



## From mere coincidences to meaningful discoveries ☆

Thomas L. Griffiths <sup>a,\*</sup>, Joshua B. Tenenbaum <sup>b</sup>

<sup>a</sup> *Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, United States*

<sup>b</sup> *Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, United States*

Received 4 August 2005; revised 12 March 2006; accepted 19 March 2006

---

### Abstract

People's reactions to coincidences are often cited as an illustration of the irrationality of human reasoning about chance. We argue that coincidences may be better understood in terms of rational statistical inference, based on their functional role in processes of causal discovery and theory revision. We present a formal definition of coincidences in the context of a Bayesian framework for causal induction: a coincidence is an event that provides support for an alternative to a currently favored causal theory, but not necessarily enough support to accept that alternative in light of its low prior probability. We test the qualitative and quantitative predictions of this account through a series of experiments that examine the transition from coincidence to evidence, the correspondence between the strength of coincidences and the statistical support for causal structure, and the relationship between causes and coincidences. Our results indicate that people can accurately assess the strength of coincidences, suggesting that irrational conclusions drawn from coincidences are the consequence of overestimation of the plausibility of novel causal forces. We discuss the implications of our account for understanding the role of coincidences in theory change.

© 2006 Elsevier B.V. All rights reserved.

---

☆ While completing this work, TLG was supported by a Stanford Graduate Fellowship and JBT by the Paul E. Newton Career Development Chair. We thank Tania Lombrozo, Tevye Krynski, and two anonymous reviewers for their comments on this manuscript, Onny Chatterjee and Davie Yoon for their help in running the experiments, and Persi Diaconis for originally inspiring our interest in coincidences.

\* Corresponding author. Tel.: +1 401 863 9563; fax: +1 401 863 2255.

E-mail address: [tom\\_griffiths@brown.edu](mailto:tom_griffiths@brown.edu) (T.L. Griffiths).

*Keywords:* Coincidences; Probabilistic reasoning; Theory change; Causal induction; Bayesian models

---

## 1. Introduction

In the last days of August in 1854, the city of London was hit by an unusually violent outbreak of cholera. More than 500 people died over the next fortnight, most of them in a small region in Soho. On September 3, this epidemic caught the attention of John Snow, a physician who had recently begun to argue against the widespread notion that cholera was transmitted by bad air. Snow immediately suspected a water pump on Broad Street as the cause, but could find little evidence of contamination. However, on collecting information about the locations of the cholera victims, he discovered that they were tightly clustered around the pump. This suspicious coincidence hardened his convictions, and the pump handle was removed. The disease did not spread any further, furthering Snow's (1855) argument that cholera was caused by infected water.

Observing clusters of events in the streets of London does not always result in important discoveries. Towards the end of World War II, London came under bombardment by German V-1 and V-2 flying bombs. It was widespread popular belief that these bombs were landing in clusters, with an unusual number of bombs landing on the poorer parts of the city (Johnson, 1981). After the war, R.D. Clarke of the Prudential Assurance Company set out to 'apply a statistical test to discover whether any support could be found for this allegation' (Clarke, 1946, p. 481). Clarke examined 144 square miles of south London, in which 537 bombs had fallen. He divided this area into small squares and counted the number of bombs falling in each square. If the bombs fell uniformly over this area, then these counts should conform to the Poisson distribution. Clarke found that this was indeed the case, and concluded that his result 'lends no support to the clustering hypothesis' (1946, p. 481), implying that people had been misled by their intuitions.<sup>1</sup>

Taken together, the suspicious coincidence noticed by John Snow and the mere coincidence that fooled the citizens of London present what seems to be a paradox for theories of human reasoning. How can coincidences simultaneously be the source of both important scientific discoveries and widespread false beliefs? Previous research has tended to focus on only one of these two faces of coincidences. Inspired by examples similar to that of Snow,<sup>2</sup> one approach has focused on conceptual

---

<sup>1</sup> Clarke's investigations were later introduced to a broader audience by Feller (1968).

<sup>2</sup> Such examples abound. In considering the apparent rotation of stars about the Earth, Aristotle viewed the coincidence between the rate of motion and the distance traversed as evidence for the existence of a single celestial sphere (Franklin, 2001, pp. 133–134). Halley would never have discovered his comet without noticing the surprising regularity in the paths and dates in a table of orbits (Cook, 1998; Hughes, 1990; Yeomans, 1991). Semmelweis might not have developed his theory of contagion without noting the similarity in the symptoms of a doctor injured during an autopsy and those of patients in his ward (Hempel, 1966). Perrin's (1913/1990) argument for the objective reality of molecules was based upon the suspiciously similar estimates of Avogadro's number produced by several quite different methods of measuring molecular magnitudes (Hacking, 1983).

analyses or quantitative measures of coincidences that explicate their role in rational inference (Horwich, 1982; Schlesinger, 1991), causal discovery (Owens, 1992) and scientific argument (Hacking, 1983). An alternative approach, inspired by examples like the bombing of London,<sup>3</sup> has analyzed the sense of coincidence as a prime example of shortcomings in human understanding of chance and statistical inference (Diaconis & Mosteller, 1989; Fisher, 1937; Gilovich, 1993; Plous, 1993). Neither of these two traditions has attempted to explain how the same cognitive phenomenon can simultaneously be the force driving human reasoning to both its greatest heights, in scientific discovery, and its lowest depths, in superstition and other abiding irrationalities.

In this paper, we develop a framework for understanding coincidences as a functional element of causal discovery. Scientific knowledge is expanded and revised through the discovery of causal relationships that enrich or invalidate existing theories. Intuitive knowledge can also be described in terms of domain theories with structures that are analogous to scientific theories in important respects (Carey, 1985; Gopnik & Meltzoff, 1997; Karmiloff-Smith, 1988; Keil, 1989; Murphy & Medin, 1985), and these intuitive theories are grown, elaborated and revised in large part through processes of causal discovery (Gopnik et al., 2004; Tenenbaum, Griffiths, & Niyogi, *in press*). We will argue that coincidences play a crucial role in the development of both scientific and intuitive theories, as events that provide support for a low-probability alternative to a currently favored causal theory. This definition can be made precise using the mathematics of statistical inference. We use the formal language of causal graphical models (Pearl, 2000; Spirtes, Glymour, & Schienens, 1993) to characterize relevant aspects of intuitive causal theories, and the tools of Bayesian statistics to propose a measure of evidential support for alternative causal theories that can be identified with the strength of a coincidence. This approach allows us to clarify the relationship between coincidences and theory change, and to make quantitative predictions about the strength of coincidences that can be compared with human judgments.

The plan of the paper is as follows. Before presenting our account, we first critique the common view of coincidences as simply unlikely events. This analysis of coincidences is simple and widespread, but ultimately inadequate because it fails to recognize the importance of alternative theories in determining what constitutes a coincidence. We then present a formal analysis of the computational problem underlying causal induction, and use this analysis to show how coincidences may be viewed as events that provide strong but not necessarily sufficient evidence for an alternative to a current theory. After conducting an experimental test of the qualitative predictions of this account, we use it to make quantitative predictions about the strength of coincidences in some of the complex settings where classic examples of coincidences occur: coincidences in space, as in the examples of John Snow and the bombing of London, and coincidences in date, as in the famous “birthday

---

<sup>3</sup> Again there are many examples. Diaconis and Mosteller (1989), Gilovich (1993), Hardy, Harvie, and Koestler (1973), and Plous (1993) all present a number of surprising coincidences that ultimately seem to be simply the work of chance.

problem”. We conclude by returning to the paradox of coincidences identified above, considering why coincidences often lead people astray and discussing their involvement in theory change.

## 2. Coincidences are not just unlikely events

Upon experiencing a coincidence, many people react by thinking something like ‘Wow! What are the chances of that?’ (e.g., Falk, 1981–1982). Subjectively, coincidences are unlikely events: we interpret our surprise at their occurrence as indicating that they have low probability. In fact, it is often assumed that being surprising and having low probability are equivalent: the mathematician Littlewood (1953) suggested that events having a probability of one in a million be considered surprising, and many psychologists make this assumption at least implicitly (e.g., Slovic & Fischhoff, 1977). The notion that coincidences are unlikely events pervades literature addressing the topic, irrespective of its origin. This belief is expressed in books on spirituality (‘Regardless of the details of a particular coincidence, we sense that it is too unlikely to have been the result of luck or mere chance’, Redfield, 1998, p. 14), popular books on the mathematical basis of everyday life (‘It is an event which seems so unlikely that it is worth telling a story about’, Eastaway & Wyndham, 1998, p. 48), and even the statisticians Diaconis and Mosteller (1989) considered the definition ‘a coincidence is a rare event’, but rejected it on the grounds that ‘this includes too much to permit careful study’ (p. 853).

The most basic version of the idea that coincidences are unlikely events refers only to the probability of a single event. Thus, some data,  $d$ , might be considered a coincidence if the probability of  $d$  occurring by chance is small.<sup>4</sup> On September 11, 2002, exactly one year after terrorists destroyed the World Trade Center in Manhattan, the New York State Lottery “Pick 3” competition, in which three numbers from 0 to 9 are chosen at random, produced the results 9–1–1 (Associated Press, September 12, 2002). This seems like a coincidence,<sup>5</sup> and has reasonably low probability: the three digits were uniformly distributed between 0 and 9, so the probability of such a combination is  $(\frac{1}{10})^3$  or 1 in 1000. If  $d$  is a sequence of ten coinflips that are all heads, which we will denote HHHHHHHHHH, then its probability under a fair coin is  $(\frac{1}{2})^{10}$  or 1 in 1024. If  $d$  is an event in which one goes to a party and meets four people, all of whom are born on August 3, and we assume birthdays are uniformly distributed, then the probability of this event is  $(\frac{1}{365})^4$ , or 1 in 17,748,900,625. Consistent with the idea that coincidences are unlikely events, these values are all quite small.

<sup>4</sup> In general, we will use upper-case letters to indicate random variables, and lower-case letters to indicate the values taken on by those variables. Here,  $d$  is a value of the random variable  $D$ .

<sup>5</sup> Indeed, many people sought explanations other than chance: the authorities responsible for the New York lottery were sufficiently suspicious that they initiated an internal investigation, and the St Petersburg Times quoted one psychologist as saying that ‘It could be that, collectively, the people in New York caused those lottery numbers to come up 9–1–1. . . If enough people all are thinking the same thing, at the same time, they can cause events to happen’ (DeGregory, 2002).

The fundamental problem with this account is that while coincidences must in general be unlikely events, there are many unlikely events that are not coincidences. It is easy to find events that have the same probability, yet differ in whether we consider them a coincidence. In particular, all of the examples cited above were analyzed as outcomes of uniform generating processes, and so their low probability would be matched by any outcomes of the same processes with the same number of observations. For instance, a fair coin is no more or less likely to produce the outcome HHTHTTHTHT as the outcome HHHHHHHHHH. Likewise, observing the lottery numbers 7–2–3 on September 11 would be no more likely than observing 9–1–1, and meeting people with birthdays on May 14, July 8, August 21, and October 4, would be just as unlikely as any other sequence, including August 3, August 3, August 3, and August 3. Using several other examples of this kind, [Teigen and Keren \(2003\)](#) provided empirical evidence in behavioral judgments for the weak relationship between the surprisingness of events and their probability. For our purposes, these examples are sufficient to establish that our sense of coincidence is not merely a result of low probability.

We will argue that coincidences are not just unlikely events, but rather events that are *less* likely under our currently favored theory of how the world works than under an alternative theory. The September 11 lottery results, meeting four people with the same birthday, and flipping 10 heads in a row all grab our attention because they suggest the existence of hidden causal structure in contexts where our current understanding would suggest no such structure should exist. Before we explore this hypothesis in detail, we should rule out a more sophisticated version of the idea that coincidences are unlikely events. The key innovation behind this definition is to move from evaluating the probability of a single event to the probability of an event of a certain “kind”, with coincidences being events of unlikely kinds. Hints of this view appear in experiments on coincidences conducted by [Falk \(1989\)](#), who suggested that people are ‘sensitive to the extension of the judged event’ (p. 489) when evaluating coincidences. [Falk \(1981–1982\)](#) also suggested that when one hears a story about a coincidence, ‘One is probably not encoding the story with all its specific details as told, but rather as a more general event “of that kind” ’ (p. 23). Similar ideas have been proposed by psychologists studying figural goodness and subjective randomness (e.g., [Garner, 1970](#); [Kubovy & Gilden, 1991](#)), and such an account was worked out in detail by [Schlesinger \(1991\)](#), who explicitly considered coincidences in birthdays. Under this view, meeting four people all born on August 3 is a bigger coincidence than meeting those born on May 14, July 8, August 21, and October 4 because the former is of the kind *all on the same day* while the latter is of the kind *all on different days*. Similarly, the sequence of coinflips HHHHHHHHHH is more of a coincidence than the sequence HHTHTTHTHT because the former is of the kind *all outcomes the same* while the latter is of the kind *equal number of heads and tails*; out of all 1024 sequences of length 10, only two are of the former kind, while there are 252 of the latter kind.

The “unlikely kinds” definition runs into several difficulties. First there are the problems of specifying what might count as a kind of event, and which kind should

be used when more than one is applicable. Like the coinflip sequence HHTHTTHTHT, the alternating sequence HTHTHTHTHT falls under the kind *equal number of heads and tails*, but it appears to present something of a coincidence while the former sequence does not. The “unlikely kinds” theory might explain this by saying that HTHTHTHTHT is also a member of a different kind, *alternating heads and tails*, containing only two sequences out of the possible 1024. But why should this second kind dominate? Intuitively, the fact that it is more specific seems important, but why? And why isn’t alternation as much of a coincidence as repetition, even though the kinds *all outcomes the same* and *alternating heads and tails* are equally specific? How would we assess the degree of coincidence for the sequence HHHHHHHTTT? It appears more coincidental than a merely “random” sequence like HHTHTTHTHT, but what “kind of event” is relevant? Finally, why do we not consider a kind like *all outcomes that begin HHTHTTHTHT* . . . , which would predict that the sequence HHTHTTHTHT is in fact the most coincidental of all? The situation becomes even more complex when we go beyond discrete events. For example, the bombing of London suggested a coincidence based upon bomb locations, which are not easily classified into kinds.

For the “unlikely kinds” definition to work, we need to be able to identify the kinds relevant to any contexts, including those involving continuous stimuli. The difficulty of doing this is a consequence of not recognizing the role of alternative theories in determining what constitutes a coincidence. The fact that certain kinds of events seem natural is a consequence of the theory-ladenness of the observer: there is no a priori reason why any set of kinds should be favored over any other. In the cases where definitions in terms of unlikely kinds do seem to work, it is because the kinds being used implicitly correspond to the predictions of a reasonable set of alternative theories. To return to the coinflipping example, kinds defined in terms of the number of heads in a sequence implicitly correspond to considering a set of alternative theories that differ in their claims about the probability that a coin comes up heads, a fact that we discuss in more detail below. Alternative theories still exist in contexts where no natural “kinds” can be found, providing greater generality for definitions of coincidences based upon alternative theories.

Finally, even if a method for defining kinds seems clear, it is possible to find counterexamples to the idea that coincidences are events of unlikely kinds. For instance, a common way of explaining why a sequence like HHHH is judged less random (and more coincidental) than HHTT is that the former is of the kind *four heads* while the latter is of the kind *two heads, two tails* (cf. Garner, 1970; Kubovy & Gilden, 1991). Since one is much more likely to obtain a sequence with two heads and two tails than a sequence with four heads when flipping a fair coin four times, the latter seems like a bigger coincidence. The probability of  $N_H$  heads from  $N$  trials is

$$P_{\text{kind}}(D) = \binom{N}{N_H} \frac{1}{2^N}, \quad (1)$$

so the probability of the *four heads* kind is  $\binom{4}{4} \frac{1}{2^4} = 0.0625$ , while the probability of the *two heads, two tails* kind is  $\binom{4}{2} \frac{1}{2^4} = 0.375$ . However, we can easily construct a sequence of a kind that has lower probability than *four heads*: the reasonably random HHHHTHTTHHHTHTHHHTHTTHHH is but one example of the *fifteen heads, eight tails* kind, which has probability  $\binom{23}{15} \frac{1}{2^{23}} = 0.0584$ .

### 3. Coincidence as statistical inference

In addition to the problems outlined in the previous section, the definition of coincidences as unlikely events seems to neglect one of the key components of coincidences: their apparent meaningfulness. This is the aspect of coincidences that makes them so interesting, and is tied to their role in scientific discoveries. We will argue that the meaningfulness of coincidences is due to the fact that coincidences are not just arbitrary low-probability patterns, but patterns that suggest the existence of unexpected causal structure. One of the earliest statements of this idea appears in Laplace (1795/1951):

If we seek a cause wherever we perceive symmetry, it is not that we regard a symmetrical event as less possible than the others, but, since this event ought to be the effect of a regular cause or that of chance, the first of these suppositions is more probable than the second. On a table we see letters arranged in this order, *C o n s t a n t i n o p l e*, and we judge that this arrangement is not the result of chance, not because it is less possible than the others, for if this word were not employed in any language we should not suspect it came from any particular cause, but this word being in use among us, it is incomparably more probable that some person has thus arranged the aforesaid letters than that this arrangement is due to chance (p. 16).

In this passage, Laplace suggested that our surprise at orderly events is a result of the inference that these events are more likely under a process with causal structure than one based purely on chance, we should suspect that a cause was involved.

The idea that coincidences are events that provide us with evidence for the existence of unexpected causal structure has been developed further by a number of authors. In the philosophy of science, Horwich (1982) defined a coincidence as ‘an unlikely accidental correspondence between independent facts, which suggests strongly, but in fact falsely, some causal relationship between them’ (p. 104), and expressed this idea formally using the language of Bayesian inference, as we do below. Similar ideas have been proposed by Bayesian statisticians, including Good (1956, 1984) and Jaynes (2003). In cognitive science, Feldman (2004) has explored an account of why simple patterns are surprising that is based upon the same

principle, viewing events that exhibit greater simplicity than should be expected under a “null hypothesis” as coincidences.

In the remainder of the paper, we develop a formal framework which allows us to make this definition of coincidences precise, and to test its quantitative predictions. Our focus is on the role of coincidences in causal induction. Causal induction has been studied extensively in both philosophy (e.g., Hume, 1739/1978) and psychology (e.g., Inhelder & Piaget, 1958). Detailed reviews of some of this history are provided by Shultz (1982; Shultz & Kestenbaum, 1985) and White (1990). Recent research on human causal induction has focused on formal models based upon analyses of how an agent should learn about causal relationships (e.g., Anderson, 1990; Cheng, 1997; Griffiths & Tenenbaum, 2005; Lopez, Cobos, Cano, & Shanks, 1998; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). These formal models establish some of the groundwork necessary for our analysis of the functional role of coincidences.

Any account of causal induction requires a means of representing hypotheses about candidate causal structures. We will represent these hypotheses using *causal graphical models* (also known as causal Bayesian networks or causal Bayes nets). Causal graphical models are a language for representing and reasoning about causal relationships that has been developed in computer science and statistics (Pearl, 2000; Spirtes et al., 1993). This language has begun to play a role in theories of human causal reasoning (e.g., Danks & McKenzie, under version; Glymour, 1998, 2001; Gopnik et al., 2004; Griffiths & Tenenbaum, 2005; Lagnado & Sloman, 2002; Rehder, 2003; Steyvers et al., 2003; Tenenbaum & Griffiths, 2001, 2003; Waldmann & Martignon, 1998), and several theories of human causal induction can be expressed in terms of causal graphical models (Griffiths & Tenenbaum, 2005; Tenenbaum & Griffiths, 2001).

A causal graphical model represents the causal relationships among a set of variables using a graph in which variables are nodes and causation is indicated with arrows. This graphical structure has implications for the probability of observing particular values for those variables, and for the consequences of interventions on the system (see Pearl, 2000 or Griffiths & Tenenbaum, 2005, for a more detailed introduction). A variety of algorithms exist for learning the structure of causal graphical models, based upon either reasoning from a pattern of statistical dependencies (e.g., Spirtes et al., 1993) or methods from Bayesian statistics (e.g., Heckerman, 1998). We will pursue the latter approach, treating theories as generators of causal graphical models: recipes for constructing a set of causal graphical models that describes the possible causal relationships among variables in a given situation. Theories thus specify the hypothesis spaces and prior probabilities that are used in Bayesian causal induction. We develop this idea formally elsewhere (Griffiths, 2005; Griffiths, Baraff, & Tenenbaum, 2004; Griffiths & Tenenbaum, 2005; Tenenbaum & Griffiths, 2003; Tenenbaum et al., in press; Tenenbaum & Niyogi, 2003), but will use it relatively informally in this paper.

These tools provide the foundations of our approach to coincidences. In this section, we use a Bayesian perspective on causal induction to develop an account of what makes an event a coincidence, and to delineate the difference between “mere”



and “suspicious” coincidences. We then provide a more detailed formal analysis of one simple kind of coincidence – coincidences in coinflips – indicating how this account differs from the idea that coincidences are unlikely events. The section ends by identifying the empirical predictions made by this account, which are tested in the remainder of the paper.

### 3.1. What makes a coincidence?

Assume that a learner has data  $d$ , and a set of hypotheses,  $h$ , each being a theory about the system that produced that data. Before seeing any data, the learner assigns *prior probabilities* of  $P(h)$  to these hypotheses. The *posterior* probability of any hypothesis  $h$  after seeing  $d$  can be evaluated using Bayes’ rule,

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_h P(d|h)P(h)}, \quad (2)$$

where  $P(d|h)$ , known as the *likelihood*, specifies the probability of the data  $d$  being generated by the system represented by hypothesis  $h$ . In the case where there are just two hypotheses,  $h_1$  and  $h_0$ , we can express the relative degree of belief in  $h_1$  after seeing  $d$  using the *posterior odds*,

$$\frac{P(h_1|d)}{P(h_0|d)} = \frac{P(d|h_1)P(h_1)}{P(d|h_0)P(h_0)}, \quad (3)$$

which follows directly from Eq. (2). The posterior odds are determined by two factors: the *likelihood ratio*, which indicates the support that  $d$  provides in favor of  $h_1$  over  $h_0$ , and the *prior odds*, which express the a priori plausibility of  $h_1$  as compared to  $h_0$ . If we take the logarithm of Eq. (3), we obtain

$$\log \frac{P(h_1|d)}{P(h_0|d)} = \log \frac{P(d|h_1)}{P(d|h_0)} + \log \frac{P(h_1)}{P(h_0)}, \quad (4)$$

in which the log likelihood ratio and the log prior odds combine additively to give the log posterior odds.

To make this analysis more concrete, consider the specific example of evaluating whether a new form of genetic engineering influences the sex of rats. The treatment is tested through a series of experiments in which female rats receive a prenatal injection of a chemical, and the sex of their offspring is recorded at birth. In the formal schema above,  $h_1$  refers to the theory that injection of the chemical influences sex, and  $h_0$  refers to the theory that injection and sex are independent. These two theories generate the causal graphical models Graph 1 and Graph 0 shown in Fig. 1. Under Graph 0, the probability that a rat is male should be 0.5, while under Graph 1, rats injected with the chemical have some other probability of being male. Imagine that in the experimental test, the first 10 rats were all born male. These data,  $d$ , would provide relatively strong support for the existence of a causal relationship, such a relationship seems a priori plausible, and as a consequence you might be inclined to conclude that the relationship exists.

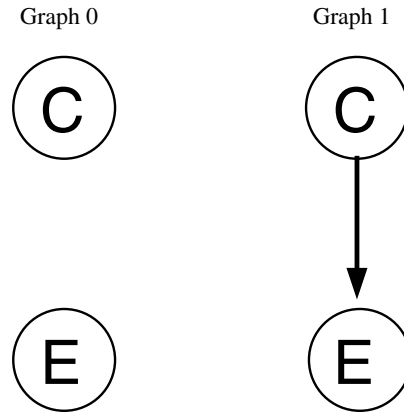


Fig. 1. Causal graphical models characterizing the relationship between two variables. *C* indicates the presence of a cause – injection of a chemical, or the thoughts of a psychic – and *E* the manifestation of the effect – a rat being born male, or a coin coming up heads. In Graph 0, cause and effect are unrelated. In Graph 1, the cause influences the effect.

Now contrast this with a different case of causal induction. A friend insists that she possesses the power of psychokinesis. To test her claim, you flip a coin in front of her while she attempts to influence the outcome. You are evaluating two hypotheses:  $h_1$  is the theory that her thoughts can influence the outcome of the coinflip, while  $h_0$  is the theory that her thoughts and the coinflip are independent. As in the previous case, these theories generate the causal graphical models Graph 1 and Graph 0 shown in Fig. 1. The first 10 flips are all heads. The likelihood ratio for these data,  $d$ , provides just as much support for a causal relationship as in the genetic engineering example, but the existence of such a relationship has lower prior probability. As a consequence, you might conclude that she does not possess psychic powers, and that the evidence to the contrary provided by the coinflips was just a coincidence.

Coincidences arise when there is a conflict between the evidence an event provides for a theory and our prior beliefs about the plausibility of that theory. More precisely, a coincidence is an event that provides support for an alternative to a current theory, but not enough support to convince us to accept that alternative. This definition can be formalized using the Bayesian machinery introduced above. Assume that  $h_0$  denotes the current theory entertained by a learner, and  $h_1$  is an alternative postulating the existence of a richer causal structure or novel causal force. In many cases of causal induction, such as establishing whether a chemical influences the sex of rats, we learn about causal relationships that seem relatively plausible, and the likelihood ratio and prior odds in favor of  $h_1$  are not dramatically in conflict. A coincidence produces a likelihood ratio in favor of  $h_1$  that is insufficient to overwhelm the prior odds against  $h_1$ , resulting in middling posterior odds. The likelihood ratio provides a measure of the *strength* of a coincidence, indicating how much support the event provides for  $h_1$ . Under this definition, the strongest coincidences can only be obtained in settings where the prior odds are equally strongly against  $h_1$ . Thus, like the test of psychokinesis, canonical coincidences typically involve data that produce a high

likelihood ratio in favor of an alternative theory in a context where the current theory is strongly entrenched.

### 3.2. *Mere and suspicious coincidences*

Up to this point, we have been relatively loose about our treatment of the term “coincidence”, relying on the familiar phenomenology of surprise associated with these events. However, when people talk about coincidences, they do so in two quite different contexts. The first is in dismissing an event as “just a coincidence”, something that is surprising but ultimately believed to be the work of chance. We will refer to these events as mere coincidences. The second context in which people talk about coincidences is when an event begins to render an alternative theory plausible. For example, Hacking’s (1983) analysis of the “argument from coincidence” focuses on this sense of coincidence, as does the treatment of coincidences in the study of vision in humans and machines (Barlow, 1985; Binford, 1981; Feldman, 1997; Knill & Richards, 1996; Witkin & Tenenbaum, 1983). We will refer to these events as *suspicious coincidences*. This distinction raises an interesting question: What determines whether a coincidence is mere or suspicious?

Under the account of coincidences outlined above, events can make a transition from coincidence to evidence as the posterior odds in favor of  $h_1$  increase. Since being considered a coincidence requires that the posterior odds remain middling, an event ceases being a coincidence and simply becomes evidence if the posterior odds increase. Consideration of the effects of the posterior odds also allows us to accommodate the difference between mere and suspicious coincidences. It is central to our definition of a mere coincidence that it be an event that ultimately results in believing  $h_0$  over  $h_1$ . Consequently, the posterior odds must be low. In a suspicious coincidence, we are left uncertain as to the true state of affairs, and are driven to investigate further. This corresponds to a situation in which the posterior odds do not favor either hypothesis strongly, being around 1 (or 0, for log posterior odds). The relationship between mere coincidences, suspicious coincidences, and unambiguous evidence for  $h_1$  is illustrated schematically in Fig. 2.

As indicated in Eqs. (3) and (4), the posterior odds in favor of  $h_1$  increase if either the prior odds or the likelihood ratio increases. Such changes can thus result in a transition from coincidence to evidence, as illustrated in Fig. 2. An example of the former was provided above: 10 male rats in a row seems like evidence in the context of a genetic engineering experiment, but 10 heads in a row is mere coincidence in a test of psychokinesis, where the prior odds are smaller.

Tests of psychokinesis can also be used to illustrate how a change in the likelihood ratio can produce a transition from mere coincidence, through suspicious coincidence, to evidence: 10 heads in a row is a mere coincidence, but twenty might begin to raise suspicions about your friend’s powers, or the fairness of the coin. At ninety heads in a row you might, like Guildenstern in Stoppard’s (1967) play, begin entertaining the possibility of divine intervention, having relatively unambiguous evidence that something out of the ordinary is taking place.

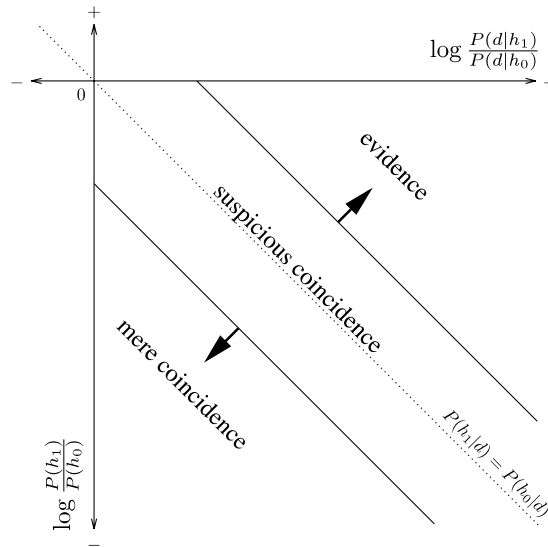


Fig. 2. The transition from coincidence to evidence can occur as a result of an increase in either the likelihood ratio or the prior odds. Between mere coincidences and evidence are suspicious coincidences, for which the posterior odds become high enough that it seems possible that  $h_1$  could actually be true.

### 3.3. Coincidences in coinflips

We have informally discussed several examples involving flipping a coin. Here, we will make these examples precise, using some tools from Bayesian statistics. This analysis helps to clarify how our framework relates to the idea that coincidences are events of unlikely kinds.

Imagine that we have two possible theories about the efficacy of psychokinesis. One theory,  $h_0$ , stipulates that there can be no relationship between thinking about a coin, and whether the coin comes up heads. Under this theory, the probability that a coin comes up heads is always 0.5. The other theory,  $h_1$ , stipulates that some people can influence the outcome of a coin toss by focussing their mind appropriately, and specifies the probability of the coin coming up heads under such influence using a parameter  $\omega$ . Given one person and one coin, each of these theories generates one causal graphical model:  $h_0$  generates Graph 0, while  $h_1$  generates Graph 1. Assume that the data,  $d$ , consist of  $N$  trials in the presence of somebody concentrating on a coin, of which  $N_H$  trials produce heads. Since  $h_0$  asserts that these outcomes are all the result of chance,  $P(d|h_0)$  is just  $\binom{N}{N_H} 0.5^N$ . Evaluating  $P(d|h_1)$  requires making assumptions about the parameter  $\omega$ .

If we assume that  $\omega = 0.9$ , indicating that we expect that a coin will come up heads far more often when it is being influenced by psychic powers,  $P(d|h_1)$  would be  $0.9^{N_H} 0.1^{N_T}$ . Consequently, a sequence like HHHH would result in a likelihood ratio  $\frac{P(d|h_1)}{P(d|h_0)}$  of  $(\frac{0.9}{0.5})^4 \approx 10.5$ , favoring  $h_1$ , while a sequence like HHTT would result

in a likelihood ratio of  $\frac{0.9^2 0.1^2}{0.5^4} \approx 0.13$ , favoring  $h_0$ . Thus, HHHH would constitute more of a coincidence than HHTT, since it provides more evidence for the low-probability theory that psychic powers exist.

The assumption that a coin will come up heads 90% of the time in the presence of a psychic makes a very specific assertion about the nature of psychic powers. More generally, we might believe that psychics have the ability to influence the probability that a coin comes up heads, but not have strong beliefs about the degree or direction of that influence. This can be expressed by defining a *distribution* over values of  $\omega$  associated with  $h_1, p(\omega|h_1)$ . In Appendix A we show that if it is assumed that  $\omega$  is uniformly distributed between 0 and 1, we obtain the likelihood ratio

$$\frac{P(d|h_1)}{P(d|h_0)} = \frac{2^N}{\binom{N}{N_H} (N+1)}, \quad (5)$$

which increasingly favors  $h_1$  as  $N_H$  deviates from  $N/2$ . This expression can be rewritten as

$$\frac{P(d|h_1)}{P(d|h_0)} = \frac{1}{P_{\text{kind}}(d)(N+1)},$$

where  $P_{\text{kind}}(d)$  is defined in Eq. (1), being the probability of a sequence of the same “kind” as  $d$ , where kinds of sequence are differentiated by the number of heads in the sequence. Consequently, the support for  $h_1$ , and the strength of the coincidence associated with  $d$ , will increase as the probability of a sequence of the same kind as  $d$  decreases. This is consistent with the “unlikely kinds” account of coincidences. This observation reveals why it is possible to construct examples that are broadly consistent with the “unlikely kinds” account of coincidences: it approximates the Bayesian solution to this problem.

Despite this connection, the Bayesian account overcomes many of the difficulties that limit the “unlikely kinds” account of coincidences. First, it provides a principled treatment of which kinds will be relevant to evaluating coincidences, and how they should be scored. This is a consequence of formulating the problem as a comparison of alternative causal theories: the relevant kinds of events are determined by the kinds of alternative causal theories that the observer implicitly considers. In our analysis of coinflipping, the kinds are differentiated by the number of heads in a sequence because  $h_1$  and  $h_0$  differ in the probabilities with which they predict a coin will produce heads. Considering other theories about possible causal mechanisms would lead to effectively considering other kinds of events, with prior probabilities proportional to the plausibility of the causal theories that generate those event kinds. For instance, the event-kind *alternating heads and tails* could be relevant because one could imagine how some kind of causal mechanism might produce such a pattern of events, although it may be harder to imagine (and thus receive a lower prior probability) than a mechanism that generates repeating sequences of coinflips. The event-kind *all outcomes that begin HHTHTTHTHT...* would almost never be considered, or would receive a

very low prior probability, because it is hard to imagine an alternative causal theory that would produce just sequences of this form.

More generally, since  $h_1$  and  $h_0$  are defined in terms of probability distributions, the Bayesian account extends naturally to continuous stimuli, as we will demonstrate later in the paper, unlike the “unlikely kinds” account. The formulation of the comparison of these hypotheses as a Bayesian inference also implicitly solves the problems with multiple kinds, and removes other technical problems. For example, the appearance of the  $(N+1)$  term in the denominator of Eq. (5) corrects for the fact that there are many more kinds of longer sequences when kinds are differentiated by the number of heads. This is the issue that made it possible for a sequence of the kind *fifteen heads, eight tails* to be less likely than a sequence of the kind *four heads*. Under Eq. (5), the former provides weaker support for  $h_1$  than the latter, as there are 24 kinds of sequence of length 23, and only 5 kinds of sequence of length 4.

### 3.4. Empirical predictions

Having given a precise definition of what constitutes a coincidence, we can evaluate how well this definition accords with human judgments. The Bayesian account presented above makes three clear empirical predictions. First, an event will be considered a coincidence when the likelihood ratio in favor of an alternative theory,  $h_1$ , is insufficient to overwhelm the prior odds against it. If either the likelihood ratio or the prior odds increase, it will ultimately come to be considered not a coincidence, but simply evidence for that theory. We test this prediction in Experiment 1. A second prediction is that the likelihood ratio in favor of  $h_1$  should indicate the strength of a coincidence. We extend our account to some of the more complex settings that have featured in arguments about the rationality of the human sense of coincidence, and assess the adequacy of the likelihood ratio in favor of  $h_1$  as a measure of the strength of coincidences in Experiments 2 and 3. Finally, our account predicts that assessing the strength of a coincidence is equivalent to assessing the evidence for a causal relationship. We test this prediction in Experiments 4 and 5.

## 4. The transition from coincidence to evidence

‘Well, Watson, what do you make of this?’ asked Holmes, after a long pause. ‘It is an amazing coincidence’.

‘A coincidence! Here is one of the three men whom we had named as possible actors in this drama, and he meets a violent death during the very hours when we know that drama was being enacted. The odds are enormous against its being coincidence. No figures could express them. No, my dear Watson, the two events are connected – *must* be connected. It is for us to find the connection’.

Sir Arthur Conan Doyle (1986a), *The adventure of the second stain*, p. 909.

What seems like a coincidence to one person can be considered compelling evidence by another.<sup>6</sup> In the analysis given above, whether an event is a coincidence or simply evidence for an alternative theory comes down to whether it ultimately justifies believing in that theory, the result of an interaction between likelihood ratio and prior (see Fig. 2). Holmes and Watson could thus differ in their construal of a violent death if they differed in the probabilities with which they thought such an event might arise independently or as the result of a connection to their case, or if they differed in the prior probability they assigned to the existence of such a connection.

Experiment 1 was designed to examine this transition from coincidence to evidence. The experiment used the two scenarios mentioned in our discussion of mere and suspicious coincidences – genetic engineering and psychokinesis – to assess whether people’s designation of events as a mere coincidence or evidence is affected by changes in the likelihood ratio and prior odds. If an event is judged “just a coincidence” when it provides insufficient support to overcome the prior, we should expect to see events with higher likelihood ratios considered a mere coincidence when people are evaluating claims about psychokinesis. More specifically, if people’s assessment of an event as a coincidence or evidence is based upon the posterior probability of  $h_1$ , we should expect to see a negative correlation between this posterior probability and the proportion of people who consider an event a coincidence. Since these predictions rely upon a subtle interaction between likelihood ratio and prior, they are inconsistent with accounts of coincidences that do not incorporate both of these components, such as the definition of coincidences as events of unlikely kinds.

## 5. Experiment 1

### 5.1. Method

#### 5.1.1. Participants

Participants were 101 undergraduates, participating for course credit. Of these participants, 24 were assigned to the *psychokinesis, posterior* condition, 20 to the *genetics, posterior* condition, 28 to the *psychokinesis, coincidence* condition, and 29 to the *genetics, coincidence* condition.

#### 5.1.2. Stimuli

Two basic cover stories were constructed that would allow the same data to be presented in different contexts. The data consisted of a table of frequencies that showed how many times a heads or tails (males or females) were produced from

---

<sup>6</sup> Coincidences played an important role in the “logical” method of deduction endorsed by Sherlock Holmes, with the notion appearing in 13 of his 60 published cases. Holmes was frequently able to solve mysteries by refusing to dismiss events as mere coincidences.

100 trials. These data showed 8 trials on which 47, 51, 55, 59, 63, 70, 87, and 99 heads (males) were obtained. Participants receiving the *psychokinesis* cover story saw:

A group of scientists investigating paranormal phenomena have conducted a series of experiments testing people who claim to possess psychic powers. All of these people say that they have psychokinetic abilities: they believe that they can influence the outcome of a coin toss. The scientists tested this claim by flipping a fair coin 100 times in front of each person as they focus their psychic energies. Under normal circumstances, a fair coin produces heads and tails with equal probability. The results of these experiments are shown below: the identities of the people are concealed with subject numbers, but you are given the number of times the coin came up heads or tails while that person was focusing their psychic energies.

while those receiving the *genetics* cover story saw:

A group of scientists investigating genetic engineering have conducted a series of experiments testing drugs that influence the development of rat fetuses. All of these drugs are supposed to affect the sex chromosome: they are intended to affect whether rats are born male or female. The scientists tested this claim by producing 100 baby rats from mothers treated with the drugs. Under normal circumstances, male and female rats are equally likely to be born. The results of these experiments are shown below: the identities of the drugs are concealed with numbers, but you are given the number of times male or female rats were produced by mothers treated with each drug.

These cover stories were presented with the data in a short questionnaire, together with further instructions on how to respond to the stimuli.

### 5.1.3. Procedure

Each participant received a questionnaire listing the eight target data sets in one of two random orders. Orthogonal to the manipulation of the cover story, participants either received the *posterior* or the *coincidence* instructions. The *posterior* instructions for the *psychokinesis* condition were

For each of the lines below, please rate HOW LIKELY you think it is that the person has psychic powers, taking into account the results of the experiment. Use a scale from 1 to 10, where 1 indicates NOT AT ALL LIKELY and 10 indicates EXTREMELY LIKELY.

Likewise, the instructions for the *genetics* condition were:

For each of the lines below, please rate HOW LIKELY you think it is that the drug affects the sex of rats, taking into account the results of the experiment. Use a scale from 1 to 10, where 1 indicates NOT AT ALL LIKELY and 10 indicates EXTREMELY LIKELY.



The eight sets of frequencies were accompanied by lines on which participants could write their responses. The *coincidence* instructions for the *psychokinesis* condition asked people to choose between a mere coincidence and evidence:

For each of the lines below, please decide whether you think the results for that person are JUST A COINCIDENCE, or COMPELLING EVIDENCE for them having psychic powers, by checking either the COINCIDENCE or the EVIDENCE box.

Similarly, the instructions for the *genetics* condition were:

For each of the lines below, please decide whether you think the results are JUST A COINCIDENCE, or COMPELLING EVIDENCE for that drug influencing the sex chromosome, by checking either the COINCIDENCE or the EVIDENCE box.

The eight sets of frequencies were listed with checkboxes to allow participants to indicate their responses.

## 5.2. Results and discussion

One participant in the *genetics* condition and two in the *psychokinesis* condition appeared to reverse the rating scale, and were eliminated from the analysis. The results are shown in Fig. 3.

The *posterior* ratings were subjected to a two-way between-within ANOVA examining the effects of condition (*psychokinesis*, *genetics*) and varying frequency. There was a main effect of condition ( $F(1,39) = 9.30$ ,  $MSE = 13.10$ ,  $p < .01$ ), a main effect of frequency ( $F(7,273) = 91.60$ ,  $MSE = 3.31$ ,  $p < .0001$ ), and an interaction between the two ( $F(7, 273) = 7.86$ ,  $MSE = 3.31$ ,  $p < .0001$ ). As can be seen from the figure, the rated probability of the conclusion went up as frequency increased, but did so earlier for the *genetics* than the *psychokinesis* condition. The same analysis was performed for the *coincidence* assessments, showing a main effect of condition ( $F(1, 55) = 18.78$ ,  $MSE = 0.18$ ,  $p < .0001$ ), a main effect of frequency ( $F(7, 385) = 99.01$ ,  $MSE = 0.08$ ,  $p < .0001$ ), and an interaction between the two ( $F(7, 385) = 39$ ,  $MSE = 0.08$ ,  $p < .0001$ ). These results are due to a similar pattern of effects: the proportion of cases classified as coincidences decreased as the frequency increased, but earlier for the *genetics* than the *psychokinesis* condition.

As predicted, there was a close correspondence between the proportion of cases classified as a mere coincidence and the mean posterior probability of the regular generating process, with a linear correlation of  $r = -0.98$ . In fact, points that are equivalent in posterior probability are also equivalent in the proportion of cases that were classified as coincidences. Examining Fig. 3 closely, it can be seen that 87 heads and 63 males produce the same results in both graphs, as do 63 heads and 59 males, and 99 heads and 70 males. This relationship holds despite the fact that responses were binary in one condition and continuous in the other, and obtained from completely different participants.

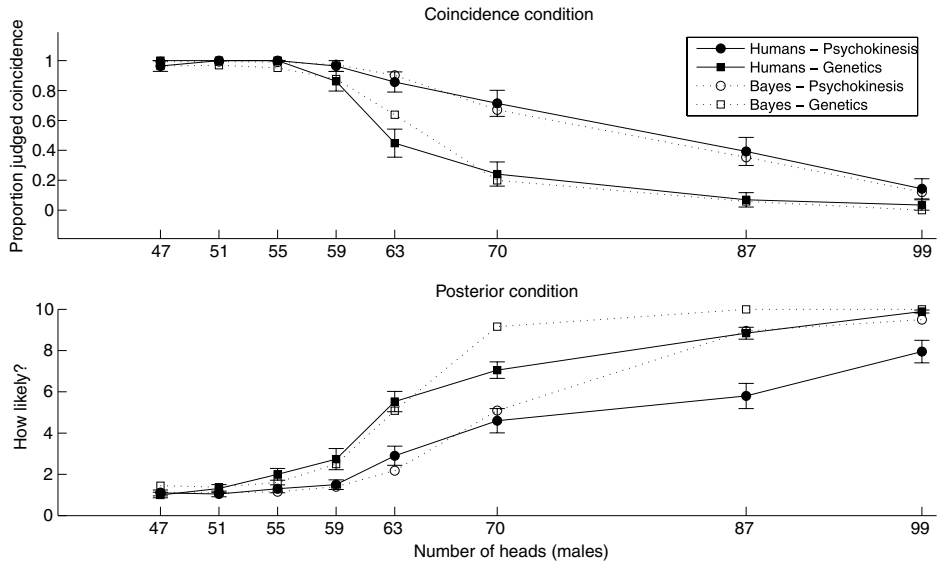


Fig. 3. Results of Experiment 1. The upper panel shows the proportion of cases judged to be coincidences in the *coincidence* condition, and the lower panel shows the mean responses in the *posterior* condition. Dotted lines show model predictions, obtained by estimating prior probabilities for each participant.

The assumption that there is a threshold on the posterior odds that determines whether an event is a coincidence or evidence, as indicated in Fig. 2, suggests that these judgments might be modeled using a sigmoid (logistic) function of the posterior odds,

$$P(\text{“evidence”} | d) = \frac{1}{1 + \exp \left\{ -g \log \frac{P(h_1 | d)}{P(h_0 | d)} - b \right\}}, \quad (6)$$

where  $g$  is the gain of the sigmoid, and  $b$  is the bias. As  $g \rightarrow \infty$ , this becomes a step function at the point  $b$ . We will assume that  $g = 1$  and  $b = 0$ , meaning that  $P(\text{“evidence”} | d)$  is equal to  $P(h_1 | d)$ . Since the likelihood ratio  $\frac{P(d | h_1)}{P(d | h_0)}$  is given by Eq. (5), we can estimate the prior odds for each participant by fitting the sigmoid function to their responses, and thus obtain the prior  $P(h_1)$ .

In the *coincidence* condition, all but one of the participants responded in a fashion consistent with thresholding the posterior odds. It was thus simple to find the value of the prior odds for each participant that maximizes the probability of their responses as predicted by Eq. (6). This results in a model fit for each participant, and the quality of these fits can be seen from the mean model predictions shown in the upper panel of Fig. 3. The median values of  $P(h_1)$  for the *psychokinesis* and *genetics* conditions were 0.0004 and 0.23, respectively.

A similar procedure can be used to estimate the prior odds directly from the posterior probabilities provided by the participants in the *posterior* condition. Again fitting a sigmoid function for each participant, this time relative to the squared error,

we obtain the fits shown in the lower panel of Fig. 3. People's more extreme probability judgments can be seen to be more conservative than those predicted by our Bayesian model, consistent with previous research (e.g., Edwards, 1968). However, this procedure yields similar median values for  $P(h_1)$ : 0.0011 in the *psychokinesis* condition and 0.20 in the *genetics* condition. Contrary to previous results illustrating deficits in the ability to combine likelihood ratios with prior odds (e.g., Kahneman & Tversky, 1972), people seem quite accurate in assessing the posterior probabilities of causal relationships. This may be a consequence of using priors that are derived from extended experience, rather than base-rates provided in an experimental scenario (Evans, Handley, Over, & Perham, 2002).

The results of this experiment are consistent with the predictions of our Bayesian account of coincidences. Data that provided the same support for  $h_1$  were judged to be coincidences if presented as the results of a test of psychokinesis, and evidence if presented as the results of a test of genetic engineering. The proportion of people who considered an event a coincidence showed a direct correspondence to the posterior probability, with the difference between the two conditions resulting from a difference in the prior probability of a causal relationship. Assuming that people are accurately evaluating the likelihood ratio in favor of  $h_1$  allows us to assess the values of these prior beliefs, which are consistent across experimental procedures and with our intuitions about the efficacy of psychic powers and genetic engineering.

## 6. The strength of coincidences

Experiment 1 suggests that the basic constituents of our definition of coincidences are correct: that events are considered a coincidence when they provide support for an alternative theory that is insufficient to convince us of its truth. We can now examine these constituents more carefully. Under this account of coincidences, the likelihood ratio indicates the strength of a coincidence, with higher likelihood ratios indicating more compelling coincidences. In the analysis given in the previous section, we assumed that the likelihood ratio given in Eq. (5) accurately captured people's assessment of the support that  $d$  gave for  $h_1$  over  $h_0$ . Whether people's assessment of the strength of coincidences corresponds to the likelihood ratio in favor of  $h_1$  more generally is an empirical question.

In exploring this question, we have the opportunity to examine people's assessment of coincidences in more realistic settings. The simplicity of coinflipping makes it an effective example with which to explore formal models, but real coincidences, such as the bombing of London, often involve more complex data and more elaborate theories. In these cases, detecting a coincidence does not just involve recognizing an unusual pattern, but doing so despite the presence of some observations that do not express that pattern. These sophisticated inductive inferences have parallels in other aspects of cognition. For example, many problems that arise in cognitive development have exactly this character, requiring a child to notice a regularity that is expressed in only a subset of the data. One such case is word learning: young children are able to learn the relationship between the use of words and the

appearance of the objects they identify, despite the fact that only about 70% of the uses of a word by a parent occur when the child is attending to the relevant object (Collins, 1977; Harris, Jones, & Grant, 1983).

We examined people’s judgments about the strength of coincidences from two different kinds of data: spatial data, consisting of the locations of bombs, and temporal data, concerning the dates of birthdays. These two cases have connections to two of the most prominent examples that are used to argue for the irrationality of human reasoning about coincidences: the bombing of London and the “birthday problem”. In each case, we investigated how well show that people’s assessment of the strength of coincidences corresponds with the rational predictions of the Bayesian account developed above.

## 7. Coincidences in space

John Snow’s inference to the cause of the Broad Street cholera outbreak and the mistaken beliefs of the populace during the bombing of London were both based upon coincidences in space —clusters in the locations of patients and bombs respectively. We will focus on coincidences that arise from patterns of bombing, looking at a measure of the strength of coincidences based upon two simple theories of bombing. Under the first theory,  $h_0$ , each bomb has its own target at a location  $L_i$ . Under the second theory,  $h_1$ , the target of each bomb is determined probabilistically: with probability  $\alpha$ , the bomb is aimed at a common target at a location  $L_c$ ; with probability  $1 - \alpha$ , the bomb has its own target, at a location  $L_i$ . We will assume that the point at which a bomb explodes has a Gaussian distribution around the location of its target, with covariance matrix  $\Sigma$ , and that targets are distributed uniformly throughout the region in which bombs fall,  $\mathcal{R}$ .

The theory  $h_0$  generates only one causal graphical model, denoted Graph 0 in Fig. 4. In this model, each bomb has a single target, and the points at which the bombs explode are independent. Using  $X_i$  to indicate the point at which the  $i$ th bomb explodes,  $d = \{x_1, \dots, x_{N_B}\}$ , where  $N_B$  is the number of bombs. In the Appendix A we show that the probability that a bomb lands in a particular location under  $h_0$  is approximately uniform over  $\mathcal{R}$ , as illustrated schematically in Fig. 5, with the likelihood for  $h_0$  being

$$P(d|h_0) \approx \left( \frac{1}{|\mathcal{R}|} \right)^{N_B}, \quad (7)$$

where  $|\mathcal{R}|$  is the area of  $\mathcal{R}$ .

The theory  $h_1$  generates  $2^{N_B}$  causal graphical models, corresponding to each partition of  $N_B$  bombs into two sets, one in which each bomb has a unique target and one in which each bomb shares a common target. Two causal graphical models generated by this theory with  $N_B = 6$  are shown in Fig. 4. Evaluating  $P(d|h_1)$  requires summing over all of these different causal models, a procedure that is discussed in

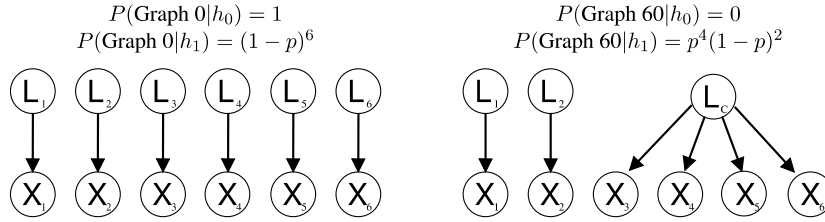


Fig. 4. Causal graphical models generated by theories of bombing.  $L_i$  indicates the location of the  $i$ th target,  $X_i$  indicates the point at which the  $i$ th bomb explodes, and  $L_c$  is the location of a target common to a subset of the bombs.

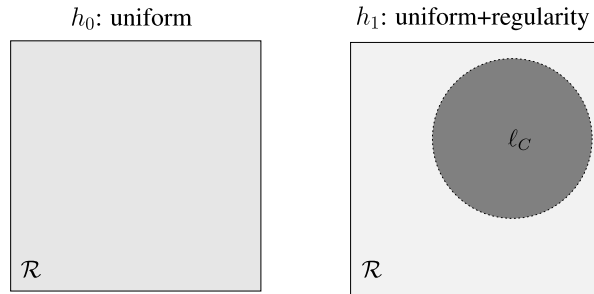


Fig. 5. Distributions over  $X_i$ , the location of the  $i$ th bomb, under theories  $h_0$  and  $h_1$ . Under  $h_0$ , the distribution is uniform over the region of interest  $\mathcal{R}$ , in this case a square indicating the city of London. Under  $h_1$ , the distribution is a mixture of a uniform distribution and a Gaussian regularity centered on the location of the common target,  $\ell_c$ . Here, the Gaussian is illustrated schematically using a circle.

Appendix A. Evaluating this probability is facilitated by the fact that  $h_1$  implies that each  $X_i$  is drawn from a *mixture* of a uniform and a Gaussian, giving

$$\begin{aligned}
 P(d|\Sigma, p, \ell_c) &= \prod_{i=1}^{N_B} [P(x_i|L_i \rightarrow X_i)P(L_i \rightarrow X_i|\alpha) + P(X|\Sigma, \ell_c, L_c \rightarrow X_i)P(L_c \rightarrow X_i|\alpha)] \\
 &= \prod_{i=1}^{N_B} \left[ \frac{1 - \alpha}{|\mathcal{R}|} + \alpha \phi_{\Sigma}(x_i, \ell_c) \right],
 \end{aligned}$$

where  $P(L_i \rightarrow X_i|\alpha)$  is the probability that  $b_i$  has a unique target,  $P(L_c \rightarrow X_i|\alpha)$  is the probability that  $b_i$  shares the common target, and  $\phi_{\Sigma}(x_i, \ell_c)$  is the probability of  $x_i$  under a Gaussian distribution with mean  $\ell_c$  and covariance matrix  $\Sigma$ . Each of these possibilities implies a different distribution for  $X_i$ , being uniform and Gaussian respectively, and their probabilities provide the weights with which these distributions are mixed, being  $1 - \alpha$  and  $\alpha$ , respectively. The resulting mixture distribution is illustrated schematically in Fig. 5. Computing  $P(d|h_1)$  reduces to the problem of computing the marginal probability of data under a mixture distribution, a problem that has been studied extensively in Bayesian statistics (e.g., Emond, Raftery, & Steele, 2001).

Eq. (7) and the procedure described in the Appendix A provide us with the means of computing  $P(d|h_0)$  and  $P(d|h_1)$ , the basic constituents of the likelihood ratio indicating the support that data  $d$  provide for  $h_1$ . Experiment 2 was designed to investigate how well this quantity predicts people’s assessment of the strength of coincidences in bombing. Participants were informed that  $h_0$  was in fact the correct account of the data, meaning that any support for  $h_1$  would constitute a coincidence. Under the Bayesian model outlined above, people’s assessment of the strength of coincidences should be strongly affected by the properties of the data  $d$ . In particular, the statistical evidence in favor  $h_1$  will be increased by the number of bombs that appear in a cluster, in both absolute and relative numbers. The location and size of the cluster should have weaker effects. We constructed a set of stimuli that varied along these dimensions, and examined whether people’s judgments of the strength of coincidences demonstrated the predicted sensitivity to these statistical properties.

## 8. Experiment 2

### 8.1. Method

#### 8.1.1. Participants

Participants were 235 undergraduates, participating for course credit.

#### 8.1.2. Stimuli

Stimuli were 12 images containing points at different locations within a 10 by 10 square, ranging from  $-5$  to  $5$  in two directions. No markers on the axes indicated this scale, but we provide the information to give meaning to the parameters listed below. Nine of these stimuli were generated from a mixture of a uniform and a Gaussian distribution, with parameters selected to span four different dimensions – number of points, proportion of points within the cluster, location of the cluster, and spread of the cluster.

The basic values of the parameters used in generating the stimuli were  $N_B = 50$ ,  $\alpha = 0.3$ ,  $\ell_c = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ , and  $\Sigma = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$ , which were varied systematically to produce the range of stimuli described above. The parameter values used to generate these stimuli are given in Table 1. The other three stimuli were generated by sampling 50 points from the uniform distribution. All 12 images are shown in Fig. 6, with repetition of the stimulus embodying the basic parameter values accounting for the presence of 15 images in the Figure. The stimuli were delivered in a questionnaire.

#### 8.1.3. Procedure

Participants completed the questionnaire as part of a booklet of other short psychology experiments. Each participant saw all 12 images, in one of six random orders. The instructions on the questionnaire read as follows:

Table 1  
Parameters used in generating the stimuli for experiment 2

Property	Parameters		
Number	$N_B = 20$	$N_B = 50$	$N_B = 200$
Proportion	$\alpha = 0.5$	$\alpha = 0.3$	$\alpha = 0.1$
Location	$\ell_c = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$	$\ell_c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\ell_c = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$
Spread	$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$	$\Sigma = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$	$\Sigma = \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{5} \end{bmatrix}$

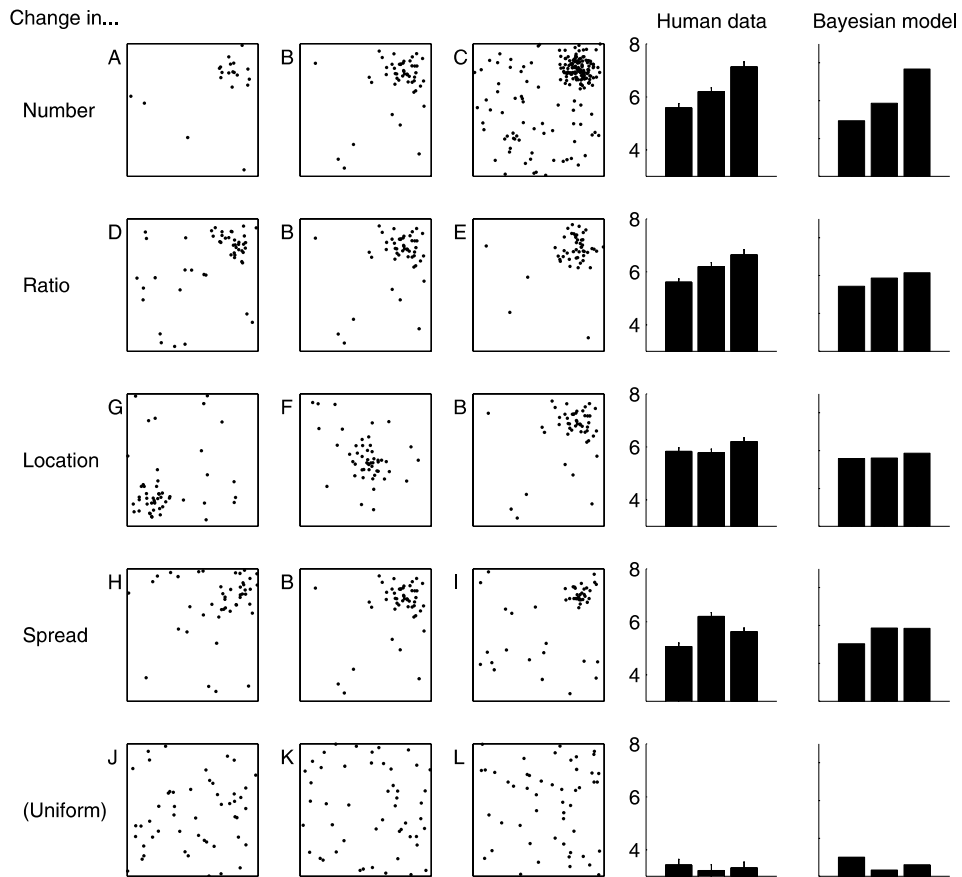


Fig. 6. Results of Experiment 2. Each line shows the three stimuli used to test the effects of manipulating one of the statistical properties of the stimulus, together with the mean judgments of strength of coincidences from human participants and the predictions of the Bayesian model. Error bars show one standard error, and letters label the different stimuli.

During World War II, the city of London was hit repeatedly by German bombs. While the bombs were found to be equally likely to fall in any part of London, people in the city believed otherwise.

Each of the images below shows where bombs landed in a particular part of London for a given month, with a single point for each bomb. On the lines at the bottom of the page corresponding to each image, please rate HOW BIG A COINCIDENCE the distribution of bombs seems to you. Use a scale from 1 to 10, where 1 means ‘Very small (or no) coincidence’, and 10 means ‘Very big coincidence’.

The images were labelled with alphabetical letters, and correspondingly labelled lines were provided at the bottom of the questionnaire for responses.

## 8.2. Results and discussion

The mean responses are shown in Fig. 6. Planned comparisons were computed for each of the manipulated variables, with statistically significant outcomes for number ( $F = 22.89$ ,  $p < .0001$ ), proportion ( $F = 10.18$ ,  $p < .0001$ ), and spread ( $F = 12.03$ ,  $p < .0001$ ), and a marginally significant effect of location ( $F = 2.0$ ,  $p = 0.14$ ). The differences observed among responses to the three sets of points generated from the uniform distribution were not statistically significant ( $F = 0.41$ ,  $p = 0.66$ ). All planned comparisons had  $df = 2,2574$ , and  $MSE = 6.21$ .

Values of  $\frac{P(d|h_1)}{P(d|h_0)}$ , were computed for each image using the method outlined in Appendix A. The predictions of the Bayesian model are shown in Fig. 6. The ordinal correlation between the raw statistical evidence and the responses was  $p = 0.965$ . The values shown in the figure are a result of the transformation  $y = \text{sign}(x)\text{abs}(x)^\gamma$  for  $x = \log \frac{P(d|h_1)}{P(d|h_0)}$  and  $\gamma = 0.32$ , which gave a linear correlation of  $r = 0.981$ .<sup>7</sup> People’s assessment of the strength of coincidences shows a remarkably close correspondence to the predictions of this Bayesian account. The main discrepancy is an overestimate of the effect of strength of coincidence for the stimulus with the least spread. This may have been a consequence of the fact that the dots indicating the bomb locations overlapped in this image, making it difficult for participants to estimate the number of bombs landing in the cluster.

## 9. Coincidences in date

How often have you been surprised to discover that two people share the same birthday? Matching birthdays are a canonical form of coincidence, and are often used to demonstrate errors in human intuitions about chance. The “birthday prob-

<sup>7</sup> We will use this non-linear scaling transformation wherever we convert a log likelihood ratio into predictions of human judgments, to accommodate the possibility that log likelihood ratio should not be mapped linearly onto the rating scale used in the experiment. The parameter of the transformation,  $\gamma$ , will be selected to maximize the fit between model and data. The same non-linear function was used by Griffiths and Tenenbaum (2005) to map log likelihood ratios to causal judgments.



lem” – evaluating the number of people that need to be in a room to provide a 50% chance of two sharing the same birthday – is a common topic in introductory statistics classes, since students are often surprised to discover that the answer is only 23 people. In general, the number of people required to have a 50% chance of a match on a variable with  $k$  alternatives is approximately  $\sqrt{k}$ , since there are  $\binom{N_p}{2} \approx N_p^2$  opportunities for a match between  $N_p$  people. Using a set of problems of this form that varied in  $k$ , Matthews and Blackmore (1995) found that people expect  $N_p$  to increase linearly with  $k$ , explaining why such problems produce surprising results. Diaconis and Mosteller (1989) argued that many coincidences are of similar form to the birthday problem, and that people’s faulty intuitions about such problems are one source of errors in reasoning about coincidences.

In this section, we will examine how people evaluate coincidences in date, through a novel “birthday problem”: assessing how big a coincidence it would be to meet a group of people with a particular set of birthdays. In contrast with the tasks that have been used to argue that coincidences are an instance of human irrationality, this is not an objective probability judgment. It is a subjective response, asking people to express their intuitions. In many ways, this is a more natural task than assessing the probability of an event. It is also, under our characterization of the nature of coincidences, a more useful one: knowing the probability of an very specific event, such as meeting people with certain birthdays, is generally less useful than knowing how much evidence it provides for the theory that a causal process was responsible for bringing that event about. By examining the structure of these subjective responses, we have the opportunity to understand the principles that guide them.

Imagine you went to a party, and met people with a set of birthdays such as {August 3, August 3, August 3, August 3}. Assume we have two possible theories that could explain this event. One theory,  $h_0$ , asserts that the presence of people at the party is independent of their birthday. This theory generates one causal graphical model for any number of people  $N_p$ , which is denoted Graph 0 in Fig. 7. The other theory,  $h_1$ , suggests that, with probability  $\alpha$ , the presence of a person at the party was dependent upon that person’s birthday. As with the theory of bombing presented above, this theory generates  $2^{N_p}$  causal graphical models for  $N_p$  people, consisting of all partitions of those people into subsets whose presence either depends or does not depend upon their birthday. Fig. 7 shows two causal graphical models generated by  $h_1$  with  $N_p = 6$ . A priori,  $h_0$  seems far more likely than  $h_1$ , so a set of birthdays that provides support for  $h_1$  constitutes a coincidence.

The data  $d$  in this setting consists of the birthdays of the people encountered at the party. Since only the people present at the party can be encountered, these are *conditional* data. If  $B_i$  indicates the birthday of the  $i$ th person and  $P_i$  indicates the presence of that person at the party, our data are the values of  $B_i$  conditioned on  $P_i$  being positive for all  $i$ . Under  $h_0$ ,  $B_i$  and  $P_i$  are independent and  $B_i$  is drawn uniformly from the set of 365 days in the year, as illustrated in Fig. 8, so we have

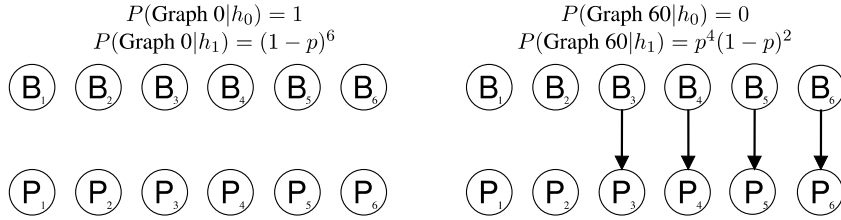


Fig. 7. Causal graphical models generated by theories of birthdays.  $B_i$  indicates the birthday of the  $i$ th person, and  $P_i$  indicates their presence at a party.

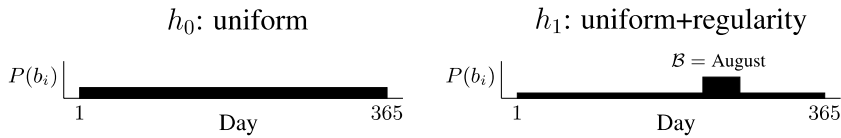


Fig. 8. Distributions over  $B_i$ , the birthday of the  $i$ th person, given their presence at a party ( $P_i = 1$ ) under theories  $h_0$  and  $h_1$ . Under  $h_0$ , the birthday is chosen from a uniform distribution over days of the year. Under  $h_1$ , the birthday is chosen from a mixture of a uniform distribution and a regularity, a uniform distribution over some subset of the year. The particular subset is determined by the filter set  $\mathcal{B}$ , which in this case corresponds to all birthdays in August.

$$P(d|h_0) = \left(\frac{1}{365}\right)^{N_p^+}, \tag{8}$$

where  $N_p^+$  is the number of people who are present at the party.

Evaluating  $P(d|h_1)$  is slightly more complicated, due to the possible dependence of  $B_i$  on  $P_i$  and the functional form of that dependence. We need to specify how people’s birthdays influenced their presence at the party. A simple assumption is that there is a “filter” set of birthdays,  $\mathcal{B}$ , and only people whose birthdays fall within that set can be present. As a first step towards evaluating  $P(d|h_1)$ , we can consider the probability of  $d$  conditioned on a particular filter. There are two possibilities for the component of the causal structure that corresponds to each person: with probability  $1 - p$ ,  $B_i$  and  $P_i$  are independent, and with probability  $\alpha$ ,  $B_i$  and  $P_i$  are dependent. If  $B_i$  and  $P_i$  are independent, the probability of  $B_i$  conditioned on  $P_i$  is just the unconditional probability of  $B_i$ , which is uniform over  $\{1, \dots, 365\}$ . If  $B_i$  and  $P_i$  are dependent, the distribution of  $B_i$  conditioned on  $P_i$  is uniform over the set  $\mathcal{B}$ , since  $P_i$  has constant probability when  $B_i \in \mathcal{B}$  and zero probability otherwise. It follows that the probability distribution for each  $B_i$  conditioned on  $P_i$  being positive is a mixture of two uniform distributions, and

$$P(d|\mathcal{B}) = \prod_{i=1}^{N_p^+} \left[ \frac{1 - \alpha}{365} + I(b_i \in \mathcal{B}) \frac{\alpha}{|\mathcal{B}|} \right], \tag{9}$$

where  $I(\cdot)$  is an indicator function that takes the value 1 when its argument is true and 0 otherwise, and  $|\mathcal{B}|$  is the number of dates in  $\mathcal{B}$ . The nature of this mixture distribution is illustrated schematically in Fig. 8.

We can use Eq. (9) to compute  $P(d|h_1)$ . If we define a prior,  $P(\mathcal{B})$  on filter  $\mathcal{B}$ , we have

$$P(d|h_1) = \sum_{\mathcal{B}} P(d|\mathcal{B})P(\mathcal{B}). \quad (10)$$

The extent to which a set of birthdays will provide support for  $h_1$  will thus be influenced by the choice of  $P(\mathcal{B})$ . We want to define a prior that identifies a relatively intuitive set of filter that might be applied to a set of birthdays to determine the presence of people at a party. An enumeration of such regularities might be: falling on the same day, falling on adjacent days, being from the same calendar month, having the same calendar date (e.g., January 17, March 17, September 17, December 17), and being otherwise close in date. With 365 days in the year, these five categories identify a total of 11,358 different filters  $\mathcal{B}$ : 365 consisting of a single day in the year, 365 consisting of neighboring days, 12 consisting of calendar months, 31 consisting of specific days of the month, and 10,585 having to do with general proximity in date (from 3 to 31 days). This is not intended to be an exhaustive set of the kinds of regularities one could find in birthdays, but is a simple choice for the values that  $\mathcal{B}$  could take on that allows us to test the predictions of the model. Given this set, we will define a prior,  $P(\mathcal{B})$ , by taking a uniform distribution over the filters in the first four categories, and giving all 10,585 filters in the fifth category as much weight as a single filter in one of the first four. Eq. (10) can then be evaluated numerically by explicitly summing over all of these possibilities.

The second term in Eq. (9) has an important implication: the influence of a filter  $\mathcal{B}$  on the assessment of a coincidence decreases as that filter admits more dates. Thus, while the set {August 3, August 3, August 3, August 3} consists of birthdays that all occur in August, the major contribution to the support for  $h_1$  having been responsible for producing this outcome is the fact that all four birthdays fall on the same day. This sensitivity to the size of the filter  $\mathcal{B}$  is equivalent to the “size principle” that plays a key role in Bayesian models of concept learning and generalization (Tenenbaum, 1999a, 1999b; Tenenbaum & Griffiths, 2001). The filtering procedure by which people come to be present at the party under  $h_1$  is one means of deriving this size principle.

We can use Eqs. (8) and (10) to compute the likelihood ratio  $\frac{P(d|h_1)}{P(d|h_0)}$  for any set of birthdays. Experiment 3 compared this likelihood ratio with human ratings of the strength of coincidence for different sets of birthdays. The key prediction is that sets of birthdays corresponding to small filters will constitute strong coincidences.

## 10. Experiment 3

### 10.1. Method

#### 10.1.1. Participants

Participants were 93 undergraduates, participating for course credit.

*10.1.2. Stimuli*

Stimuli were sets of dates, chosen to allow assessment of the degree of coincidence associated with some of the regularities enumerated above. Fourteen potential relationships between birthdays were examined, using two choices of dates. The sets of dates included: 2, 4, 6, and 8 apparently unrelated birthdays for which each date was chosen from a different month, 2 birthdays on the same day, 2 birthdays in 2 days across a month boundary, 4 birthdays on the same day, 4 birthdays in one week across a month boundary, 4 birthdays in the same calendar month, 4 birthdays with the same calendar dates, and 2 same day, 4 same day, and 4 same date with an additional 4 unrelated birthdays, as well as 4 same week with an additional 2 unrelated birthdays. These dates were delivered in a questionnaire. One of the two choices of dates, in the order specified above, was

February 25, August 10  
 February 11, April 6, June 24, September 17  
 January 23, February 2, April 9, July 12, October 17, December 5  
 February 22, March 6, May 2, June 13, July 27, September 21, October 18,  
 December 11  
 May 18, May 18  
 September 30, October 1  
 August 3, August 3, August 3, August 3  
 June 27, June 29, July 1, July 2  
 January 2, January 13, January 21, January 30  
 January 17, April 17, June 17, November 17  
 January 12, March 22, March 22, July 19, October 1, December 8  
 January 29, April 26, May 5, May 5, May 5, May 5, September 14, November 1  
 February 12, April 6, May 6, June 27, August 6, October 6, November 15,  
 December 22  
 March 12, April 28, April 30, May 2, May 4, August 18

*10.1.3. Procedure*

Participants completed the questionnaire as part of a booklet of other short psychology experiments. Each participant saw one choice of dates, with the regularities occurring in one of six random orders. The instructions on the questionnaire read as follows:

All of us have experienced surprising events that make us think ‘Wow, what a coincidence’. One context in which we sometimes encounter coincidences is in finding out about people’s birthdays. Imagine that you are introduced to various groups of people. With each group of people, you discuss your birthdays. Each of the lines below gives the birthdays of one group, listed in calendar order.

Please rate how big a coincidence the birthdays of each group seem to you. Use a scale from 1 to 10, where 1 means ‘Very small (or no) coincidence’, and 10 means ‘Very big coincidence’.

The sets of dates were then given on separate lines, in calendar order within each line, with a space beside each set for a response.

## 10.2. Results and discussion

The mean responses for the different stimuli are shown in Fig. 9. The birthdays differed significantly in their judged coincidentalness ( $F(13,1196) = 185.55$ ,  $MSE = 3.35$ ,  $p < .0001$ ). The figure also shows the predictions of the Bayesian model. The ordinal correlation between the likelihood ratio  $\frac{P(d|h_1)}{P(d|h_0)}$  and the human judgments was  $\rho = 0.921$ . The values shown in the Figure were obtained using  $\gamma = 0.60$ , and produced a linear correlation of  $r = 0.958$ .

The predictions of the Bayesian model correspond closely to people’s judgments of the strength of coincidences. Each of the parts of this model – the size principle, the set of filter, and the prior over filter  $P(\mathcal{B})$  – contributes to this performance. Fig. 9 illustrates the contributions of these different components: the panel labelled ‘Without sizes’ shows the effect of removing the size principle; ‘Uniform  $P(\mathcal{B})$ ’ shows the effect of removing  $P(\mathcal{B})$ ; and ‘Unit weights’ shows the effect of removing both of these elements of the model and simply giving equal weight to each filter  $\mathcal{B}$  consistent with  $B_i$ . We will discuss how each of these modifications reduces the fit of the model to the data, but the basic message is clear: simply specifying a set of regularities is not sufficient to explain people’s judgments. The model explains many of the subtleties of people’s performance on this task as the result of rational statistical inference.

The ‘Without sizes’ model shown in Fig. 9 replaces the  $\frac{\alpha}{|\mathcal{B}|}$  term in Eq. (9) with just  $p$ , removing the effect of the size principle. The model fit is significantly worse, with a rank-order correlation of  $\rho = 0.12$ , and  $\gamma = 1.00$  giving a linear correlation of  $r = -0.079$ . The worse fit of this model illustrates the importance of the size of the extension of the judged event in determining the strength of a coincidence, consistent with Falk’s (1981–1982, 1989) results. This effect can be seen most clearly by examining the stimuli that consist of four dates: {August 3, August 3, August 3, August 3} is more of a coincidence than {January 17, April 17, June 17, November 17}, which is in turn more of a coincidence than {January 2, January 13, January 21, January 30}. This ordering is consistent with the size of the regularities they express: a set of four birthdays falling on August 3 cover only one date, August 3, while there are 12 dates covered by the set corresponding to dates falling on the 17th day of the month, and 31 dates covered by the set corresponding to dates in January.

The size of the extension of the set is not the only factor influencing the predictions of the Bayesian model. While the size of  $\mathcal{B}$  is important in determining  $P(d|h_1)$ , the prior  $P(\mathcal{B})$  also has a large effect. In the basic model,  $P(\mathcal{B})$  gives less weight to the extremely large number of regularities corresponding to intervals of

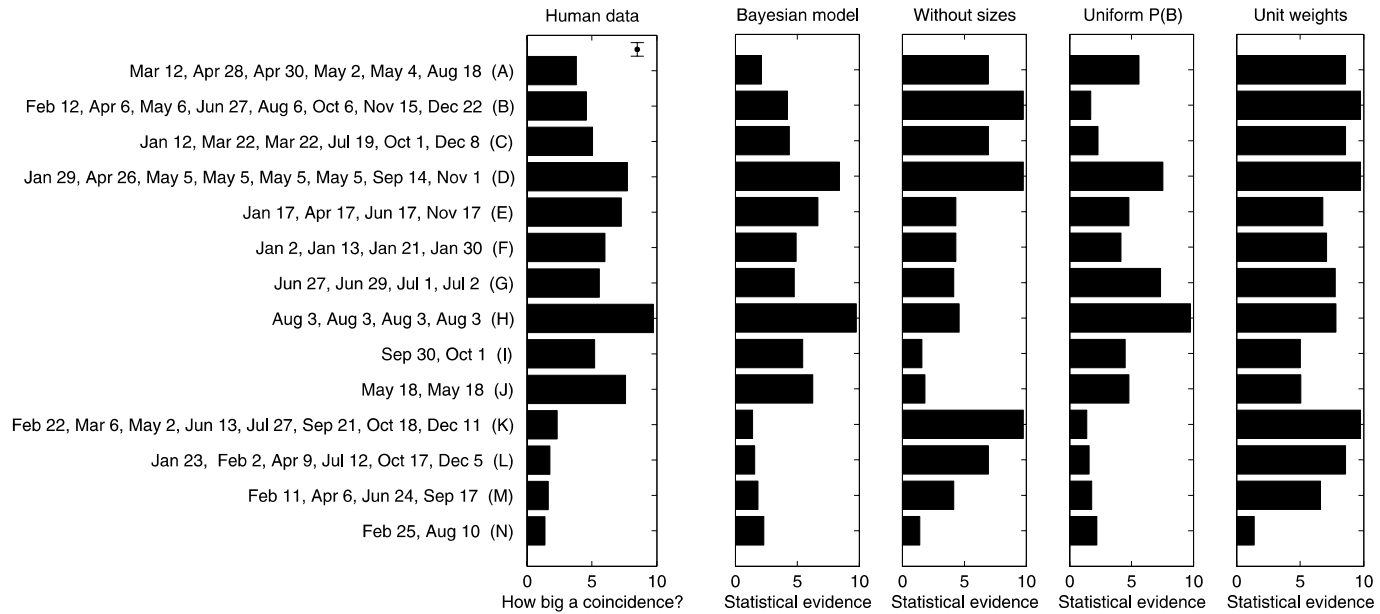


Fig. 9. The leftmost panel shows the mean judgment of the strength of coincidences from human participants in Experiment 3. Error bars indicating one standard error in either direction are shown in the upper right hand corner of the panel. The second panel shows the predictions of the Bayesian model, the third shows the consequences of removing the size principle, and the third shows the consequences of using a uniform prior on filter,  $P(\mathcal{B})$ . The fifth panel shows the combined effects of these two omissions, illustrating the performance of the model when each filter  $\mathcal{B}$  contributes equally to  $P(d|h_1)$ .

between 3 and 31 days. The importance of this prior over sets is illustrated by the “Uniform  $P(\mathcal{B})$ ” model, which gives equal probability to all of the filter  $\mathcal{B}$ . This model gives too much weight to the filter that correspond to intervals of dates, resulting in a fit of  $\rho = 0.776$ , and  $r = 0.806$  with  $\gamma = 0.80$ . The main error made by this model is not predicting the apparent equivalence of {January 17, April 17, June 17, November 17} and {June 27, June 29, July 1, July 2}, despite the fact that the former is of size 12 and the latter of size 7. In the basic model, the effect of the sizes of the regularities is overwhelmed by  $P(\mathcal{B})$ , corresponding to the fact that dates falling within seven days over a month boundary is not a particularly salient regularity.

The effects of the size principle and  $P(\mathcal{B})$  interact in producing the good performance of the basic Bayesian model. These two factors determine which regularities influence the strength of a coincidence. Simply having a sensible set of filters provides no guarantee of a good model of coincidence judgments. This can be seen in the “Unit weights” model, in which all filter  $\mathcal{B}$  are given unit weight, removing the size principle and using a uniform prior  $P(\mathcal{B})$ . The model gives a fit of  $\rho = 0.099$ , and  $r = 0.158$  with  $\gamma = 0.002$ . In this model, the major contributors to the strength of a coincidence are the number of dates and their proximity.

The main discrepancy between the basic Bayesian model and the data is the ordering of the random dates. The model predicts that the longer lists of unrelated dates should be considered less of a coincidence, while people seem to believe the opposite. To explore this curious effect further, we conducted a second survey with a separate group of 73 undergraduates, showing them a subset of 8 of the 14 stimuli used in the experiment that included the four sets of random dates. The participants were asked to rate the strength of the coincidences, as before, and to state why they gave the rating they did. Of the 73 participants, 49 did not identify any kind of pattern in the random dates, 23 noted a regularity, and one gave a high rating because of a match with her own birthday. The regularity identified by the 23 subjects had to do with the fact that the “random” birthdays were suspiciously evenly spaced throughout the year, not overlapping at all in month or date. This slight discrepancy is thus due to the fact that people are sensitive to regularities that were not included in our simple model.

## 11. Causes and coincidences

Experiments 2 and 3 show that the likelihood ratio in favor of  $h_1$  is a good predictor of people’s assessment of the strength of coincidences, as predicted by our account of coincidences. Since this likelihood ratio is intended to measure the evidence for a theory, a further prediction of our account is that the strength of coincidences should correlate with the strength of evidence for that theory in contexts where a causal relationship is more plausible. To test this hypothesis, we conducted two experiments using stimuli with the same statistical structure as those used in Experiments 2 and 3, but explicitly asking people to make judgments about the probability that a hidden cause was present. Our account predicts that people’s judgments

of the evidence for causal structure should correspond to their assessments of the strength of coincidences.

## 12. Experiment 4

### 12.1. Method

#### 12.1.1. Participants

Participants were 156 undergraduates, participating for course credit.

#### 12.1.2. Stimuli

The stimuli were those used in Experiment 2.

#### 12.1.3. Procedure

The experimental procedure was identical to that used in Experiment 2, except participants were provided with a different set of instructions. The instructions changed the context from one in which they were explicitly evaluating the strength of coincidences to one in which they were evaluating the evidence in favor of a hidden cause. The instructions read as follows:

A researcher in Madagascar is studying the effects of environmental resources on the location of lemur colonies. She has studied twelve different parts of Madagascar, and is trying to establish which areas show evidence of being affected by the distribution of resources in order to decide where she should focus her research. Each of the images below shows the locations of lemur colonies in one of the areas the researcher has studied. For each image, please rate HOW LIKELY you think it is that there is some underlying cause influencing the places where the lemurs choose to live. Use a scale from 1 to 10, where 1 means ‘very UNLIKELY to have an underlying cause’, and 10 means ‘very LIKELY to have an underlying cause’.

### 12.2. Results and discussion

As in Experiment 2, planned comparisons were computed for each of the manipulated variables, with statistically significant outcomes for number ( $F = 54.91$ ,  $p < .0001$ ), proportion ( $F = 54.27$ ,  $p < .0001$ ), location ( $F = 13.07$ ,  $p < .0001$ ) and spread ( $F = 51.10$ ,  $p < .0001$ ). The differences observed among responses to the three sets of points generated from the uniform distribution were not statistically significant ( $F = 0.64$ ,  $\rho = 0.47$ ). All planned comparisons had  $df = 2, 1705$ , and  $MSE = 3.11$ . The mean responses are shown in Fig. 10, together with the mean responses from Experiment 2. The two sets of responses are extremely similar, with a linear correlation of  $r = 0.995$  and a rank-order correlation of  $\rho = 0.993$ .



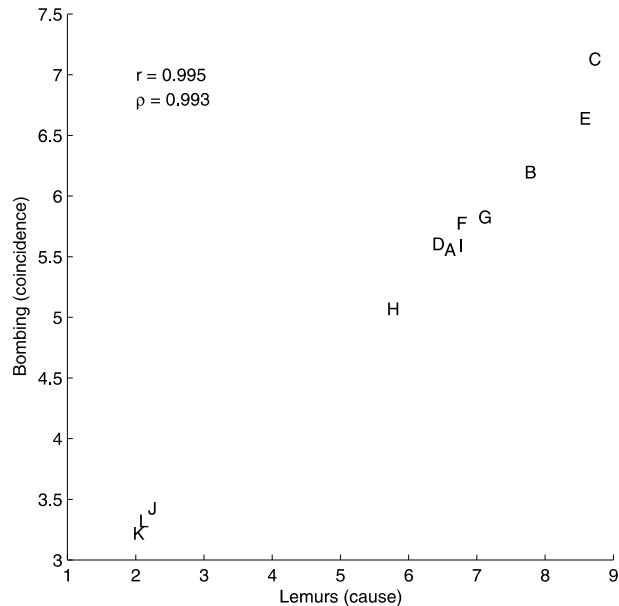


Fig. 10. Results of Experiment 4. The scatterplot indicates the close relationship between the mean ratings of the likelihood of an underlying cause behind lemur colony locations and the mean ratings of the size of a coincidence in bomb locations, using the same locations as stimuli. The letters indicate which stimulus was used, under the key from Fig. 6.

## 13. Experiment 5

### 13.1. Method

#### 13.1.1. Participants

Participants were 120 undergraduates, participating for course credit.

#### 13.1.2. Stimuli

The stimuli were those used in Experiment 3.

#### 13.1.3. Procedure

The experimental procedure was identical to that used in Experiment 3, except participants were provided with a different set of instructions. The instructions changed the context from one in which they were explicitly evaluating the strength of coincidences to one in which they were evaluating the evidence in favor of a hidden cause. The instructions read as follows:

A parcel-shipping company has been keeping meticulous records on the habits of its customer base over the past year. For each customer who sent more than one package, the company recorded the date on which each of those packages was sent. The company's marketing department is trying to figure out why

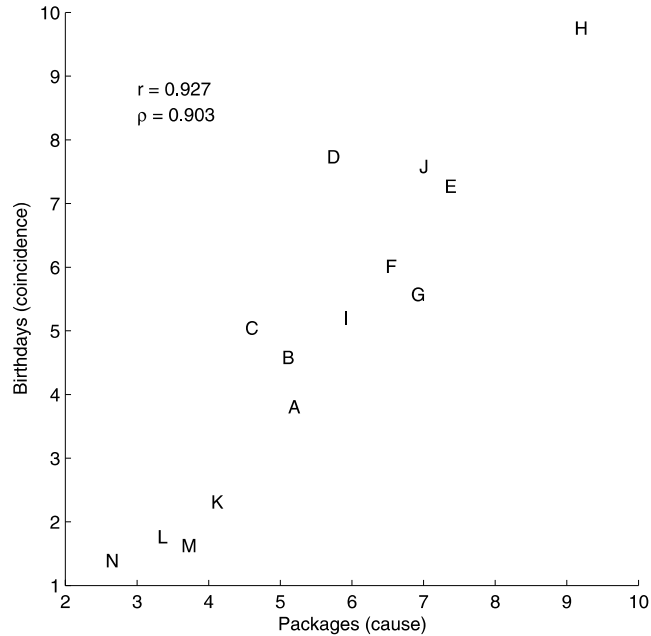


Fig. 11. Results of Experiment 5. The scatterplot indicates the close relationship between the mean ratings of the likelihood of an underlying cause behind a set of package shipments and the mean ratings of the size of a coincidence in a set of birthdays, using the same dates as stimuli. The letters indicate which stimulus is plotted, using the key from Fig. 9.

different customers shipped packages when they did. They believe that for some customers, there is some underlying cause, reason, or occasion common to some of their shipments that explains why those packages were sent on the particular days that they were. In contrast, for other customers, each package sent was independent of the others, with no common underlying cause, reason, or occasion. The shipping company would like to identify those customers whose shipments had an underlying cause, in order to offer them special discounts in the future.

The dates on which several customers sent packages are shown in calendar order below. Each set of dates corresponds to one customer's record of shipments for the year; each date corresponds to a single shipment by that customer. For each customer, please rate **HOW LIKELY** you think it is that there is some underlying cause, reason or occasion responsible for **SOME OF** their shipments. The alternative is that all the customer's shipments are independent, and none of them have a common cause. Use a scale from 1 to 10, where 1 means 'very **UNLIKELY** some shipments have an underlying cause', and 10 means 'very **LIKELY** some shipments have an underlying cause'.

### 13.2. Results and discussion

As in Experiment 3, there was an overall effect of the set of dates, ( $F(13,1547) = 36.53$ ,  $MSE = 10.55$ ,  $p < .0001$ ). The mean responses are shown in Fig. 11, plotted against the mean responses from Experiment 3. The overall pattern of responses is very similar in the two experiments, with a linear correlation of  $r = 0.927$  and an ordinal correlation of  $\rho = 0.903$ .

The only stimuli that deviate from the otherwise strong linear relationship between the results of the two experiments are C and D, which both contain a regularity together with several unrelated dates. This difference may have been a result of less willingness to accept partial regularities as indicating the presence of an underlying cause when reasoning about packages. We evaluated this hypothesis by examining the consequences of changing the assumptions about the distribution of  $\alpha$ , the proportion of dates that should be drawn from the regularity, in our Bayesian model. Assuming a uniform distribution over  $\alpha$ , as was done for the birthday data from Experiment 3, results in a linear correlation between model and data of  $r = 0.840$  with  $\gamma = 0.59$ . If instead we assume that  $\alpha$  is drawn from a distribution that is peaked near 1, we obtain a better fit to the data. For example, assuming that  $\alpha$  follows a Beta(9.99,0.01) distribution, which has a mean at  $\alpha = 0.999$ , gives  $r = 0.916$  with  $\gamma = 1$ . This is consistent with people having a stronger expectation that a causal relationship would affect all dates in the case of the packages.

## 14. General discussion

We defined a coincidence as an event that provides evidence for an alternative to a current theory, but not enough evidence to convince us to accept that alternative. More formally, a coincidence is an event where the posterior odds in favor of a hypothesis  $h_1$  over our current beliefs  $h_0$  remain middling as the consequence of a high likelihood ratio and low prior odds. This definition makes three predictions: that an event can transform from a coincidence to unambiguous evidence for an alternative theory as the prior odds or likelihood ratio increase; that the likelihood ratio indicates the strength of a coincidence; and that the strength of a coincidence should be the same as the amount of evidence that an event provides in favor of that alternative theory. Our experiments support these predictions. In Experiment 1, people's interpretation of an event as coincidence or evidence was directly affected by manipulating the prior odds and likelihood ratios of different stimuli. In Experiments 2 and 3, the likelihood ratios associated with different kinds of structure embedded in noise predicted people's judgments about the strength of coincidences. In Experiments 4 and 5, these judgments correlated almost perfectly with people's assessments of the evidence an event provided for a particular theory.

We began this paper by observing an apparent paradox associated with coincidences: that the same events seem to be involved in both our most grievous errors of reasoning, and our greatest causal discoveries. Our account of coincidences provides some insight into this paradox. Under our definition, coincidences provide an

opportunity to make a discovery that is inconsistent with our current account of how the world works. The low prior odds in favor of  $h_1$  indicates that this theory is rendered implausible by the remainder of a learner's knowledge, while the high likelihood ratio suggests that  $h_1$  should be taken seriously. The ultimate outcome of accepting the conclusion suggested by a coincidence depends on the truth of one's current theory. If one's current theory is true, then one will be led to a false conclusion. If one's current theory is false, then one might make a significant discovery.

Formulated in these terms, it becomes clear that the utility of attending to coincidences depends upon the state of our knowledge. If our understanding of the world is accurate, then coincidences can only be false alarms: cases where events that arise by chance provide support for an alternative theory,  $h_1$ . Our susceptibility to being misled by coincidences is thus partly a consequence of our success in causal discovery making one of the major sources of clues redundant. For anybody with a less accurate account of how the world works than a modern adult, such as an early scientist or a young child, coincidences are a rich source of information as to how a theory might be revised, and should be given great attention. This account also explains why many of the most compelling coincidences, such as the September 11 lottery results, are associated with mysticism. Since  $h_0$  represents the sum of our knowledge of nature,  $h_1$  will have to postulate the existence of a supernatural force.

When combined with the results of our experiments, this view of coincidences provides the opportunity to gain a deeper understanding of their role in both false conclusions and meaningful discoveries. In the remainder of the paper, we will discuss these two aspects of coincidences more detail, considering the implications of our results for claims about human rationality, and how coincidences play a role in theory change.

#### 14.1. *The locus of human irrationality*

'Singular coincidence, Holmes. Very smart of you to notice it, but rather uncharitable to suggest that it was cause and effect.'

Sir Arthur Conan Doyle (1986b), *The adventure of the dying detective*, p. 396.

In the Bayesian approach to causal induction outlined in this paper, causal inferences are the result of combining two kinds of information: the evidence that the particular data  $d$  provide for a theory; and the a priori plausibility of the existence of that structure. These two kinds of information are expressed by the likelihood ratio and the prior odds in Eqs. (3) and (4). Under this approach, three factors could lead to errors in evaluating the existence of causal structure: failing to evaluate the evidence provided by a particular event, failing to accurately assess the plausibility of the suggested theory, or failing to combine these two sources of information appropriately. Using this framework, we can ask which of these factors is responsible for the false conclusions about causal structure that people sometimes reach when they experience coincidences.

The results of the experiments presented above can be used to identify the locus of human irrationality with respect to coincidences. Experiment 1 showed that people

could integrate prior knowledge with statistical evidence appropriately in evaluating coincidences. Experiments 2 and 3 (as well as Experiments 4 and 5) indicate that people are very good at assessing the support that an event provides for a theory: the likelihood ratio in favor of  $h_1$  gave a remarkably good fit to human judgments. Thus, of the three factors that could lead to errors, one remains. Our results suggest that when people are led to believe theories that are false, they do so as a consequence of over-estimating the a priori plausibility of those theories, as reflected in the prior odds.

The suggestion that people can accurately assess the evidence that a set of events provides for a theory is consistent with some of the ideas that appear in the literature on judgment and decision making. Tversky and Koehler (1994) argued that many of the irrational aspects of people's probability judgments can be understood by viewing these judgments as reflecting the support that a set of observations provide for a particular hypothesis. In order to use this information, people have to be able to actually compute some measure of support. While various measures have been suggested, a Bayesian measure of support similar to our measure of evidence has been found to provide reasonable results on at least some cognitive tasks (Koehler, White, & Grondin, 2003). This is consistent with the results of Experiments 2 and 3 (and Experiments 4 and 5). However, accurately assessing the support for a theory does not guarantee a valid conclusion about the truth of that theory, just as accurate results from a statistical analysis do not guarantee a valid conclusion. Reaching the right conclusion requires having well-calibrated priors.

One suggestive hypothesis as to why people might over-estimate the a priori plausibility of certain theories comes from developmental psychology. Gopnik and Meltzoff (1997) argue that the scientific behavior of adults is an extension of the capacity for causal discovery that is essential for the cognitive development of children. It is quite understandable that children might be willing to believe the theories suggested by coincidences, since they are surrounded by events that really do involve novel causal relationships. Small children are justified in being conspiracy theorists, since their world is run by an inscrutable and all-powerful organization possessing secret communications and mysterious powers – a world of adults, who act by a system of rules that children gradually master as they grow up. If our scientific capacities really are for solving these childhood mysteries, then our disposition to believe in the existence of unexpected causal relationships might lag behind our current state of knowledge, leading us to see causes where none exist.

Further opportunities for erroneous inferences are provided by cases where suspicious coincidences are not tested through further investigation. If we examine the contexts in which coincidences lead people to false beliefs, we see that many of them involve situations where it is hard to conduct convincing experiments that invalidate a hypothetical causal relationship. Synchronicity, extrasensory perception, and other paranormal forces are all quite slippery subjects of investigation, for which it is challenging to construct compelling experimental tests (e.g., Diaconis, 1978). The bombing of London involved a similarly untestable hypothesis, compounded by the fear and uncertainty associated with being under attack. The cases where coincidences have resulted in rational discoveries, in science and detective stories, are all cases

where a coincidence suggests a hypothesis which can be established through further investigation. Without this kind of detailed investigation, all but the most compelling coincidences should be treated as nothing more than suspicious.

#### 14.2. Events and non-events

‘Is there any point to which you would wish to draw my attention?’

‘To the curious incident of the dog in the night-time’.

‘The dog did nothing in the night-time’.

‘That was the curious incident’, remarked Sherlock Holmes.

Sir Arthur Conan Doyle (1986b), *Silver Blaze*, p. 472.

Traditional explanations of why certain coincidental events, such as meeting an acquaintance in a distant place, should not be considered surprising focus on the fact that when we experience such events, we tend not to consider all of the other moments at which such an event could have occurred, but did not. This explanation is based upon the idea that coincidences are unlikely events: once the large number of opportunities for an improbable event to occur are taken into account, the probability that it would occur on any one of them becomes quite high, and thus we should not be surprised when such an event occurs. Our Bayesian framework can clarify this argument. In particular, it can be used to distinguish between two properties of events that are conflated in these traditional explanations: being surprising, and justifying the conclusion that a causal relationship exists.

Assume that in addition to the event  $d$ , we have a set of “non-events”  $d^*$ , which are more probable under the theory  $h_0$  than under the theory  $h_1$ . In our Bayesian framework, these non-events should influence the assessment of  $d$  as a coincidence by affecting the prior odds. Throughout the rest of the paper, we have described the prior odds as reflecting the a priori plausibility of an alternative theory. However, judgments are a priori only in the sense that they describe people’s beliefs without knowledge of  $d$  – they will still be informed by all other available evidence. The prior odds used in a Bayesian inference reflect the prior probability of two hypotheses, taking into account all sources of evidence other than the data that is being considered in that inference. Thus, if the other evidence that is available is  $d^*$ , the posterior odds in favor of  $h_1$  will be

$$\frac{P(h_1|d, d^*)}{P(h_0|d, d^*)} = \frac{P(d|h_1)P(h_1|d^*)}{P(d|h_0)P(h_0|d^*)} \quad (11)$$

where  $d$  and  $d^*$  are assumed to be independent, conditioned on  $h_1$  or  $h_0$ .

Comparing Eq. (11) with Eq. (3), the key difference is that the non-events,  $d^*$ , are taken into account in determining the prior odds when such information is available. Since  $d^*$  is more consistent with  $h_0$  than  $h_1$ , the prior odds in favor of  $h_1$  will be decreased. Under our definition of a coincidence, the event associated with  $d$  will still be considered a coincidence, provided the likelihood ratio

in favor of  $h_1$  is sufficiently high. However, taking  $d^*$  into account will result in a significant decrease in the posterior odds. Thus, the influence of many unfulfilled opportunities for an event to occur is not to decrease its potential to be surprising, but to lessen the extent to which one should believe that the suggested causal relationship actually exists.

In the previous section, we argued that our results suggest that human irrationality concerning coincidences could be localized in miscalibrated prior odds. The importance of non-events in determining these prior odds provides another explanation for why they might be miscalibrated. Detecting non-events – that is, being aware of all of the moments when an event fails to occur – requires significantly more effort than noticing that an event actually took place. Under-estimating the number of such non-events would lead to an overly permissive prior probability in favor of theories that predict novel causal structure. Thus, one reason why many people reach less rational conclusions as a result of coincidences than those drawn by Sherlock Holmes may be that, unlike Holmes, most of us fail to notice when dogs do not bark in the night-time.

#### 14.3. *Coincidences and theory change*

Many cognitive scientists have suggested that the growth and organization of knowledge can be understood by examining similar processes in scientific theories (Carey, 1985; Gopnik & Meltzoff, 1997; Karmiloff-Smith, 1988; Keil, 1989; Murphy & Medin, 1985). One of the major problems that arises in this “theory theory” is understanding the process of theory change. The formal analyses we have presented in this paper have characterized coincidences as involving data that provide support for a theory that has low a priori probability. Coincidences thus constitute an opportunity to discover that one’s current theory of how the world works is false. This characterization of coincidences suggests that they may play an important role in theory change, similar to the role of “anomalies” in accounts of scientific discovery in philosophy of science.

The theory draws extensively upon work in philosophy of science, and in particular upon Kuhn’s (1970) analysis of science in terms of a succession of scientific revolutions. One of the major topics of Kuhn’s work is the factors contributing to scientific discovery and subsequent theoretical change. Principal among these factors is the growing awareness of “anomalies”, with Kuhn (1970) claiming that ‘discovery commences with the awareness of anomaly, i.e., with the recognition that nature has somehow violated the paradigm-induced expectations that govern normal science’ (p. 52). Kuhn (1970) argued that the process of discovery often follows a particular course:

Initially, only the anticipated and usual are experienced even under circumstances where anomaly is later to be observed. Further acquaintance, however, does result in awareness of something wrong or does relate the effect to something that has gone wrong before. That awareness of anomaly opens

a period in which conceptual categories are adjusted until the initially anomalous has become the anticipated. At this point the discovery has been completed. (p. 64)

Anomalies can also be responsible for large-scale theoretical change, inducing a crisis that is resolved by the development of a new theory. However, [Kuhn \(1970\)](#) noted that ‘if an anomaly is to evoke crisis, it must usually be more than just an anomaly’ (p. 82).

Anomalous scientific results can be of two kinds. The strongest kind of anomaly is an event that is impossible under a particular scientific theory, having zero probability. Such an event contributes infinite evidence against the theory, and suggests that it should be replaced. However, most anomalies are of a different kind: events that are improbable under a theory. [Salmon \(1990\)](#) suggested that a Bayesian approach to comparing theories might be consistent with Kuhn’s characterization of theory change. Salmon characterized an anomaly as ‘a phenomenon that appears to have a small, possibly zero, likelihood given that theory’ (1990, p. 193). This assertion is similar to the claim that coincidences are unlikely events, defining anomalies only in terms of their probability under the current theory and not considering alternatives. Just as we can construct cases in which events are equally unlikely but not equally coincidental, we can construct cases in which events are equally unlikely but not equally anomalous. A full account of anomalies needs to compare this likelihood with some alternative, as in our account of coincidences.

The consistency of [Salmon’s \(1990\)](#) statistical definition of an anomaly with the accounts that appear in the literature on coincidences suggests that there may be some correspondence between the two notions. Kuhn’s informal characterization of anomalies is very similar to our intuition behind our formal definition of coincidences: anomalies are patterns of results that suggest a structure not predicted by the current theory, which can come to motivate theoretical change once sufficient evidence mounts. [Kuhn’s \(1970, p. 64\)](#) description of the process by which anomalies lead to discoveries bears a remarkable similarity to the process by which mere coincidences become suspicious. Initially, a few surprising coincidences will be dismissed as the result of chance. However, as one comes to consider the possibility of other processes being involved, and as the number of coincidences increases, the evidence provided by this set of events begins to promote suspicions. Further exploration of the source of these events might reveal an unexpected causal relationship. Once one is aware of this relationship, the events that were previously coincidences become anticipated, and merely provide further evidence for a known relationship. Likewise, the statement that crises are provoked by anomalies that are not just anomalies expresses the same sentiment as our notion of suspicious coincidences – in order to result in a change in beliefs, a coincidence must be more than just a coincidence.



## 15. Conclusion

Coincidences pose an intriguing paradox, playing key roles both in significant discoveries and in propagating false beliefs. Resolving this paradox requires going beyond the common idea that coincidences are just unlikely events, and considering their relationship to causality. We have argued that coincidences are events that naturally arise in the process of causal induction, providing support for an alternative to a current theory, but not enough support to convince us to accept that alternative. We encounter coincidences when our data provide evidence that goes against our expectations, and are central to the process of making new discoveries. By attending to coincidences, we have the opportunity to discover that our beliefs are false, and to develop more accurate theories. Our sensitivity to coincidences is not just a source of curious tales and irrational conclusions – it is one of the cognitive capacities that makes causal discovery possible, both in science and everyday life.

## Appendix A

### A.1. Bayesian coinflipping

If  $d$  is a sequence of TV coinflips producing  $N_H$  heads, then  $P(d|h_0) = \binom{N}{N_H} \left(\frac{1}{2}\right)^N$ .

To compute,  $P(d|h_1)$ , we need to define a prior distribution on  $\omega$ , the probability that a given flip comes up heads under  $h_1$ . If we define a prior distribution  $P(\omega|h_1)$ , then we can compute

$$\begin{aligned} P(d|h_1) &= \int_0^1 P(d|\omega, h_1)P(\omega|h_1)d\omega \\ &= \int_0^1 \binom{N}{N_H} \omega^{N_H} (1-\omega)^{N-N_H} P(\omega|h_1) d\omega. \end{aligned} \quad (12)$$

One possibility is to take  $P(\omega|h_1)$  to be a Beta( $r, s$ ) distribution over the range  $[0, 1]$ , with

$$P(\omega|h_1) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \omega^{r-1} (1-\omega)^{s-1},$$

where  $\Gamma(r) = \int_0^\infty x^{r-1} \exp\{-x\} dx$  is the generalized factorial function, with  $\Gamma(r) = (r-1)!$  for integer values of  $r$  (Boas, 1983). Substituting this distribution into Eq. (12), we obtain

$$\begin{aligned} P(d|h_1) &= \binom{N}{N_H} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \int_0^1 \omega^{N_H+r-1} (1-\omega)^{N-N_H+s-1} d\omega \\ &= \binom{N}{N_H} \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \frac{\Gamma(N_H+r)\Gamma(N-N_H+s)}{\Gamma(N+r+s)}. \end{aligned}$$

In the case where  $r = s = 1$ , corresponding to a uniform distribution over values of  $\omega$  between 0 and 1, this reduces to

$$P(d|h_1) = \frac{1}{N+1},$$

from which it follows that the likelihood ratio in favor of  $h_1$  is

$$\frac{P(d|h_1)}{P(d|h_0)} = \frac{2^N}{\binom{N}{N_H}(N+1)},$$

which increasingly favors  $h_1$  as  $N_H$  deviates from  $N/2$ .

### A.2. Evaluating Bayes factors for bombing

Under  $h_0$ , each bomb has its own target, and targets are distributed uniformly over the region  $\mathcal{R}$ . The location at which a bomb lands follows a Gaussian distribution, with the mean being the target of the bomb and a covariance matrix  $\Sigma$ . Using  $X_i$  to indicate the location of the  $i$ th bomb and  $L_i$  to indicate the location of the  $i$ th target, we have

$$P(x_i|h_0) = \int_{\mathcal{R}} P(x_i|\ell_i)P(\ell_i)d\ell_i = \int_{\mathcal{R}} \phi_{\Sigma}(x_i, \ell_i) \frac{1}{|\mathcal{R}|} d\ell_i \approx \frac{1}{|\mathcal{R}|},$$

where  $x_i$  is the value taken by  $X_i$ ,  $\ell_i$  is the value taken by  $L_i$ , and  $\phi_{\Sigma}(x, \ell)$  is the value of the multivariate Gaussian density with mean  $\ell$  and covariance matrix  $\Sigma$  at point  $x$ . The approximation in the last line is a consequence of the fact that if  $\ell_i$  is near the boundary of  $\mathcal{R}$ , some of the mass of  $P(x_i|\ell_i)$  will fall outside  $\mathcal{R}$ , making values of  $x_i$  near the boundary slightly less likely. This effect will be negligible if  $\mathcal{R}$  is large relative to  $\Sigma$ , so  $P(x_i|h_0)$  is well approximated by a uniform distribution over  $\mathcal{R}$ . Consequently, with  $d$  consisting of the locations of  $N_B$  bombs,  $d = \{x_1, \dots, x_{N_B}\}$ , we have

$$P(d|h_0) \approx \left(\frac{1}{|\mathcal{R}|}\right)^{N_B},$$

since the locations of the bombs are assumed to be independent.

Under  $h_1$ , the target of each bomb is determined probabilistically: with probability  $\alpha$ , the bomb is aimed at a common target at a location  $L_c$ ; with probability  $1 - \alpha$ , the bomb has its own target, at a location  $L_i$ . Computing  $P(d|h_1)$  for  $d = \{x_1, \dots, x_{N_B}\}$  requires evaluating

$$P(d|h_1) = \sum_{i=0}^{2^{N_B}-1} \left[ \int_{\mathcal{R}} \int P(d|\Sigma, \ell_c, \text{Graph } i)P(\Sigma)P(\ell_c)d\Sigma d\ell_c \right] \times \int_0^1 P(\text{Graph } i|\alpha)P(\alpha)d\alpha.$$

we will explain how this sum can be computed evaluating the bracketed term for Graph  $2^{N_B} - 1$ , in which all  $X_i$  are drawn from a single Gaussian distribution, summing over the mean  $\ell_c$  and covariance  $\Sigma$ , and then discussing how the result

can be generalized. More details on the kind of computations performed in this section can be found in [Minka \(2001\)](#).

We will assume a uniform prior on  $\ell_c$ , with  $P(\ell_c) = 1/|\mathcal{R}|$  for  $\ell_c \in \mathcal{R}$ , and an inverse Wishart prior on  $\Sigma$  with parameters  $\mathbf{I}, k$ , where  $\mathbf{I}$  is the  $d$ -dimensional identity matrix. Under this prior,

$$P(\Sigma) = \frac{1}{c_{kd}|\Sigma|^{k+d+1/2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1}) \right\},$$

where  $c_{kd} = 2^{kd/2} \pi^{d(d-1)/4} \prod_{j=1}^d \Gamma((k+1-j)/2)$ . Given  $\Sigma$  and  $\ell_c$ , we have

$$\begin{aligned} P(d|\Sigma, \ell_c, \text{Graph } 2^{N_B} - 1) &= \frac{1}{|2\pi\Sigma|^{N_B/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N_B} (x_i - \ell_c)^T \Sigma^{-1} (x_i - \ell_c) \right\} \\ &= \frac{1}{|2\pi\Sigma|^{N_B/2}} \exp \left\{ -\frac{N_B}{2} (\bar{x} - \ell_c)^T \Sigma^{-1} (\bar{x} - \ell_c) - \frac{1}{2} \text{tr}(\mathbf{S}\Sigma^{-1}) \right\}, \end{aligned}$$

where  $\bar{x}$  is  $\frac{1}{N_B} \sum x_i$  and  $\mathbf{S} = \sum_{i=1}^{N_B} (x_i - \bar{x})^T$ .

Using these definitions, we can express our integral as

$$\begin{aligned} P(d|\text{Graph } 2^{N_B} - 1) &= \int_{\mathcal{R}} \int P(d|\Sigma, \ell_c, \text{Graph } 2^{N_B} - 1) P(\Sigma) P(\ell_c) d\Sigma d\ell_c \\ &= \frac{1}{|\mathcal{R}| c_{kd} (2\pi)^{dN_B/2}} \int \frac{1}{|\Sigma|^{(N_B+k+d+1)/2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \text{tr}((\mathbf{S} + \mathbf{I})\Sigma^{-1}) \right\} \\ &\quad \times \left[ \int_{\mathcal{R}} \exp \left\{ -\frac{N_B}{2} (\bar{x} - \ell_c)^T \Sigma^{-1} (\bar{x} - \ell_c) \right\} d\ell_c \right] d\Sigma. \end{aligned}$$

The bracketed integrand has the form of a Gaussian. The result is upper bounded by  $|2\pi/N_B|^{1/2}$ , with the tightness of the bound increasing with the size of  $\mathcal{R}$  relative to  $\Sigma/N_B$ .

This reduces the outer integral to

$$\begin{aligned} P(d|\text{Graph } 2^{N_B} - 1) &\approx \frac{1}{|\mathcal{R}| c_{kd} (2\pi)^{d(N_B-1)/2} N_B^{d/2}} \int \frac{1}{|\Sigma|^{(N_B+k+d)/2}} \exp \left\{ -\frac{1}{2} \text{tr}((\mathbf{S} + \mathbf{I})\Sigma^{-1}) \right\} d\Sigma \\ &= \frac{1}{|\mathcal{R}| \pi^{d(N_B-1)/2} N_B^{d/2} |\mathbf{S} + \mathbf{I}|^{(N_B+k-1)/2}} \prod_{j=1}^d \frac{\Gamma((N_B+k-j)/2)}{\Gamma((k+1-k)/2)}, \end{aligned} \quad (13)$$

where the result follows from the fact that the integrand has the form of an inverse Wishart distribution. The expression given in Eq. (13) is a measure of the ‘‘Gaussianity’’ of  $d$ : it is the probability of  $d$  being produced from some Gaussian distribution, assessed under the priors specified on  $\ell_c$  and  $\Sigma$ .

This result can be extended to allow us to evaluate the probability of  $d$  under any other graph. For each graph, the  $X_i$  can be partitioned into two sets: those that have their own target and those that share the common target. The  $X_i$  that have their own

target each have probability  $\frac{1}{|\mathcal{A}|}$ . The probability of the  $X_i$  that share a target can be computed using Eq. (13).

The integral over  $\alpha$  is straightforward to evaluate. Taking  $P(\alpha)$  to be uniform over  $[0,1]$ , the probability of Graph  $i$ , in which  $N_B^+$  bombs share a common target and  $N_B^-$  bombs have their own targets, is

$$\begin{aligned} P(\text{Graph } i) &= \int_0^1 P(\text{Graph } i|\alpha)P(\alpha) d\alpha \\ &= \int_0^1 \alpha^{N_B^+} (1-\alpha)^{N_B^-} d\alpha \\ &= \frac{\Gamma(N_B^+ + 1)\Gamma(N_B^- + 1)}{\Gamma(N_B + 2)}, \end{aligned} \quad (14)$$

following a similar analysis to that given above for coinflipping.

Combining the values of  $P(d|\text{Graph } i)$  obtained via Eq. (13) with  $P(\text{Graph } i)$  from Eq. (14), we need only evaluate the sum over the  $2^{N_B}$  possible graph structures. This can be done by Monte Carlo simulation. The results shown in Fig. 6 were computed using a form of importance sampling designed for finding the Bayes factors of mixture distributions (Emond et al., 2001). The model predictions shown in the figure use 100,000 samples and  $k = 4$ .

## References

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum. Associated Press (September 12, 2002). N.Y. lottery drawing pops up 9-1-1.
- Barlow, H. (1985). Cerebral cortex as a model builder. In D. Rose & V. G. Dobson (Eds.), *Models of the visual cortex* (pp. 37–46). Wiley.
- Binford, T. O. (1981). Inferring surfaces from images. *Artificial Intelligence*, 17, 205–244.
- Boas, M. L. (1983). *Mathematical methods in the physical sciences* (2nd ed.). New York: Wiley.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Cheng, P. (1997). From covariation to causation: a causal power theory. *Psychological Review*, 104, 367–405.
- Clarke, R. D. (1946). An application of the Poisson distribution. *Journal of the Institute of Actuaries (London)*, 72.
- Collins, G. M. (1977). Visual co-orientation and maternal speech. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction*. London: Academic Press.
- Cook, A. (1998). *Edmond Halley: Charting the heavens and the seas*. Oxford: Clarendon Press.
- Danks, D., & McKenzie, C. R. M. (under revision). *Learning complex causal structures*.
- DeGregory, L. (September 24, 2002). 9-1-1 numerology. *St Petersburg Times*.
- Diaconis, P. (1978). Statistical problems in ESP research. *Science*, 201, 131–136.
- Diaconis, P., & Mosteller, F. (1989). Methods for studying coincidences. *Journal of the American Statistical Association*, 84, 853–861.
- Doyle, A. C. (1986a). *Sherlock Holmes: The complete novels and stories* (Vol. 1). New York: Bantam.
- Doyle, A. C. (1986b). *Sherlock Holmes: The complete novels and stories* (Vol. 2). New York: Bantam.
- Eastaway, R., & Wyndham, J. (1998). *Why do buses come in threes? The hidden mathematics of everyday life*. New York: Wiley.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment*. New York: Wiley.

- Emond, M. J., Raftery, A. E., & Steele, R. (2001). *Easy computation of Bayes factors and normalizing constants for mixture models via importance sampling* (Tech. Rep. No. 398). University of Washington.
- Evans, J. S. B. T., Handley, S. J., Over, D. E., & Perham, N. (2002). Background beliefs in Bayesian reasoning. *Memory & Cognition*, *30*, 179–190.
- Falk, R. (1981–1982). On coincidences. *Skeptical Inquirer*, *6*(2), 24–25.
- Falk, R. (1989). Judgment of coincidences: mine versus yours. *American Journal of Psychology*, *102*, 477–493.
- Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170.
- Feldman, J. (2004). How surprising is a simple pattern? Quantifying “Eureka!”. *Cognition*, *93*, 199–224.
- Feller, W. (1968). *An introduction to probability theory and its applications*. New York: Wiley.
- Fisher, R. A. (1937). *The design of experiments*. London: Oliver & Boyd.
- Franklin, J. (2001). *The science of conjecture: Evidence and probability before Pascal*. Baltimore, MD: John Hopkins University Press.
- Garner, W. R. (1970). Good patterns have few alternatives. *American Scientist*, *58*, 34–52.
- Gilovich, T. (1993). *How we know what isn't so: The fallibility of reason in everyday life*. New York: Free Press.
- Glymour, C. (1998). Learning causes: psychological explanations of causal explanation. *Minds and Machines*, *8*, 39–60.
- Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.
- Good, I. J. (1956). The surprise index for the multivariate normal distribution. *The Annals of Mathematical Statistics*, *27*, 1130–1135.
- Good, I. J. (1984). A Bayesian approach in the philosophy of inference. *British Journal for the Philosophy of Science*, 161–166.
- Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.
- Gopnik, A., & Meltzoff, A. N. (1997). *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Griffiths, T. L. (2005). *Causes, coincidences, and theories*. Unpublished doctoral dissertation, Stanford University.
- Griffiths, T. L., Baraff, E. R., & Tenenbaum, J. B. (2004). Using physical theories to infer hidden causal structure. In *Proceedings of the 26th annual meeting of the cognitive science society*.
- Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.
- Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.
- Hardy, A., Harvie, R., & Koestler, A. (1973). *The challenge of chance*. New York: Random House.
- Harris, M., Jones, D., & Grant, J. (1983). The nonverbal content of mothers' speech to infants. *First Language*, *4*, 21–31.
- Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In M. I. Jordan (Ed.), *Learning in graphical models* (pp. 301–354). Cambridge, MA: MIT Press.
- Hempel, C. G. (1966). *Philosophy of natural science*. New York: Prentice-Hall.
- Horwich, P. (1982). *Probability and evidence*. Cambridge: Cambridge University Press.
- Hughes, D. W. (1990). Edmond Halley: his interest in comets. In N. J. M. Thrower (Ed.), *Standing on the shoulders of giants: a longer view of Newton and Halley* (pp. 324–372). Berkeley and Los Angeles: University of California Press.
- Hume, D. (1739/1978). *A treatise of human nature*. Oxford: Oxford University Press.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge and Kegan Paul.
- Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge: Cambridge University Press.
- Johnson, D. (1981). *V-1, V-2: Hitler's vengeance on London*. New York: Stein & Day.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.
- Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind and Language*, *3*, 183–195.

- Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- Knill, D. C., & Richards, W. A. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.
- Koehler, D. J., White, C. M., & Grondin, R. (2003). An evidential support accumulation model of subjective probability. *Cognitive Psychology*, *46*, 152–197.
- Kubovy, M., & Gilden, D. (1991). Apparent randomness is not always the complement of apparent order. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 115–127). Washington, DC: American Psychological Association.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Lagnado, D. A., & Sloman, S. (2002). Learning causal structure. In *Proceedings of the twenty-fourth annual meeting of the cognitive science society*. Erlbaum.
- Laplace, P. S. (1795/1951). *A philosophical essay on probabilities* (F.W. Truscott & F.L. Emory, Trans.). New York: Dover.
- Littlewood, J. E. (1953). *A mathematician's miscellany*. London: Methuen.
- Lopez, F. J., Cobos, P. L., Cano, A., & Shanks, D. R. (1998). The rational analysis of human causal and probability judgment. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 314–352). Oxford: Oxford University Press.
- Matthews, R. A. J., & Blackmore, S. J. (1995). Why are coincidences so impressive?. *Perceptual and Motor Skills* *80*, 1121–1122.
- Minka, T. (2001). *Inferring a Gaussian distribution*. <<http://www.stat.cmu.edu/~minka/papers/gaussian.html>>.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*, 289–316.
- Owens, D. (1992). *Causes and coincidences*. Cambridge: Cambridge University Press.
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.
- Perrin, J. (1913/1990). *Atoms* (D.L. Hammick, Trans.). Connecticut: Ox Bow Press.
- Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.
- Redfield, J. (1998). *The Celestine vision: Living the new spiritual awareness*. Sydney, Australia: Bantam.
- Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141–1159.
- Salmon, W. C. (1990). Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In C. W. Savage (Ed.), *Scientific theories* (Vol. XIV). Minneapolis, MN: University of Minnesota Press.
- Schlesinger, G. N. (1991). *The sweep of probability*. Notre Dame, IN: University of Notre Dame Press.
- Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, *47*, Serial no. 194.
- Shultz, T. R., & Kestenbaum, N. R. (1985). Causal reasoning in children. *Annals of Child Development*, *2*, 195–249.
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 544–551.
- Snow, J. (1855). *On the mode of communication of cholera*. London: John Churchill.
- Spirtes, P., Glymour, C., & Schienens, R. (1993). *Causation prediction and search*. New York: Springer-Verlag.
- Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453–489.
- Stoppard, T. (1967). *Rosencrantz and Guildenstern are dead*. New York: Grove Press.
- Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts?. *Cognition* *87*, 55–71.
- Tenenbaum, J. B. (1999a). Bayesian modeling of human concept learning. In M. S. Kearns, S. A. Solla, & D. A. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–65). Cambridge, MA: MIT Press.
- Tenenbaum, J. B. (1999b). *A Bayesian framework for concept learning*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

- Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems 13* (pp. 59–65). Cambridge, MA: MIT Press.
- Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal induction. In *Advances in neural information processing systems 15* (pp. 35–42). Cambridge, MA: MIT Press.
- Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (in press). Intuitive theories as grammars for causal inference. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. Oxford: Oxford University Press.
- Tenenbaum, J.B., & Niyogi, S. (2003). Learning causal laws. In *Proceedings of the 25th annual meeting of the cognitive science society*. Erlbaum.
- Tversky, A., & Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review*, *101*, 547–567.
- Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher & S. J. Deny (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 1102–1107). Mahwah, NJ: Erlbaum.
- White, P. A. (1990). Ideas about causation in philosophy and psychology. *Psychological Bulletin*, *108*, 3–18.
- Witkin, A. P., & Tenenbaum, J. M. (1983). On the role of structure in vision. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 481–543). Academic Press.
- Yeomans, D. K. (1991). *Comets: a chronological history of observation, science, myth and folklore*. New York: Wiley.