it into a theory closer to the present one).

13.2.1.1 Empirical disconfirmations of classic probability theory

The classic probability theory of surprise has been disconfirmed in several studies. First, Teigen and Keren (2003) developed a series of experiments in which they showed that surprise and probability do not always exhibit an inverse relationship. In their Erik-racing scenarios, in which the probability of Erik winning a race against several competitors is systematically varied, they showed that higher probability events were sometimes rated as more surprising than lower probability events. Teigen and Keren (2003) argued that it was not the absolute probability of the outcome *per se* that was important, but the contrast between the outcome and an expectation. For example, if there are seven runners in the race, each having a 14% chance of winning and Erik wins, then our surprise is low because the outcome does not contrast with a more likely alternative. However, when the lead runner is perceived as being very likely to win, with Erik trailing behind in second place (i.e., the leader has a much higher chance of winning than Erik), it is the contrast between the outcome of Erik winning and the expectation of him losing that makes his win surprising. This *contrast hypothesis* clearly points to a comparison process that must surely have some role in surprise, perhaps as a mechanism used in explanation formation; indeed, Kahneman and

Miller's (1986) proposals on counterfactual thinking seem to require a similar comparison process between outcomes and counterfactual norms. However, as we have seen, Maguire et al. (2011) explicitly recorded subjective probabilities for different outcomes in their scenarios and found in a regression analysis that they did not predict levels of surprise, using either the probabilities of the expected outcomes or contrasting probabilities (see Experiment 4, Maguire et al., 2011). Finally, in Experiment 6c we saw that a similar assessment of subjective probabilities did not correspond to the pattern of surprise judgments for the cueing/miscueing manipulation (Experiments 6b and 6c). These disconfirmations of the Classical probability theory suggest that its broad intuitive plausibility is not supported by the evidence.

13.2.1.2 Ad hoc extensions to the classic theory

If we set aside these empirical wrinkles for the classic probability theory, it could be argued that it can account for the effects reported here with the following added propositions:

(i) known outcomes are higher-probability outcomes than less-known outcomes, therefore are considered less surprising;

(ii) given enabling information (partial explanations) modify the probabilities of outcomes (either predictively or retrospectively) and hence change surprise;

(iii) instructional effects to explain versus comprehend scenarios change the probabilities of outcomes (perhaps by increasing the perceived likelihood of the outcome after explaining);

(iv) the number-of-explanations effect modifies the probabilities of outcomes (though we cannot plausibly see why);

(v) cueing versus miscueing keywords modify the probabilites of stated outcomes and hence change perceived surprise (although, again, it is not wholly clear why, and in fact, no evidence for this was found in Experiment 6c).

Some of these propositions appear more plausible then others. At best, this extended theory is an after-the-fact, computational-level description of what has been observed. At worst, it is a bundle of ad-hoc additions that are not natural or obvious parts of the original formulation. As we said in the introduction, though the theory may claim to account for the present findings after the fact, there is nothing in the classic probability theory that would lead us to propose the present experiments. Rather, they emerge from the MEB theory of surprise, as do the exploration of new measures (such as explanation productivity and proportion-of-agreement). But, perhaps the biggest problem with this ad hoc saving of the classic probability theory is that it fails to recognise that probability accounts need to be radically reconceived if they are to work.

13.2.1.3 Shaping a new probability theory of surprise

We believe that an adequate probability theory of surprise, one that goes beyond the classic theory, is non-trivial to formulate. One option would be to adopt a Bayesian framework. Itti and Baldi (2005, 2009, Baldi & Itti, 2010) have proposed such an account in AI to account for surprise, albeit largely for simple perceptual stimuli, and Griffiths & Tenenbaum (2007) have proposed a Bayesian theory of how people handle coincidences, that is not irrelevant to surprise. Clearly, some form of belief-updating within a Bayesian framework could be advanced to explain the effects of enabling conditions or, perhaps, even the effect of explaining the outcome. However, these proposals assume (i) that probabilities can be meaningfully assigned to information items in the scenario, (ii) to any inferences made about the scenario, and (iii) that these can then be meaningfully related to expected probabilities for the outcome. These questions may be the most significant stumbling block for the development of any probabilistic account (cf., more general observations made by Jones & Love, 2011). Here we consider briefly what these questions mean for the two best Bayesian accounts in the literature.

Griffiths and Tenenbaum (2007) formally defined a probabilistic account for coincidences, which do not always correspond to simple low probability events, using a sophisticated Bayesian framework for causal induction. However, this theory relies on situations for which alternative theories (i.e., expectations) have been developed, so that prior beliefs and likelihoods can be calculated. So, one of the problems for this account is elaborating all the alternative explanations for an outcome, before that outome is known; recall, Kahneman & Miller's observation that many explanations are constructed restrospectively, and could not plausibly assumed to be present before the surprising outcome occurs.

Baldi and Itti's (2010*)* Bayesian theory of surprise, mathematically defines surprise as the effect that an event has on an observer; specifically, surprise is defined as the distance between prior and posterior belief distributions (see also Itti & Baldi, 2005, 2009). They have shown this theory of surprise to work well in predicting human gaze by computing surprise over images and video stimuli in a computer vision system using a neural network architecture. However, Itti and Baldi (2009) note that a consistent definition of surprise (using a Bayesian framework), must involve prior and posterior distributions to capture subjective expectations. So, again, for this theory prior beliefs (i.e., expectations) necessarily need to be computed so that the change between prior and posterior belief distributions can be calculated. As such, the theory cannot account for instances of surprise for which expectations are not computed in advance. Current probabilistic accounts seem to be necessarily ad hoc; once we know what has occurred, and what has been explained, then perhaps we could assign probabilities to the parts of the scenario. Indeed, as we have seen (in Experiments 6c), even when this can be done it does not always do a good job at prediction.[10]

Perhaps, the real fundamental theoretical problem for probabilistic accounts is the need for their computational-level proposals to be supplemented by additional algorithmic-level propositions. Following Jones and Love's (2011) suggestions, such a Bayesian account would have to be supplemented with a mechanism, an algorithmic-level description that accounts for the impacts of various empirical factors on the components of the Bayesian model. We believe that such a complete account would end up having to subsume many of the precepts of the current theory (e.g., about how the search in memory is directed, how relevant concepts are retrieved, and how these are built into explanations for outcomes). So, a new, more successful probability theory would appear to hinge on incorporating many of the propositions of the current account. Obviously, we would welcome such an account, should it prove to be feasible. For now, we will simply note that, such an account does not exist in the literature, is hard to imagine, and lies outside the scope of this paper.

13.2.2 Several shades of sense-making

Clearly, the MEB theory of surprise sits more firmly within the sense-making tradition. As we shall see, many of its precepts are echoed in previous work, though it advances these ideas in several respects.

Kahneman and Miller's (1986) sense-making proposals originally inspired the current theory. Obviously, their emphasis is more on norms and counterfactual thinking than on surprise. Their main relevance here is for how surprise triggers counterfactual thinking. We have seen that specific evidence of the contrastive effects of counterfactuals on surprise

[10] Recently, our colleagues (see, e.g., Maguire, Moser, Maguire & Keane, 2013) have argued that surprise needs a probabilistic account that makes no assumptions about prior probabilities; one that instead uses Algorithmic Information Theory (e.g., Kolmogorov Complexity). This may be a fruitful direction to pursue, but these ideas have not yet been seriously reviewed by the community.

remains elusive, though clearly more experimental work is required before one abandons this proposal. The main idea we take from their approach is the insight that surprise does not necessarily hinge on prediction but can arise from retrospective sense-making. Clearly, prediction could play a role, but it does not necessarily have to occur for an outcome to be deemed surprising, or for the explanation process to operate to resolve these surprising outcomes.

Thagard's (2000) coherence theory is a general theory of coherence as constraint satisfaction that has been applied to many aspects of human thinking. Thagard has developed models of explanatory coherence that capture the semantic connectedness of ideas in a parallel constraint satisfaction model (called ECHO) and of emotional coherence (in a model called HOTCO). Within HOTCO, surprise is characterised as a metacoherence emotion which "reflects a judgment that a situation has occurred differently from what was expected" (Thagard, 2000, p. 194). As we see it, Thagard's account sees surprise occurring when the coherence of one's expectations deviate from the coherence of what actually occurs, with this coherence including explanatory coherence, and the discrepancy linked to an emotion component. Thagard's account may provide us with a sketch of the key, computational-level, informational constraints on the goodness of explanations, and may provide a model of the competition processes that select the best explanation among many for a surprising outcome. The theory says less about the factors, captured by the current theory, that establish the representations over which coherence is computed. Furthermore, empirically, it is much harder to examine aspects of explanation competition, though this clearly needs to be explored in future work.

Maguire et al.'s (2011) sense-making account stresses integration as a comprehension step, which may involve explanation, to handle surprising events. This account is closest to the present theory and explicitly developed the idea that given information would ease integration and, hence, reduce surprise. The present account says more about the contents of memory, the factors that impact the retrieval of relevant information and the construction of explanations and explores many new measures for profiling surprise. The present theory also includes the novel idea that surprise is a metacognitive sense of the cognitive work done to explain the surprising event.

As we saw earlier, surprise has been implicated as a factor that influences hindsight bias. In this literature, *Pezzo's (2003)* work comes closest to the present theory; he proposed that surprise in hindsight bias involves making sense of an incongruent outcome and hints at it being a metacognitive "feeling" (e.g., "if an outcome produces an effortful search that is not successful, this should reduce, remove, or even reverse hindsight bias, and produce resultant surprise levels that are relatively high", Pezzo, 2003, p. 424). However, for Pezzo, surprise is a factor for exploring hindsight bias, so he says much less about what the sense-making process involves; he sketches it as a process of dealing with surprise resulting from expectation-disconfirmation (similar to that found in probability theories). In this literature, others have also suggested that surprise may be used metacognitively as an indicator to trigger a sense-making process to resolve prediction-outcome discrepancies, ultimately affecting hindsight bias (Müller & Stahlberg, 2007; Ofir & Mazursky, 1997). We would argue that the MEB theory presented here significantly fleshes out these suggestions. Finally, it should be noted that there are many useful experimental-paradigm analogies between studies of hindsight bias and surprise (for parallel manipulations see Choi & Nisbett, 2000; Pezzo, 2003; Touroutoglou and Efklides, 2010); though the two phenomena are clearly separate (one can be shown to be a factor influencing the other) they seem to be influenced by several similar factors (e.g., availability of knowledge, ease of processing, sense-making).

14. Conclusion

Why are some events more surprising than others? Why are we less surprised to hear that Amy Winehouse, Margaret Thatcher or Kurt Cobain died, but more surprised when we hear of Lady Diana or Michael Jackson's death? We propose that surprising outcomes have to be explained in order to make them sensible. In this paper, we have argued that every surprise scenario cues information in memory that is retrieved to explain how the outcome could have occurred and to resolve the surprise.

To belabour our somewhat morbid example, according to our theory, Jackson's death was surprising for two main reasons. First, because he did not fit the young-self-destructive-rock-star stereotype, few people had "ready-made" explanations to account for his death. Second, there was not a lot of relevant knowledge available and it was accordingly hard to explain the death, which in turn influenced the surprise experienced. Of course, in time, we began to hear about Jackson's bizarre history of illness and medication by his personal doctor. If this enabling information (or partial explanation) had been known before the death, it would have been easier to explain and we simply would not have been as surprised.

Many other events involving the deaths of younger rock-stars are, unfortunately, less surprising, like those of Amy Winehouse, Kurt Cobain or Janis Joplin. On our account, these events are less surprising because there is a lot of available relevant knowledge to resolve the surprising event, to explain easily what occurred. For someone like Amy Winehouse there was the whole panoply of prior knowledge about the young, talented, drug-abusing rock-star; this is a very well-known, though surprising, outcome. Furthermore, just before her death we had regular newspaper reports with enabling information reporting her most recent episode of public drunkenness (in Winehouse's case at a concert several days before her death). In short, Winehouse's death was easier to find an explanation for, and so was less surprising.

The tragedy of sudden death can be more, or less, surprising depending on how much knowledge we have about the situation surrounding that death, specific given information about these people's lives, and the cognitive work required to explain what has occurred. These are the key components that impact how surprised we feel for these and other surprising events, a feeling that depends on our metacognitive sense of the explanatory work necessary to understand surprising events in the world.

References

Anderson, J, R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Adler, J. E. (2008). Surprise. *Educational Theory, 58*(2), 149-173.

Ash, I. K. (2009). Surprise, memory, and retrospective judgment making: Testing cognitive reconstruction theories of the hindsight bias effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(4),* 916-933.

Bae, B.C., & Young, R. M. (2008). A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In U. Spierling & N. Szilas (Eds.), *Proceedings of ICIDS, LNCS, 5334* (pp. 156-167). Heidelberg: Springer.

Bae, B. C., & Young, R. M. (2009). Suspense? Surprise! or How to generate stories with surprise endings by exploiting the disparity of knowledge between a story's reader and its characters. In I. A. Iurgel, N. Zagalo & P. Petta (Eds.), *Proceedings of ICIDS, LNCS, 5915* (pp. 304-307). Berlin: Springer.

Baldi, P., & Itti, L. (2010). Of bits and wows: a Bayesian theory of surprise with applications to attention. *Neural Networks*, *23*(5), 649-666.

Byrne, R. M. J. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences, 6*(10), 405-445.

Byrne, R. M. J. (2005). *The rational imagination: How people create alternatives to reality.* Cambridge, MA: MIT Press.

Byrne, R. M. (2007). Precis of the rational imagination: How people create alternatives to reality. *Behavioral and Brain Sciences*, *30*(5), 439-452.

Carli, L. L. (1999). Cognitive reconstruction, hindsight, and reactions to victims and perpetrators. *Personality and Social Psychology Bulletin, 25*(8), 966–979.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145-182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Choi, I., & Nisbett, R. E. (2000). Cultural psychology of surprise: holistic theories and recognition of contradiction. *Journal of Personality and Social Psychology*, *79*(6), 890-905.

Connell, L., & Keane, M. T. (2006). A model of plausibility. *Cognitive Science*, *30*(1), 95-120.

Darwin, C. R. (1872). *The expression of the emotions in man and animals.* London: John Murray.

Dehghani, M., Iliev, R., & Kaufmann, S. (2012). Causal explanation and fact mutability in counterfactual reasoning. *Mind & Language, 27*(1), 55-85.

Dixon, J., & Byrne, R. M. J. (2011). Counterfactual thinking about exceptional actions. *Memory and Cognition, 39*(7), 1317–1331.

Durkin, K. (2011). The self-explanation effect when learning mathematics: a meta-analysis. *Science*, *26*(2), 147-179.

Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–9.

Foster, M. I. & Keane, M. T. (2013). Surprise! You've got some explaining to do. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2321-2326). Austin, TX: Cognitive Science Society.

Frank, S. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In N. Taatgen and H. van Rijn (Eds.), *Proceedings of the 31st*

Annual Conference of the Cognitive Science Society, (pp. 1139–1144). USA: Cognitive Science Society.

Gernsbacher, M. A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A. (1991). Cognitive processes and mechanisms in language comprehension: The structure building framework. *Psychology of Learning and Motivation*, *27*, 217-263.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes*, *23*(3), 265-304.

Gendolla, G. H. (1997). Surprise in the context of achievement: The role of outcome valence and importance. *Motivation and Emotion*, *21*(2), 165-193.

Gendolla, G. H., & Koller, M. (2001). Surprise and motivation of causal search: How are they affected by outcome valence and importance? *Motivation and Emotion*, *25*(4), 327-349.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science, 3*(2), 371-398.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163–189.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395.

Griffiths, T. L., & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition, 103*(2), 180-226.

Haton, J. P., Keane, M., & Manago, M. (Eds.) (1995). *Advances in Case-Based Reasoning* (Vol. 984). Amsterdam: Springer Verlag.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: John Wiley & Sons.

Itti, L., & Baldi, P. F. (2005). Bayesian surprise attracts human attention. In Y. Weiss, B. Scholköpf. & J. Platt (Eds.), *Advances in neural information processing systems* (pp. 547-554). Cambridge, MA : MIT Press.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*(10), 1295-306.

Izard, C. (1977). *Human emotions.* Plenum Press, New York.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111*(3), 640-661.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34*(4), 169-231.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136-153.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, *18*(3), 387-438.

Khemlani, S. S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta psychologica*, *139*(3), 486-491.

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2011). Harry Potter and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition*, *39*(3), 527-535.

Keil, F. C. (2006). Explanation and understanding. *Annual Reviews of Psychology, 57,* 227-254.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge UK: Cambridge University Press.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological bulletin, 110*(3), 499-519.

Laird, J. E. (2012). *The Soar cognitive architecture.* Cambridge, MA: MIT Press

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33,* 1-64.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Landauer, T. K., & Dumais, S. T. (1996). How come you know so much? From practical problem to theory. In D. Hermann, C. McEvoy, M. Johnson, & P. Hertel (Eds.), *Basic and applied memory: Memory in context* (pp.105-126). Mahwah, NJ: Erlbaum.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.

Leake, D. B. (1991). Goal-based explanation evaluation. *Cognitive Science, 15*(4), 509-545.

Loewenstein, J., & Heath, C. (2009). The Repetition-Break plot structure: A cognitive influence on selection in the marketplace of ideas. *Cognitive Science,33*(1), 1-19.

Lorini, E., Castelfranchi, C. (2006). The unexpected aspects of surprise. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(6), 817-835.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. *Cognition, 99*(2), 167-204.

Lombrozo, T. (2012). Explanation and abductive inference. In K.J. Holyoak & R.G. Morrison (Eds.), *Oxford Handbook of Thinking and Reasoning* (pp. 260-276), United Kingdom: Oxford University Press

Macedo, L. (2010). The practical advantage of surprise-based agents (Extended Abstract). In W. van der Hoek, G. A. Kaminka, Y. Lesperance, M. Luck, & S. Sen (Eds.), *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems* (pp. 1401-1402). Toronto: IFAMAS.

Macedo, L., & Cardoso, A. (2001). Creativity and surprise. In G. Wiggins (Ed.), *Proceedings of the AISB'01 Symposium on Artificial Intelligence and Creativity in Arts and Science* (pp. 84-92). United Kingdom: University of York.

Macedo, L., Reisenzein, R., & Cardoso, A. (2004). Modeling forms of surprise in artificial agents: Empirical and theoretical study of surprise functions. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 588-593). Mahwah, NJ: Erlbaum.

Macedo, L., Cardoso, A., Reisenzein, R., Lorini, E., & Castelfranchi, C. (2009). Artificial surprise. In Vallverdú, J. & Casacuberta, D. (Eds.) *Handbook of research on synthetic emotions and sociable robotics: New applications in affective computing* *and artificial intelligence* (pp. 267-291). United Kingdom: Information Science Reference.

Maguire, R., & Keane, M. T. (2006). Surprise: Disconfirmed expectations or representation-fit? In R. Son (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 531-536). Hillsdale, NJ: Erlbaum.

Maguire, R., Maguire, P., & Keane, M. T. (2011). Making sense of surprise: An investigation of the factors influencing surprise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(1), 176-186.

Maguire, P., Moser, P., Maguire, R., & Keane, M. T. (2013). A computational theory of subjective probability [Featuring a proof that the conjunction effect is not a fallacy]. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society* (pp. 960-965). Austin, TX: Cognitive Science Society.

Mandel, D. R. (2003). Effect of counterfactual and factual thinking on causal judgments. *Thinking & Reasoning, 9*(3), 245-265.

McCloy, R., & Byrne, R. M. (2002). Semifactual "even if" thinking. *Thinking & reasoning*, *8*(1), 41-67.

McEleney, A., & Byrne, R. M. (2006). Spontaneous counterfactual thoughts and causal explanations. *Thinking & Reasoning*, *12*(2), 235-255.

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*(1), 51-62.

McNamara, D. S., & Scott, J. L. (1999). Training reading strategies. *Proceedings of the Twenty-First Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion, 21*(3), 251-274.

Müller, P. A., & Stahlberg, D. (2007). The role of surprise in hindsight bias: A metacognitive model of reduced and reversed hindsight bias. *Social Cognition, 25*(1), 165-184.

Munnich, E., Milazzo, J., Stannard, J., & Rainford, K. (2012). Can causal sense-making benefit foresight, rather than biasing hindsight? In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2669-2674). Austin, TX: Cognitive Science Society.

Munnich, E., Ranney, M., A. & Song, M. (2007). Surprise, surprise: The role of surprising numerical feedback in belief change. In D.S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 503-508). Mahwah, NJ: Erlbaum.

Nestler, S., & Egloff, B. (2009). Increased or reversed? The effect of surprise on hindsight bias depends on the hindsight component. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1539-1544.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: A memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,* 1200–1210.

Ofir, C., & Mazursky, D. (1997). Does a surprising outcome reinforce or reverse the hindsight bias? *Organizational Behavior and Human Decision Processes, 69*(1), 51-57.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331.

Pezzo, M. V. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory, 11*(4/5), 421-441.

Pezzo, M. V. & Pezzo, S. P. (2007). Making sense of failure: A motivated model of hindsight bias. *Social Cognition, 25*(1), 147-164.

Piaget, J. (1952). *The origins of intelligence in children.* New York: W W Norton & Company.

Plutchik, R. (1991). *The emotions.* New York: University Press of America.

Ramscar, M., Dye, M., Gustafson, J. W., & Klein, J. (2013). Dual routes to cognitive flexibility: Learning and response-conflict resolution in the dimensional change card sort task. *Child Development, 84*(4), 1308-1323.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience, 4*(3), 193-202.

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion, 14*(1), 1-38.

Reisenzein, R., & Studtmann, M. (2007). On the expression and experience of surprise: no evidence for facial feedback, but evidence for a reverse self-inference effect. *Emotion, 7*(3), 601-611.

Roese, N. J. (1997). Counterfactual thinking. *Psychological bulletin*, *121*(1), 133.

Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology, 32*(3), 197-227.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science, 7*(5), 411-426.

Roy, M., & Chi, M. T. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learnng.* (pp. 271-286). New York: Cambridge University Press.

Schkade, D. A., & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes, 49*(1), 105–123.

Schank, R. C. (1986). *Explanation patterns: Understanding mechanically and creatively.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schank, R. C., Kass, A., & Riesbeck, C. K. (Eds.) (1994). *Inside case-based explanation.* UK: Lawrence Erlbaum Associates.

Schützwohl, A. (1998). Surprise and schema strength. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1182-1199.

Schützwohl, A., & Reisenzein, R. (1999). Children's and adult's reactions to a schema-discrepant event: A developmental analysis of surprise. *International Journal of Behavioral Development, 23*(1), 37-62.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*(2), 195-202.

Stiensmeier-Pelster, J., Martini, A., & Reisenzein, R. (1995). The role of surprise in the attribution process. *Cognition and Emotion, 9*(1), 5-31.

Silvia, P. (2009). Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts, 3*(1), 48-51.

Teigen, K. H., & Keren, G. (2002). When are successes more surprising than failures? *Cognition and Emotion, 16*(2), 245-268.

Teigen, K. H., & Keren, G. (2003). Surprises: Low probabilities or high contrasts? *Cognition, 87*(2), 55-71.

Thagard, P. (2000). *Coherence in thought and action.* London, UK: MIT Press.

Touroutoglou, A., & Efklides, A. (2010). Cognitive interruption as an object of metacognitive monitoring: Feeling of difficulty and surprise. In A. Efklides & P. Misailidi (Eds.), *Trends and prospects in meta-cognition research* (pp. 171-208). New York: Springer.

Tomkins, S. S. (1962). *Affect, imagery, consciousness.* (Vol. 1). New York: Springer.

Tsang, N. M. (2013). Surprise in social work education. *Social Work Education: The International Journal, 32*(1), 55-67.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232.

Wasserman, D., Lempert, R. O., & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psychology Bulletin, 17*(1), 30–35.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56*(2), 161-169.

Williams, J. J., Lombrozo, T., & Rehder, B. (2010). Why does explaining help learning? Insight from an explanation impairment effect. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society

Williams, J. J., Lombrozo, T., & Rehder, B. (2013). The hazards of explanation: Overgeneralization in the face of exceptions. *Journal of Experimental Psychology: General*, *142*(4), 1006-1014.

Wilson, T. D., & Gilbert, D. T. (2008). Explaining away: A model of affective adaptation. *Perspectives on Psychological Science, 3*(5), 370-386.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 386.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185.