Coincidence Judgment in Causal Reasoning: How Coincidental Is This?

Mark K. Johansen[a**] Magda Osman[b]

** Corresponding author

a. School of Psychology, Cardiff University, Tower Building, 70 Park Place, Cardiff, CF10 3AT, UK, JohansenM@cardiff.ac.uk

b. Centre for Biological and Experimental Psychology, School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, E1 4NS, UK, m.osman@qmul.ac.uk

Declaration of interest: none.

Abstract

Given the important conceptual connections between cause and coincidence as well as the extensive prior research on causality asking, “how causal is this?”, the present research proposes and evaluated a psychological construction of coincidentality as the answer to the question, “how coincidental is this?” Four experiments measured the judgment properties of a reasonably large set of real coincidences from an initial diary study. These judgements included coincidentality and an array of other judgments about event uncertainty, hypothesis belief and surprise as predictors of coincidentality consistent with and supporting our prior definition of coincidence (Johansen & Osman, 2015): “coincidences are surprising pattern repetitions that are observed to be unlikely by chance but are nonetheless ascribed to chance since the search for causal mechanisms has not produced anything more plausible than mere chance.” In particular, we evaluated formal models based on judgements of uncertainty, belief and surprise as predictors to develop a model of coincidentality. Ultimately, we argue that coincidentality is a marker for causal suspicion/discovery in terms of a flag that a new, unknown causal mechanism *may* be operating.

Keywords: coincidence judgment; causal reasoning; causal discovery

Coincidence Judgment in Causal Reasoning: How Coincidental Is This?

1. Introduction

"That's just a coincidence!" Coincidences are a common part of daily life and have attracted a considerable body of psychological research (e.g. Beitman, 2009; Crandall, Backstrom, Cosley, Suri, Huttenlocher, & Kleinberg, 2010; Diaconis & Mosteller, 1989, 2006; Falk & MacGregor, 1983; Falk, 1989; Griffiths & Tenenbaum, 2007; Watt, 1991). However, a substantial part of this research (see the review in Johansen and Osman, 2015) connects coincidental experiences with various kinds of irrationality (Blackmore, 1992; Blackmore & Troscianko, 1985; Hanley, 1984, 1992; Matthews & Stones, 1989; Mock & Weisberg, 1992) ties them to paranormal beliefs (Blackmore, 1984; Bressan, 2002; Brugger, Regard, Landis & Grave, 1995; Glicksohn, 1990; Houran & Lange, 1996; Tobacyk, 1995; Tobacyk & Milford, 1983) and other well-established biases, such as in probability judgments (Falk & MacGregor, 1983; Falk, 1989). The dominant emphasis in prior research on coincidences is closely aligned with the many demonstrations that human judgement and decision making are nonoptimal relative to objective normative standards.

The perspective taken in the present research is an alternative that disagrees with the bias's literature, not in terms of content, but rather in emphasis (though see the notable exceptions of Dessalles, 2008, and especially Griffiths & Tenenbaum, 2007). The emphasis here is on using human judgments about real coincidences as a way of evaluating the *psychological* mechanisms underlying causal reasoning, learning and discovery from the perspective that these mechanisms are substantively rational and adaptive. By combining the key family resemblance-properties in prior definitions of coincidence (including Coleman & Beitman, 2009; Diaconis & Mosteller, 1989; Griffiths & Tenenbaum, 2007; Henry, 1993; Mill, 1843), we proposed a new definition for the psychological concept of coincidence which provides the evaluative purpose of the present research: "coincidences are surprising pattern [co-occurrence/"co-inCIDence"] repetitions that are observed to be unlikely by chance but are nonetheless ascribed to chance since the search for causal mechanisms has not produced anything more plausible than mere chance” (Johansen & Osman, 2015, p. 34).

According to this definition, coincidences are improbable and surprising, but these are not simply synonyms for coincidence. Coincidences tend to be improbable events, but improbable events aren’t necessarily coincidences (Griffiths & Tenenbaum, 2007). For example, rolling a die six times and getting 464255 might be just as improbable as 666666 but not nearly as surprising (Maguire, Moser, Maguire & Keane, 2019) or as coincidental (Johansen & Osman, 2015). Similarly, coincidences are usually surprising (Falk, 1989; Falk & MacGregor, 1983; Johansen & Osman, 2015) but events that are surprising (e.g. an unexpected firecracker or birthday party) are not necessarily coincidences (Johansen & Osman, 2015). So in the context of coincidences, surprise and improbability are in relation to co-occurrence/co-incidence patterns that capture attention and demand to be explained.

Contextualized by this definition, and in contrast to the typical view of coincidences as indicators of biased cognition, there are two overarching mistakes/biases that can occur when reasoning about new causal mechanisms, caricatured as believers and skeptics (alternatively as “sheep” and “goats”, Bressan, 2002; Brugger, Landis & Regard, 1990; Matthews & Blackmore, 1995). The first, believer-mistake, is to infer the existence of a causal mechanism when it does not actually exist. The second, skeptic-mistake, is to infer the nonexistence of a causal mechanism which actually does exist. Further, both perspectives are potentially biases that trade off risk in the context of given data: a believer runs the risk of applying inaccurate causal mechanisms, and a skeptic runs the risk of failing to apply a causal mechanism to situations where it does actually apply. From a somewhat different perspective, the danger for both is reaching a categorical conclusion, “coincidence!” or “cause!”, too quickly. Both tend to be biased to accept the explanation in hand when they should probably consider other explanations.

The believer and skeptic biases invoke a third caricature of the curious enquirer; starting with the same perception as believers and skeptics of a surprising co-occurrence/co-incidence and its apparent improbability by chance, the curious enquirer, because of greater consideration of possible mechanisms, is slower to make a coincidence versus cause attribution and does so more tentatively. In highlighting an alternative perspective from skeptic and believer, the emphasis here is not that coincidences indicate biased cognition but rather the opposite. This alternative, enquirer perspective exposes a way of thinking about coincidences as a part of the cognitive processing that happens *before* additional evidence that will ultimately adjudicate which causal mechanisms do and don't exist. As such, we propose that coincidences are an integral, unavoidable and useful part of rational causal learning and reasoning (Griffiths & Tenenbaum, 2007; Johansen & Osman, 2015), in particular as a flag to the curious enquirer that new causal knowledge *might* be available and further evaluation *may* be warranted. So coincidence judgment, especially in answer to "how coincidental is this?", represents an under-evaluated psychological approach to understanding the adaptive cognitive mechanisms for causal reasoning.

Despite the integral conceptual connections between coincidence and cause and the extensive literature on causality judgment, which asks "how causal is this?", there has been relatively little evaluation of coincidence judgment, especially the question "how coincidental is this?", and even less emphasizing the adaptive rather biased aspects of such judgments. More specifically, there are many well-formalized mathematical models of causal-learning and reasoning (Cheng, 1997; Jenkins & Ward, 1965; Navarro & Kemp, 2017; Pacer, 2016; Pearl, 2000; Spirtes, Glymour, & Scheines, 2000; Rescorla & Wagner, 1972, Tauber Navarro, Perfors & Steyvers, 2017), but there have been very few formal models that have been specifically developed to predict coincidence judgment. The most important prior model is what we term "the Bayesian ratio model of coincidence strength"[1] (Griffiths & Tenenbaum, 2007). This model positions coincidences in the Bayesian conceptual

[1] . This is our name rather than theirs for the "Bayesian model[s]" they argued for, but we ultimately propose a model that also has Bayesian probabilities and so need a terminological distinction.

framework as applied to causal reasoning (see also Griffiths, 2017; Tauber, Navarro, Perfors & Steyvers, 2017). It has several of the same key components but also some fundamental differences to the model we develop below, so it is important to summarize this model in detail. But the common emphasis of both is on coincidences as a part of the adaptive mechanisms for causal reasoning.

1.1 Bayesian Conceptualization of Coincidences

The Bayesian ratio model of coincidence strength is directly implied by and was developed to support Griffiths and Tenenbaum's definition of coincidence as "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" (2007, p. 180). For example, a "thought-contact coincidence", thinking of contacting someone but they contact you first, does provide genuine evidence for a paranormal hypothesis of a telepathic connection. Nevertheless, most people quickly dismiss this evidence as insufficient because prior belief summarizes so much evidence against this hypothesis, that is, because the prior probability for this hypothesis is extremely small. We evaluated this definition and our proposed definition starting in Experiment 1 with the ratio model. In particular, our empirical evaluation was in terms of predicting coincidentality judgments (how coincidental is this?) from judgments of other concepts implicated in these definitions.

The Bayesian ratio model and the various belief and uncertainty judgments in our experiments are specified in relation to the general Bayesian perspective on causal hypotheses as formalized in relation to Bayes' famous theorem, Equation 1. The emphasis here is not on Bayes' theorem as a mathematical tool (for reversing conditional probabilities), or even as a normative probability standard per se but rather as psychological theory with mappings to some of the key components in causal reasoning: People have prior beliefs about the world before events happen, and they update those beliefs after experiencing those events (Lewandowsky, Griffiths and Kalish, 2009). Formally, the belief that a hypothesis, H, is true *after* the occurrence of events/data, d, maps to the conditional probability of the hypothesis being true given that the data has occurred, p(H|d), equation

1 left side. And this equates to the prior belief in the hypothesis *before* the events/data, as the probability p(H), multiplied by the current evidence as the conditional probability of the data occurring under the (counterfactual) assumption that the hypothesis is definitely true, p(d|H), divided by the overall probability of the data, p(d). So the theorem is being used as a psychological theory for relating beliefs, expectations and evidence as mapped to probabilities.

Equation 1. Bayes' theorem. $\mathrm{p(H|d)} = \frac{\mathrm{p(d|H) \cdot p(H)}}{\mathrm{p(d)}}$

Even more importantly, Griffiths and Tenenbaum (2007) proposed a construct called "coincidence strength" in relation to evaluating potential coincidences and formalized it as the support for and evidence of one causal mechanism as compared to the support and evidence for another. (Note that this is about the degree of belief in the existence of a causal mechanism as distinct from the concept of causal strength (e.g. Cheng 1997) which is about the strength of the connection between cause and effect not the existence of that connection *per se*.) In the present context, it is most useful to consider the alternative "causal" mechanism as *chance* co-occurrence in the absence of a connecting mechanism. This Bayesian belief updating approach (see also Gill, 2014; Pearl, 1982; Pearl, 2014) starts with *prior* relative beliefs about the hypotheses and uses the current evidence, the data *likelihoods*, to generate updated, *posterior* beliefs in the hypotheses. Here degree of belief is an estimated probability of a hypothesis being true, and data likelihood is the estimated probability of data co-occurrence if a given hypothesis is definitely true. Thus coincidence strength is the relative extent to which one mechanism is supported over the other after the co-incidence has occurred.

Across multiple psychological judgment experiments using manipulated coincidences together with many computationally derived probabilities, Griffiths and Tenenbaum (2007) showed that coincidence strength is related to the relative posterior hypothesis support as a ratio. It is also related to the data likelihood ratio, the probability of the data in the context of the two hypotheses, detailed below.

Because the results from our experiments will show that coincidence strength is an important predictor of judged coincidentality, we describe Griffiths and Tenenbaum's (2007) ratio model of coincidence strength, Equation 2, in detail. Coincidence strength is proportional to the relative support, log odds form of Bayes' Theorem in relation to the evaluation of and belief in causal and chance hypotheses. The emphasis in Griffiths and Tenenbaum (2007) was on evaluating two specific hypotheses, but the emphasis here is on contrasting causal and chance hypotheses overall. In detail, Equation 2 says that the judged coincidence strength, C_s, of a co-incidence is related to the relative extent of the posterior belief in and support for the causal hypothesis, p(Ca|d), versus for the chance hypothesis, p(Ch|d), expressed as the logarithm of their ratio, $\log\left[\frac{p(Ca|d)}{p(Ch|d)}\right]$. The probabilities in this odds ratio form both derive directly from Bayes' theorem, Equation 1, and the log odds form has the important property of making the posterior log odds the sum of the log data likelihood odds plus the prior belief odds. This leads to an extremely useful Bayes' space conceptualization of coincidence strength, Figure 1 (and see Figure 2 in Griffiths & Tenenbaum, 2007).

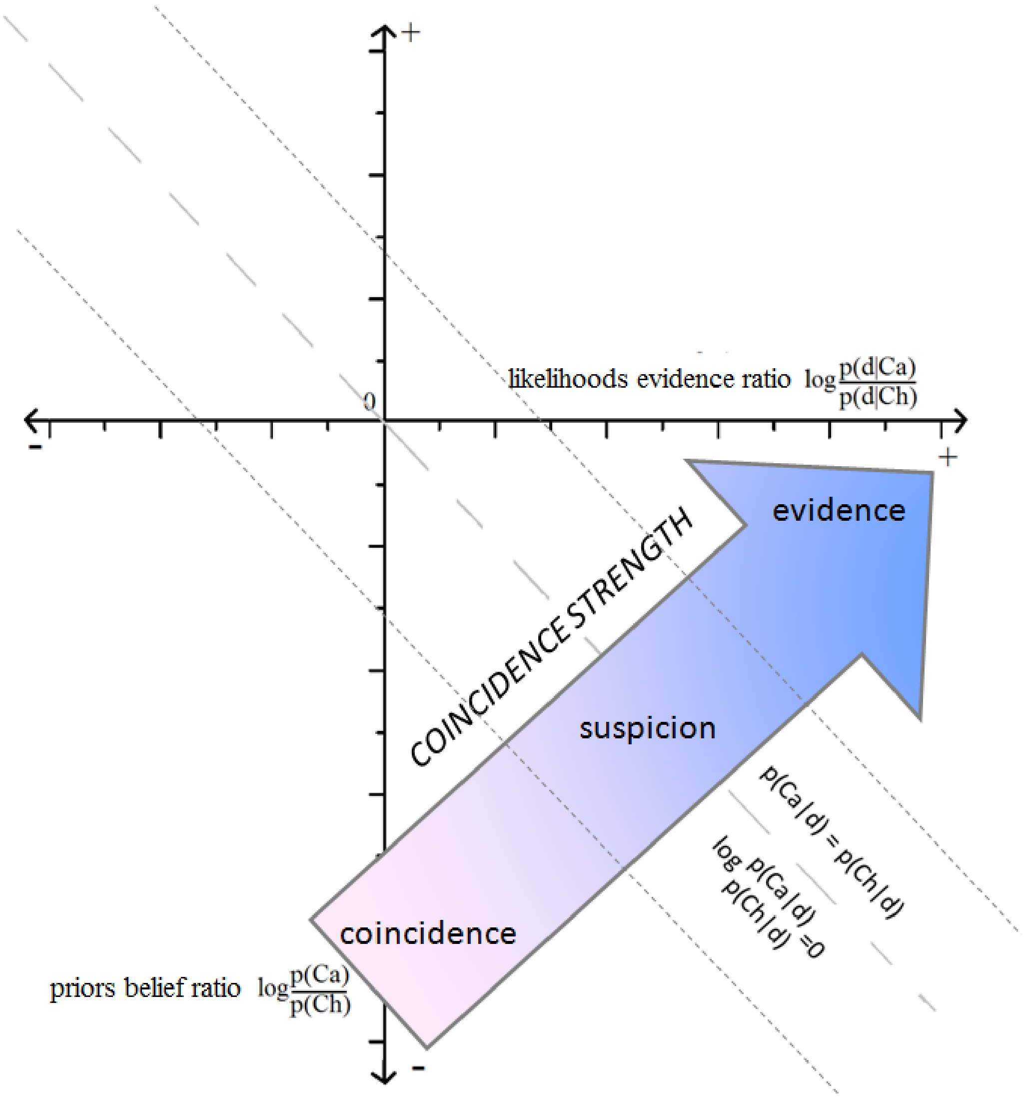
Equation 2: $C_s \sim \log\left[\frac{p(Ca|d)}{p(Ch|d)}\right] = \log\left[\frac{p(d|Ca)}{p(d|Ch)}\right] + \log\left[\frac{p(Ca)}{p(Ch)}\right]$

Coincidence strength has a clear interpretation in Bayes' space as shown in Figure 1, where the y-axis is the relative prior beliefs in the causal and chance hypotheses as the log of the ratio of their respective probabilities, and the x-axis is the relative likelihoods of the coincidence via the causal and chance hypotheses as the log ratio of these probabilities. Then each point in the space, as a combination of an x-axis co-ordinate and a y-axis co-ordinate, corresponds to a coincidence strength as per Equation 2. Events toward the bottom left hand of the space are "just coincidence". They correspond to a priors ratio (on the y-axis) favoring the chance hypothesis over the causal hypothesis; $p(Ch) > p(Ca)$, the ratio $p(Ca) / p(Ch)$ is a small number less than 1, and so the log of $p(Ca)/p(Ch)$ is negative. Further, the log likelihoods ratio on the x-axis only slightly favors the data as occurring due to the causal hypothesis over the chance; p(d|Ca) is only slightly larger than

$p(d|Ch)$, $p(d|Ca)$ / $p(d|Ch)$ is only slightly larger than 1, so log $p(d|Ca)/p(d|Ch)$ is only moderately positive. Thus the only slightly positive log likelihoods ratio isn't sufficient to overcome the substantially negative log priors ratio, resulting in relatively low coincidence strength. Coincidence strength, then, increases diagonally from the bottom left to the top right of the space; going from the region of "just coincidence" to "suspicious coincidence" (separated by a small dotted line) corresponds to the positive log likelihoods ratio doing a better job of overcoming the negative log priors ratio, and a better job still for "evidence" region (to use Griffiths and Tenenbaum's terminology). The dashed line through the origin from the top left to the bottom right of the space corresponds to a posteriors ratio of one, that is equal belief in the causal and chance hypotheses, and the increase in coincidence strength, from just coincidence to suspicious coincidence to evidence, occurs orthogonal to this negative diagonal.

As a concrete example, consider the simple thought-contact coincidence mentioned earlier, in relation to the Bayes' space in Figure 1: Prior belief in the causal hypothesis that just thinking of someone will cause (via telepathic communication) them to get in touch is low (for most people) and correspondingly prior belief in the chance hypothesis that such co-incidences are just due to chance is high (it just so happens that you thought of them at the same time they thought of you). Thus the prior relative belief ratio for the causal to chance hypothesis is a number substantially less than 1, $\frac{p(Ca)}{p(Ch)}$, because belief in the causal hypothesis is substantially less than the chance hypothesis. The natural log of this priors ratio is then a negative number corresponding to a value on the y-axis toward the bottom in the figure. Now consider the data likelihoods of the thought-contact event given the causal and chance hypothesis respectively. These are hypothetical conditional probabilities; under the counterfactual assumption that the (paranormal) causal hypothesis is true, what is the probability that the thought-contact event occurs? The implication is that $p(d|Ca)$ is a relatively large probability. (Note, though, that this probability need not be that near 1 just as a causal mechanism definitely existing doesn't necessarily mean that the cause always produces the effect.)

Figure 1. Coincidence strength, the posteriors belief ratio for the causal and chance hypotheses, in Bayes' space. The x-axis is the log of the data likelihoods ratio for the causal and chance hypotheses respectively. The y-axis is the log of the priors ratio of beliefs in the causal and chance hypotheses respectively. The blue arrow indicates increasing coincidence strength. See main text for details. Figure adapted partly from Figure 2 in Griffiths and Tenenbaum (2007).



Correspondingly, the probability of the thought-contact event if the chance hypothesis is true can't be all that high, especially for example if the friend was from childhood and not heard from in twenty years. So $p(d|Ch)$ is implied to be a relatively small probability. Thus the data likelihoods odds ratio, $\frac{p(d|Ca)}{p(d|Ch)}$, is a number substantially bigger than 1 because the data are a lot more likely via the causal hypothesis than the chance hypothesis. The natural log of likelihoods ratio is then a positive number corresponding to a value on the x-axis toward the right in the figure. The combination of the small log priors ratio toward the bottom of the y-axis, $\log\frac{p(Ca)}{p(Ch)}$, with the large log likelihoods ratio to the right on the x-axis with the likelihoods ratio, $\frac{p(d|Ca)}{p(d|Ch)}$, puts coincidences in general predominantly in the bottom right quadrant of the space and with an intermediate posteriors ratio near 1 corresponding to a log posteriors ratio near 0, $\log\frac{p(Ca|d)}{p(Ch|d)} = \log\frac{p(d|Ca)}{p(d|Ch)} + \log\frac{p(Ca)}{p(Ch)}$. This fits neatly with Griffiths and Tenenbaum's (2007) definition of coincidences as events which provide some support for an alternative, e.g. paranormal, hypothesis over some other hypothesis such as chance but not enough to overcome prior beliefs against the alternative hypothesis. And the larger the log posterior ratio, the greater the evidence for the causal mechanism and thus the greater coincidence strength. So again coincidence strength as relative evidence for the causal hypothesis over the chance increases diagonally from the bottom left to the top right of the space, Figure 1. As subsequent results from our experiments will show, this space provides a partial, but incomplete account of coincidentality, in response to the question "how coincidental is this?" We will argue causality and coincidentality are not equivalent constructs but rather correspond to different things.

Before giving an overview of our research, a real example of an actual coincidence, the solar eclipse coincidence, is useful to contextualize the key purposes of this research and preview a key conclusion: The remarkably close co-incidence of the apparent size of the moon and sun during certain kinds of solar eclipse in the only place in the universe where we know there is life has been widely noted; had either been slightly bigger/smaller, slightly closer/farther, etc., eclipse alignment wouldn't be nearly as good or as surprising as it is. A common believer perspective, especially for advocates of various religions, is to argue that the co-incidence is too unlikely by chance (d|Ch is judged small) to actually be chance (Ch|d is believed small), and the lack of a standard causal mechanism linking good eclipses and life (Ca|d and d|Ca are small) means that a deity must have caused both. In reaction, a common skeptic perspective is to admit this co-incidence is fairly unlikely by chance (d|Ch), but then argue as follows: the lack of a clear causal mechanism (Ca|d and d|Ca are small) implies this is just a chance event, i.e. a coincidence (Ch|d is relatively large because of the large prior belief in chance, p(Ch)). They also tend to point to the many attributes the earth has as providing many possible places surprising coincidences might have occurred thus increasing the probability of such co-incidences in at least some of these just by chance, i.e. p(Ch|d) is large even though d|Ch for the specific co-incidence is small. In contrast to both of these approaches to interpreting this co-incidence, the emphasis here is on the curious enquirer who, like the others, notes the surprising improbability of the co-incidence by chance (small d|Ch) as well as by well-established causal mechanisms (small d|Ca). However unlike the skeptics and believers, the curious enquirer is extremely suspicious that there might be an as yet unclear causal mechanism operating, but they are also actively looking for and wanting *a lot* more evidence, as there is currently only one place known to have life, the earth. And it is worth emphasizing that, in contrast to research comparing human/psychological judgment against a normative standard, the true/normative values for the key judged probabilities do not tend to be objectively known for real co-incidence situations involving causal suspicion. The objective truth isn't yet clear. As such, while co-incidences occur objectively in the world, coincidences are perceptions associated with beliefs and judgments and so are psychological phenomena in peoples' minds.

1.2 Research Purpose and Overview

The data-driven purpose of this research was to evaluate the basic psychological judgment properties of coincidences using a reasonably large sample of real, naturally occurring coincidences from a diary study. The rationale for this, in contrast to most prior judgment studies on coincidences, was to avoid using artificially fabricated or manipulated coincidences so as to evaluate the judgment processes involved with the conceptual richness of real events, as self-judged coincidences by the participants that had experienced them (Diary study). And the central judgment we asked all of our participants to provide, in contrast to the large prior literature on causality, was coincidentality as the answer to the basic question, "how coincidental is this?"

The theory-driven purpose of our empirical work was to support our psychological definition by developing a model of coincidentality based on psychological judgments for concepts implicated by the definition. But it is worth highlighting that we did not give participants this or any other definition of coincidence. Four experiments progressively evaluated the conceptual components in our definition of coincidence and their relations to judged coincidentality: Experiment 1 measured degree belief in the truth of causal and chance hypotheses as related to the part of the definition "coincidences are…. ascribed to chance [p(Ch|d) is high]… search… has not produced anything more plausible than mere chance [p(Ca|d) is low]". Experiment 2 assessed event likelihood (probability) in the context of causal and chance hypotheses as related to the part of the definition, "coincidences are … unlikely by chance [d|Ch is low]". Experiment 3 elicited judgments of event surprisingness, "coincidences are surprising…", and evaluated several ways of formalizing surprise based on event likelihood judgments. Finally, Experiment 4 evaluates all key judgments in a within participants design so as to assess the relations between all the key components of the definition.[2]. Ultimately, the purpose was to support the psychological construct of coincidentality as a cognitive marker that new causal mechanisms *might* be operating and further evaluation *may* lead to new, adaptive causal knowledge.

2. Materials Diary Study

This study collected a sample of naturally occurring self-reported coincidences. These were used as materials for subsequent judgment experiments.

[2] These experiments also had other judgment and/or trait tasks, usually small and at the end, the results of which we do not report in the interests of space and focus These included a variety of clinical and personality measures and small judgment tasks such as judgments about contrived coincidences with known probabilities. Methodologically, the core judgment tasks were unlikely to have been substantively influenced by these other tasks as they were mostly completed after the core judgment tasks.

2.1 Methods

2.2.1 Participants

Twenty-five students from University of Surrey and University College London completed this experiment in exchange for partial course credit or an option to enter a £50 lottery draw.

2.2.2 Procedure

Each participant was asked to record any coincidences they experienced in a booklet "in as much detail as possible." The diary covered a five-week period.

2.2 Results and Discussion

Participants recorded 102 coincidences, examples in Table 1. Five of these were eliminated (as potentially offensive), and the set of 97 coincidences used in subsequent experiments was edited only slightly to fix typographical errors/spelling mistakes. A key aspect of these results is that, unlike most prior research, the participants were actively watching for coincidences rather than retrospectively reporting them (Bressan, 2002).

3. Experiment 1. Coincidentality and the Likeliness of Causal/Chance Hypotheses Given the Data

This experiment evaluated three key judgments—coincidentality, causal likeliness and chance likeliness--for the diary set of coincidences. These judgments were answers to the questions, How coincidental is this?, How likely due to chance? and How likely due to causal mechanism?, all

Table 1. Example coincidences from the diary study

Blond-guy coincidence (typically judged most coincidental in the set and marked by **red circles** in subsequent figures.**)**	"In January I was in NZ. I was in a Belgian restaurant with my parents when we noticed an English family, particularly a blond guy that was my age drinking next to a bus. When I moved to Guildford I met the same guy again because we were going to the same gym."
Weekend-alarm coincidence (typically judged least coincidental, **blue squares**.)	"At the week-end I woke up at almost exactly the same time my alarm goes off on a weekday."
Thought-contact coincidence	"Having been discussing the unlikelihood of bumping into a certain person that I have not seen for ages around campus, I bumped into them on several occasions later that day."

on 1 to 100 judgment scales. Having both causal and chance likeliness judgments as estimates of belief in the posterior hypotheses for cause and chance respectively allowed direct assessment of the Bayesian ratio model, Equation 2.

3.1 Methods

3.1.1 Participants

Thirty-three undergraduate psychology students at Cardiff University completed this experiment in exchange for course credit as part of a participant panel.

3.1.2 Materials, Design and Procedure

In overview, judgment tasks were presented as a series of worksheets in Microsoft Excel with instructions to work through the sheets in order. The instructions for each task were in the top left cell of each worksheet, and the order of the coincidences on each sheet was randomized.

Participants completed three separate judgment tasks on the 97 diary study coincidences: 1) coincidentality, 2) causal likeliness and 3) chance likeliness--in counterbalanced order. The coincidentality rating task was to answer the question, “how coincidental is this?” for each of the 97 coincidences on a scale from 1, not at all coincidental, to 100, extremely coincidental. The causal likeliness instructions were, “On a scale from 1 to 100 please specify how likely the events in each description are to have been causally connected where 1 is not at all likely to have been causally connected and 100 is extremely likely to have been causally connected.” The chance likeliness task instructions were, “On a scale from 1 to 100, please specify how likely the events in each description are to have occurred by chance where 1 is not at all likely to have occurred by chance and 100 is extremely likely to have occurred by chance.”

3.1.3 Data tabulation and modeling analysis

The handling of missing, ambiguous and repetitive responses as well as judgment scale reversals by a minority of participants for this and subsequent experiments is described in detail in

Appendix A. But to summarize: Averaged data in figures and modeling results was based on all data except that a small number of individual responses that were substantially outside of the judgment scales (1 to 100), e.g. 606, were deleted. However, combined plots of individual participant responses, e.g. Figure 2 bottom panels, only included participants with no scale reversals, as did individual participant model fits. It is worth noting that prior research, including Griffiths and Tenenbaum (2007), has also eliminated scale-reversed participants. However, such scale reversals were relatively rare here, e.g. only four participants out of 33 apparently reversed the coincidentality scale in this experiment. Also we report correlation, r, for judgment relations and R^2 for model fits.

3.2 Results and Discussion

Coincidentality was negatively related to causal likeliness for both data averaged across participants by coincidence, r = -0.75, and all (nonreversed) individual participant judgements combined together, r = - 0.37. Coincidentality was positively related to chance likeliness for both averaged data, r = + 0.90, and combined (nonreversed) data, r = +0.58. Further, since causal and chance hypotheses are mutually exclusive, increased support for one conceptually corresponds to decreased support for the other, consistent with the observed negative relationship between causal and chance likeliness rates for averaged data, r = -0.77, and combined (nonreversed) data, r = -0.44. Thus, the causal and chance likeliness judgments are consistent with these being measures of support for and degree of belief in the Bayesian posterior hypotheses for cause and chance.

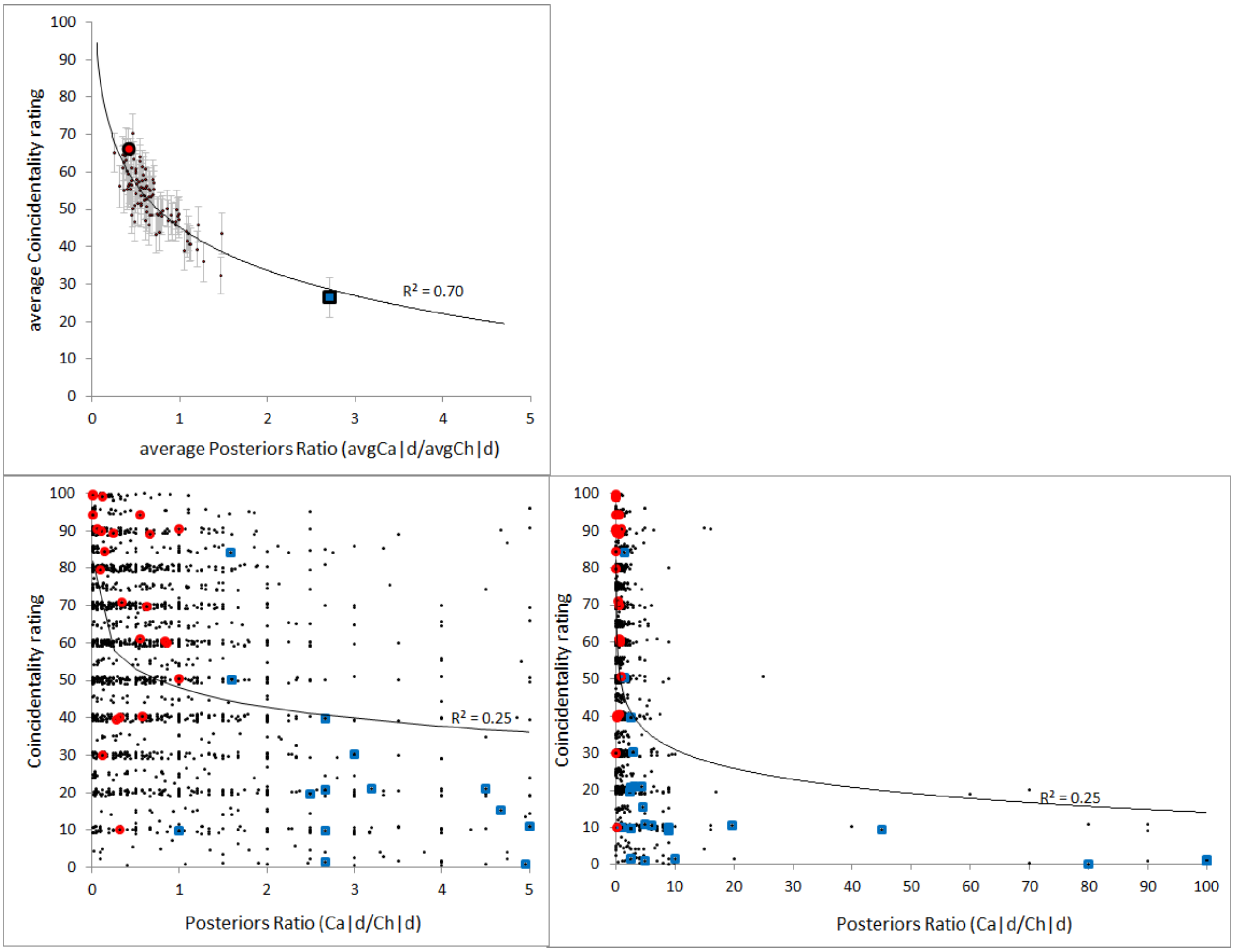
Combining the causal and chance likeliness judgments together in a posteriors hypothesis belief ratio, Equation 2, coincidentality was *inversely* related to this ratio, Figure 2; it accounted for 70% of the variance in coincidentality for averaged data, R^2 = 0.70 top panel, for 25% of the variance of the combined (nonreversed) results, R^2 = 0.25 in the bottom panels, and an average of 26% of the variance for individual (nonreversed) participants (R^2 = 0.26, model details in Appendix B). Posterior ratios greater than 1 correspond to a balance of support favoring causal mechanisms over chance, so for example the ‘weekend-alarm’ coincidence (Table 1) had an averaged posteriors ratio of slightly less than 3 (the blue square Figure 2 top panel), supporting the causal over the chance hypothesis. Posteriors ratios less than 1 correspond to a balance of evidence favoring chance over causal mechanisms, so for example the 'blond-guy' coincidence had a posterior ratio of less than 0.5 thus supporting the chance hypothesis (the red circle Figure 2 top panel). Since these events were reported coincidences, the majority had posterior ratios less than 1.

The results in Figure 2 show that judged coincidentality is *inversely* related to the posteriors ratio and thus inversely related to the construct of coincidence strength, equation 1. Further the trend line in the figure suggests this relationship is logarithmic. So the first form of the coincidentality model, Equation 3, is that coincidentality is related to the log of the inverted posteriors belief ratio. And to account for individual differences in scale use, the model has two scaling parameters; Co_scale, to accommodate participants who restricting their responding to a subsection of a judgment scale, and a centrality parameter, Co_mid, to account for participants whose sets of coincidentality judgments were not centered on the middle of judgment scale.

Equation 3. $\mathbf{C}_{\boldsymbol{\emptyset}} \sim \mathrm{C}_{\mathrm{o\_scale}} \cdot \log_{\mathrm{e}}\left[\frac{1}{\mathrm{C_s}}\right] + \mathrm{C}_{\mathrm{o_{mid}}}, \mathrm{C_s} \approx \frac{\mathrm{p}(\mathbf{Ca}|\mathbf{d})}{\mathrm{p}(\mathbf{Ch}|\mathbf{d})} = \frac{1-\mathrm{p}(\mathbf{Ch}|\mathbf{d})}{\mathrm{p}(\mathbf{Ch}|\mathbf{d})}$

Taken together, these results indicate the stability of coincidentality (how coincidental is this?) as a psychological construction. In particular, this stability is indicated by coherence with other judgments, e.g. the predictiveness of coincidence strength. These results support the component of our definition of coincidences as events which are "ascribed to chance [p(Ch|d) is high] since the search for causal mechanisms has not produced anything [p(Ca|d) is low] more plausible than chance [the inverted posterior ratio p(Ch|d)/p(Ca|d) is large]." However, what is apparently missing is support for that aspect of our definition of coincidence as "unlikely by chance...." This is implied in the combined (nonreversed) judgment results for coincidentality versus coincidence strength in the *bottom* panels of Figure 2: when belief in the causal hypothesis posterior is large relative to the chance posterior, coincidentality is constrained to be low, consistent with things that are definitely causally connected not being coincidences; but when the causal posterior is low, coincidentality is not correspondingly constrained and can be high or low. Both the proposed definition of coincidence and the data in the bottom panel of Figure 2 suggest that another factor contributes to coincidentality when the causal posterior is not large (relative to the chance posterior). And the representation of coincidence strength, Equation 2, in Bayes' space, Figure 1, makes plausible suggestions for what

Figure 2. Average coincidentality versus the ratio of the average causal likeliness (Ca|d) to the average chance likeliness (Ch|d), top panel, and combined plots of all nonreversed participants' judgments, bottom panels, from Experiment 1. Error bars are standard error. Note that the left bottom panel is a magnification of the right bottom panel for posterior ratios between 0 and 5. The trend line is logarithmic. Red circles are the blond-guy coincidence and blue squares are the weekend-alarm coincidence from Table 1. A small amount of noise jitter has been added to the data points in this and subsequent combined data panels to more clearly convey data density.



this might be in terms of the data likelihoods; the likelihood of the co-incidence under the (counterfactual) assumption that the causal hypothesis is definitely true, p(d|Ca), and the likelihood of the coincidence given that the chance hypothesis is definitely true, p(d|Ch). As described in the Introduction, it is conceptually important to distinguish data likelihoods given the hypotheses, p(d|Ca) and p(d|Ch), from the posterior probabilities of the hypotheses given the data, p(Ca|d) and p(Ch|d). This was examined in Experiment 2.

4. Experiment 2. Coincidentality and Data Likelihood Given the Causal and Chance Hypotheses

The primary purpose of this experiment was to add judgments of data likeli*hood* for the causal and chance hypotheses. Experiment 1's likeli*ness* judgments were about posteriors hypotheses beliefs, e.g. a preference for the chance hypothesis over the causal hypothesis such as for a thought-contact coincidence that does *not* produce posterior belief in a paranormal causal hypothesis about telepathic connection. But even if posterior belief/likeliness is low, data likelihoods for paranormal causal hypotheses can be high for coincidences, e.g. if telepathic connections are assumed to definitely exist, the probability of such coincidences might be genuinely high. To emphasize, both Bayes' theorem (Equation 1) and coincidence strength (Equation 2) suggest that likelinesses and likelihoods are different concepts. A key motivation for these judgments is to address the distinction in our definition of coincidence "unlikely by chance, but nonetheless ascribed to chance...." And a related benefit is that these measures then allow coincidentality to be evaluated in relation to the conceptually powerful Bayes' space (Griffiths & Tenenbaum, 2007) in Figure 1, and to Griffiths and Tenenbaum's definition of coincidence as events where there is a contrast between prior beliefs and present evidence.

4.1 Methods

4.1.1 Participants

The forty participants were from a psychology participant panel at Cardiff University, 17 as part of a paid panel and 23 in exchange for partial course credit.

4.1.2 Materials, Design and Procedure

As with Experiment 1, Experiment 2 measured coincidentality, causal likeliness as a measure of posterior belief in the causal hypothesis and chance likeliness as posterior belief in the chance hypothesis separately on all 97 diary study coincidences. In addition, causal data likelihood and chance likelihood were also separately assessed on the full set. Thus Experiment 2 contained judgments of both posterior hypothesis probabilities and both coincidence likelihoods given those hypotheses. Causal likelihood, as the probability of the data given the causal hypothesis (d|Ca), was measured using the following (counterfactual) instructions: “Assume you know for sure that the events in each description are the result of a causal mechanism connecting them. However keep in mind that some causally connected events are less or more likely to occur than others. On a scale from 1 to 100, please specify how likely the events in each description are to have occurred where 1 is not at all likely and 100 is extremely likely to have occurred.” Similarly, chance likelihood as the probability of the data given the chance hypothesis (d|Ch) was assessed using similar instructions starting, “Assume you know for sure that the events in each description occurred solely as a result of chance. However keep in mind that some chance events are less or more likely to occur than others. On a scale from 1 to 100. . . .” As in previous experiments, the coincidences were in different random orders for each task and participant.

All participants completed the coincidentality rating task. However, to keep the experiment to a reasonable length, only half of the participants completed each probability rating task; that is, half of the participations did the causal posterior task (Ca|d) and the chance likelihood task (d|Ch) while the other half of the participants did the chance posterior task (Ch|d) and the causal likelihood task (d|Ca). Thus, each participant did three separate ratings tasks on the full diary set. The first two tasks were coincidentality and a posterior ratings task (either Ca|d or Ch|d) with order counterbalanced across participants. The third rating task was always a likelihood rating task, either causal (d|Ca) or chance likelihood (d|Ch).

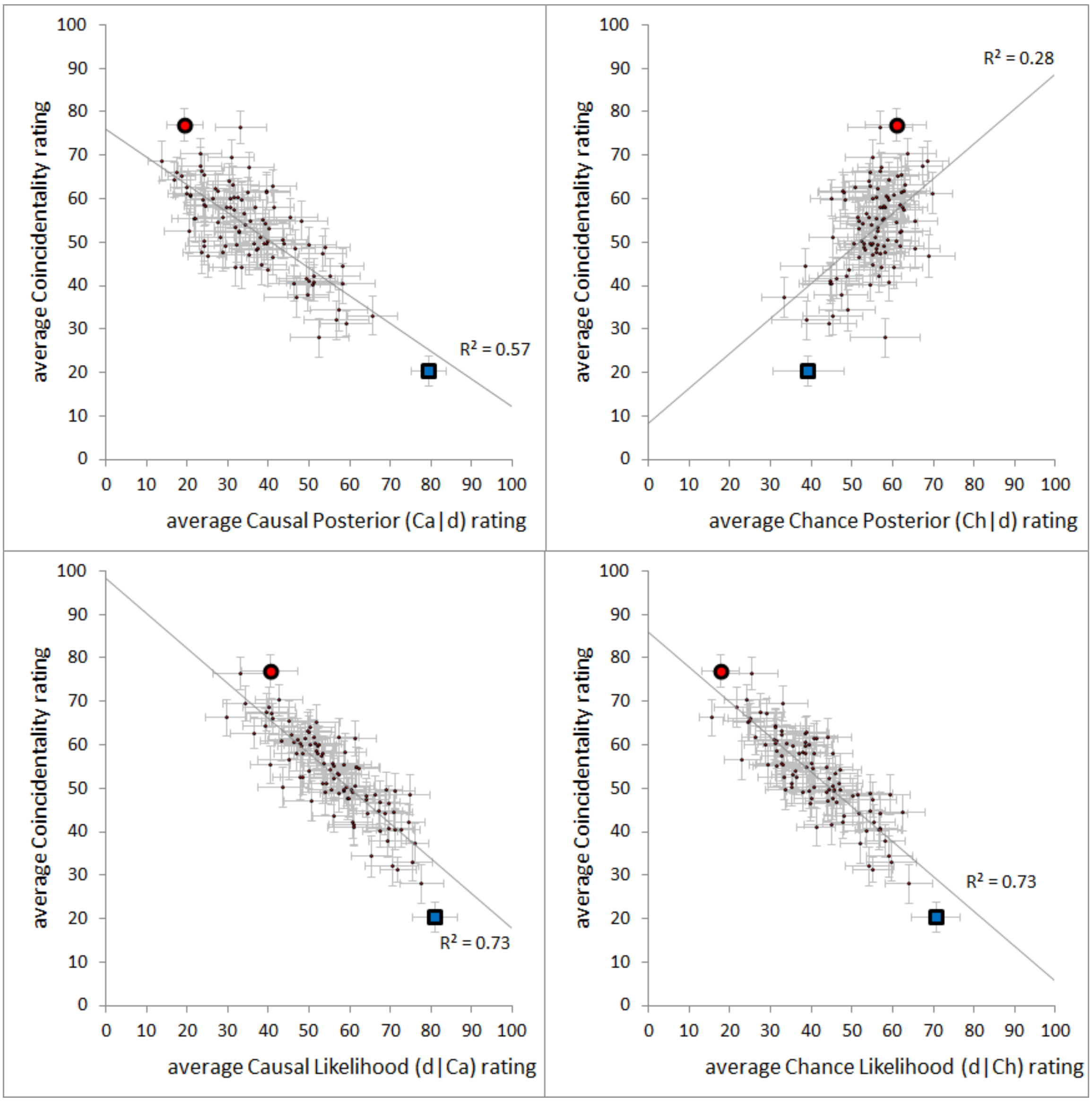
4.2 Results and Discussion

Experiment 2 replicated Experiment 1 for coincidentality and the posteriors judgments again indicating the stability of these constructs; there was a negative relationship between coincidentality and the causal posterior for averaged data, r = -0.76 shown in Figure 3's top left panel, and a positive relationship between coincidentality and the chance posterior, r = 0.53, top right panel. The key additional results were that coincidentality was negatively related to average causal likelihood, r = -0.85 shown in Figure 3's bottom left panel, and, critically, that coincidentality was *negatively* related to chance likelihood, r = -0.85, bottom right panel.

Taken together, the difference in the relationships between coincidentality and, respectively, the chance posteriors (Figure 3 top right panel) and the chance likelihoods (bottom right panel) supports our definition of coincidence as "unlikely by chance but… ascribed to chance…", i.e. coincidences are "unlikely by chance" in terms of a relatively small chance likelihood, p(d|Ch), "but nonetheless ascribed to chance" in terms of a relatively large chance posterior, p(Ch|d), because "the search for causal mechanisms has not produced anything", the causal posterior being small, "more plausible than mere chance", relative to the chance posterior.

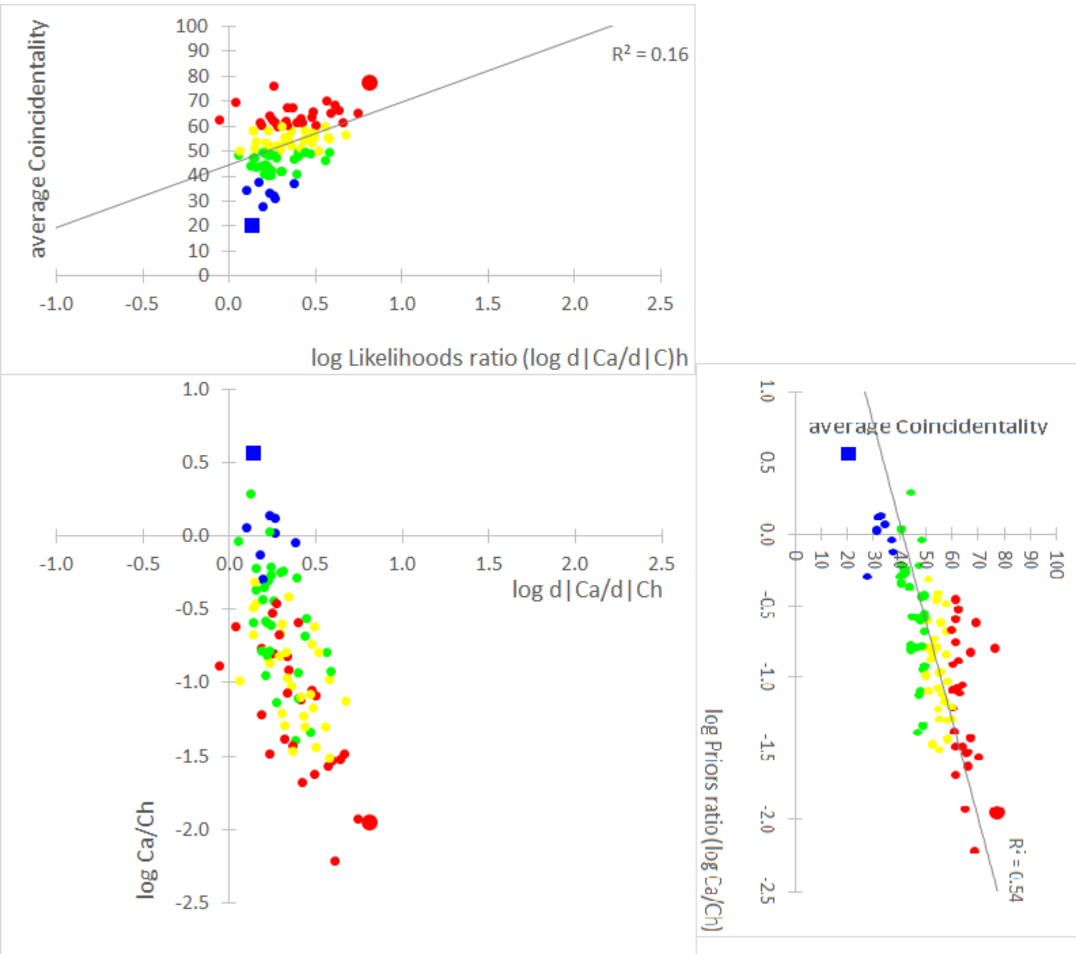
Having judgments of both posterior hypothesis probabilities and both coincidence likelihoods means that all three of the key Bayesian ratios can be specified (equation 2): the posteriors ratio of the causal to chance hypotheses, p(Ca|d)/p(Ch|d); the likelihoods ratio for the coincidence given the cause and chance hypothesis, p(d|Ca)/p(d|Ch) and the priors ratio for the causal and chance hypotheses, p(Ca)/p(Ch), which can be inferred via Bayes' theorem (as the posteriors ratio divided by the likelihoods ratio). Having all three ratios allows coincidentality to be represented in the conceptual Bayes' space (Griffiths and Tenenbaum, 2007) described in the Introduction (Figure 1) and shown for the present data in the central panel of Figure 4. In Figure 4 coincidentality is color coded in relation to the two key extreme coincidences (Table 1), the 'blond-guy' coincidence (the large red circle), and the 'weekend-alarm' coincidence (the large blue square). The top panel explicitly shows this color coding with coincidentality, y-axis, versus the log of the coincidence likelihoods ratio, x-axis, and the far right panel shows coincidentality, x-axis, versus the log of the priors ratio, y-axis. The bottom left panel then shows color-coded coincidentality in Bayes' space with the log priors ratio on the x-axis, aligned with the x-axis in the top panel, and the log likelihoods ratio on the y-axis, aligned with the y-axis in the far right panel. This representation supports

Figure 3. Coincidentality versus each of the four probability judgments--the causal and chance posteriors and the causal and chance likelihoods--for averaged judgments from Experiment 2. Error bars are standard error. Trendlines are linear with percentage of variance accounted for. Red circles are the blond-guy coincidence and blue squares are the weekend-alarm coincidence from Table 1.



Griffiths and Tenenbaum's (2007) claim and definition that coincidences tend to be events where the likelihood ratio is positive, that is, data points to the right of zero on the x-axis favoring the data as the result of a causal hypothesis, but the priors ratio is negative, data points below zero on the y-axis favoring the chance hypothesis; thus most of the coincidences are in the bottom right hand quadrant,

Figure 4. Average coincidentality color coded in Bayes' space (Figure 1) composed of the log of the priors ratios of the averages, y-axis, and the log of the likelihoods ratios of the averages, the x-axis of the central panel, for data from Experiment 2. The top panel shows the color coding of coincidentality, y-axis, in relation to the log likelihoods ratio, the x-axis, red is most coincidental and blue is least. And the right panel shows a similar coding of coincidentality for the priors ratio. Large Red circles are the blond-guy coincidence, and large blue squares are the weekend-alarm coincidence from Table 1.



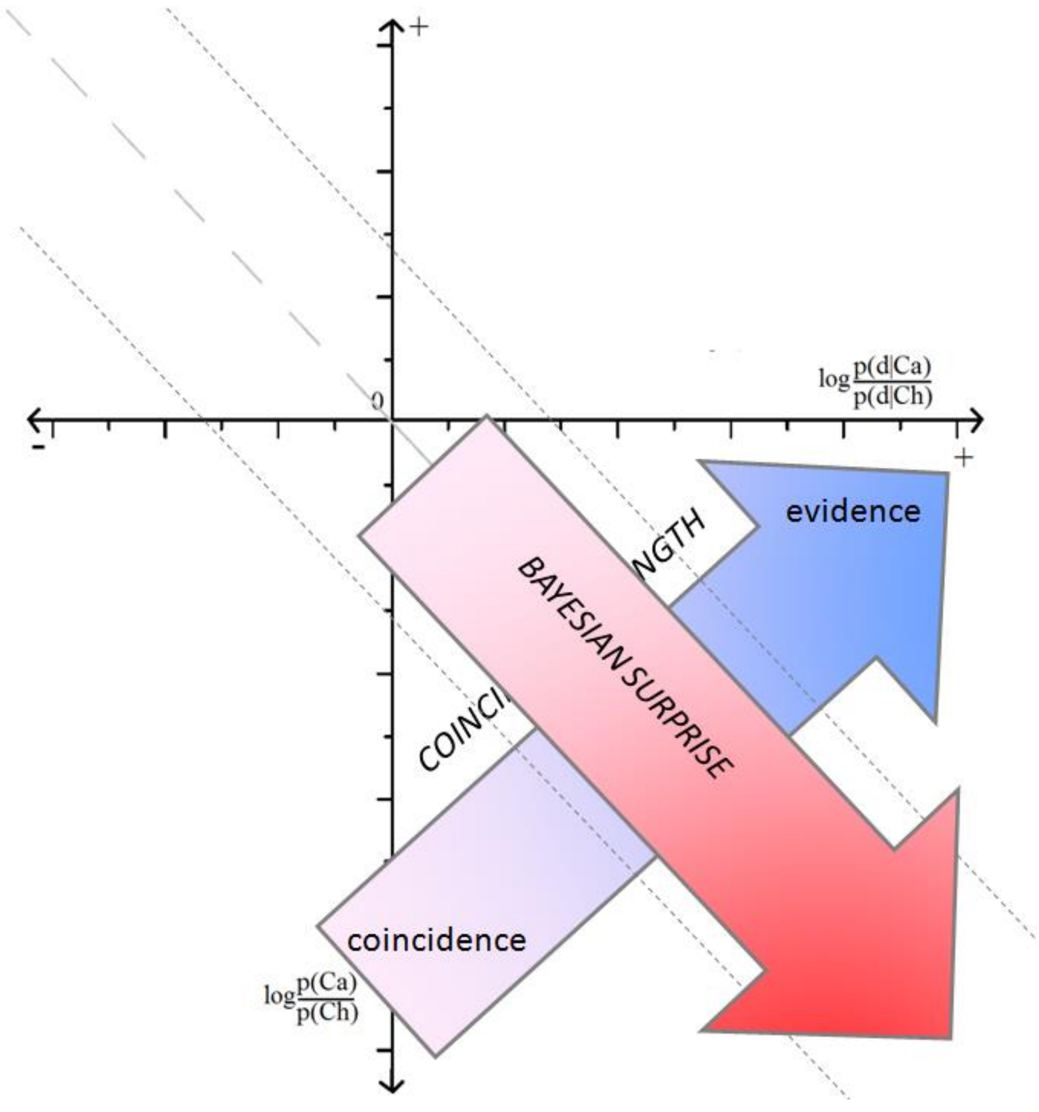
with only a handful of probably causal events in the top right hand quadrant, notably the weekend-alarm coincidence (Table 1) as the most causal of the events.

An important aspect of representing coincidentality in Bayes' space (Figure 4) is that coincidentality seems to increase roughly along a negative diagonal from the top left, the weekend-alarm coincidence, to the bottom right, the blond-guy coincidence. A particular benefit of Bayes' space is that it corresponds to a specific formalization for the concept that is explicitly missing from the posteriors ratio as related to coincidentality, namely surprise. Surprise features prominently in our definition of coincidence in terms of "surprising pattern repetitions," and intuitively/anecdotally surprise is a very common property of coincidences. This Bayesian surprise, shown in an updated version of Bayes' space in Figure 5, was described by Griffiths and Tenenbaum (2007, and relatedly by Baldi & Itti, 2010) as the extent of the conflict/disagreement between expectations (prior beliefs) and current events (data likelihoods). In detail, this conflict is between the likelihoods ratio and priors ratio in terms of the priors ratio favoring the chance hypothesis but the likelihoods ratio favoring the data as from the causal hypothesis. Thus, Bayesian surprise increases toward the bottom right hand of the space since points progressively farther toward the bottom right correspond to progressively more positive log likelihoods ratios on the x-axis (favoring the data as from the causal hypothesis) but progressively more negative values of the priors ratio on the y-axis (favoring the chance hypothesis). So the larger the likelihoods ratio and the smaller the priors ratio the greater Bayesian surprise. This suggests Bayesian surprise is based on a ratio of these two ratios, Equation 4 right hand side, that is, the likelihoods ratio divided by the priors ratio; the larger the likelihood ratio in the numerator and the smaller the priors ratio in the denominator, the bigger the ratio of ratios and the higher Bayesian surprise. Conceptually, the reason the posteriors ratio, Equation 2, isn't a measure of Bayesian surprise is that a posteriors ratio of one can arise because the likelihoods and priors ratios are both 1 *or* because the priors ratio is very small and the likelihoods ratio is very large.

Thus, coincidence strength by itself doesn't indicate surprise, but the extent of the conflict between prior beliefs and present events does, Equation 4.

To facilitate evaluation of Bayesian surprise, the bottom left panel of Figure 6 presents an updated Bayes' space where the x-axis is Bayesian surprise, the log ratio of the likelihoods ratio to the priors ratio, i.e. a ratio of ratios, and the y-axis is then the log of the posteriors ratio. This is not a new space but is rather a rotation of the original space, in Figure 4, by 45 degrees counter-clockwise. But now the additional smaller panels associated with each axis in the main panel on the bottom left

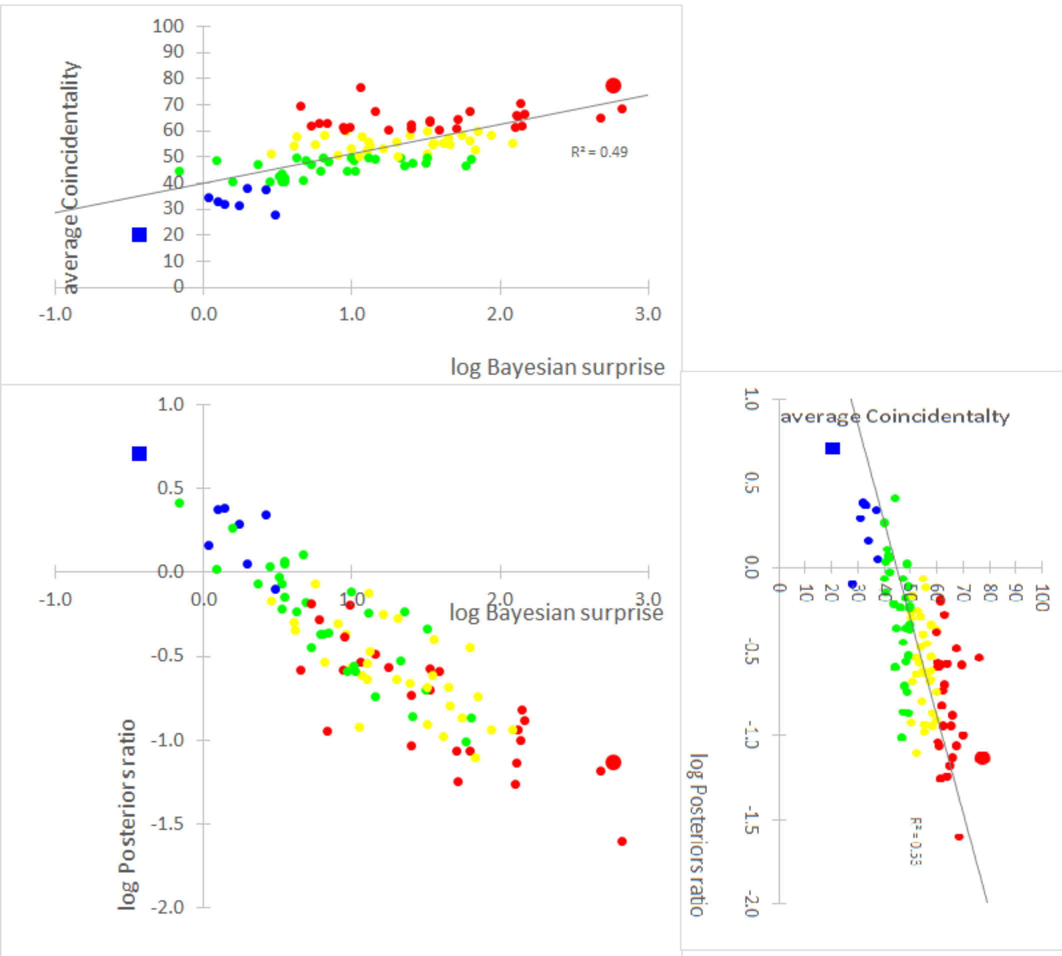
Figure 5. Bayesian surprise as the extent of the disagreement between the prior belief and data likelihoods ratios the y and x axes respectively in the Bayes' space from Figure 1 in the Introduction.



show color-coded coincidentality in relation to Bayesian surprise (top panel) and to the posterior ratio (right panel). Bayesian surprise accounted for R^2 = 0.49 of the variance in coincidentality for averaged data (Figure 6 top panel), coincidence strength for R^2 = 0.53 (right panel) and the combination of both for R^2 = 0.55 (Appendix B). This updated model is in Equation 4.

Equation 4. $\mathbf{C_{\emptyset}} \sim C_{o\_scale} \cdot \log_e\left[S_u \cdot \frac{1}{C_s}\right] + C_{o\_mid},\ C_s \approx \frac{p(\mathbf{Ca}|\mathbf{d})}{p(\mathbf{Ch}|\mathbf{d})}, \mathbf{S_u} \sim B_s \approx \frac{\left(\frac{p(\mathbf{d}|\mathbf{C_a})}{p(\mathbf{d}|\mathbf{C_h})}\right)}{\left(\frac{p(\mathbf{C_a})}{p(\mathbf{C_h})}\right)}$

Figure 6. Average coincidentality from Experiment 2 color coded in rotated Bayes' space, with the log of Bayesian surprise on the x-axis and log coincidence strength on the y-axis of the central panel. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to log Bayesian surprise on the x-axis in the top panel and in relation to the log posteriors ratio on the y-axis in the right panel. Large Red circles are the blond-guy coincidence, and large blue squares are the weekend-alarm coincidence from Table 1.



In summary, Experiment 2 replicated and extended the prior experiment in terms of predicting coincidentality. The results support the contrast between the chance likelihood and the chance posterior in the definition of coincidence in terms of "unlikely by chance but nonetheless ascribed to chance". Furthermore, these results replicate the (negative) relationship between coincidentality and coincidence strength and show that Bayesian surprise is also a good predictor of coincidentality. The conceptual and empirical importance of surprise for coincidences strongly suggested measuring judgments of surprise, Experiment 3.

5. Experiment 3. Coincidentality and Surprise

While not a synonym for coincidence (e.g. a sudden loud bang need not be a coincidence), surprise is a very typical attribute of coincidences anecdotally, in prior definitions including ours (e.g. Henry, 1993; Johansen & Osman, 2015) and in prior coincidence research (e.g. Falk & MacGregor, 1983). Surprise is an emotion associated with unexpected events and responding to such events with heightened awareness (expressed as widened eyes, etc.) and vigilance (Izard, 1991). And surprise, in terms of an unexpected event, is implicated as a basic part of associative learning in terms of prediction error e.g. as formalized in the Rescorla-Wagner model (Rescorla & Wagner, 1972). Relatedly, surprise can act as an attractor/driver for attention (e.g. Baldi & Itti, 2005). Foster and Keane (2015; 2019) propose a metacognitive explanation-based theory of surprise where surprise is inversely related to the availability of plausible explanations, similar to the argument we make for coincidences in terms of the initial lack of mechanisms that are plausible. From their perspective, surprise as a continuous construct is a metacognitive marker anticipating the amount of effort it will take to find the right explanation and eliminate the surprise. Our view in the context of coincidences is more constrained in that surprise is the lack of plausible explanations as related to the improbability of the mechanisms in hand; sometimes coincidences that initially are very surprising on reflection resolve with a sudden, low effort realization of a plausible explanation. Foster and Keane (2015) also argue against a probability-based theory of surprise. Computational complexity theory provides a related perspective (Schmidhuber, 2009) on psychological surprise as a metacognitive anticipation of the potential for data compressibility; in the context of the goal to accurately keep a record of events in reality, truly random events tend to require very long descriptions corresponding to data that are relatively uncompressible. In contrast, events with a good theoretical account are well summarized by that account, and hence the data are highly compressible in terms of only needing to store the account rather than all the different data points. From this perspective, surprise anticipates the potential for such a good account before the mechanisms underlying that compressibility have been fully worked out. So surprise can be seen as occurring for an intermediate state between completely random (uncompressible) and completely predictable (very compressible) data (Schmidhuber, 2009). While coincidences tend to be fairly unique events with relatively little immediate compressibility because the data are so limited, they do imply the potential for such compressibility in the future in terms of an unknown mechanism. And this is related to the concept of a “randomness deficiency” in data (Li & Vitányi, 2019) in terms of apparent patterns in the data which don’t obviously correspond to plausible known mechanisms. And the idea of randomness deficiency eliciting surprise (Maguire, et al. 2019) as a contrast between an apparent pattern and chance resonates strongly with our definition of coincidence (Johansen & Osman, 2015). Finally, the pervasive idea of surprise as a contrast between expectations and events arises naturally in the Bayesian framework. We’ve formalized this to align with Bayes’ space in Figure 5 (Griffiths & Tenenbaum, 2007) but related accounts have been in terms of the difference between the distributions of the posteriors and priors (Baldi & Itti, 2010). Overall these different perspectives on surprise all implicate its importance for adaptive behavior in the face of uncertain events. As such, co-occurrence surprise is an important property of coincidences.

The purpose of this experiment was to evaluate judged surprise and its relationships with coincidentality and the various uncertainty judgements, which were used to specify several derived measures of surprise including Bayesian surprise. In addition, coincidences sometimes acquire importance in terms of fortune, luck or fate in the context of various paranormal beliefs because of the personal benefit or harm attached to them (Beitman, 2009; Henry, 1993). More generally outcome valence in terms of benefit and harm has a pervasive role in causal reasoning (e.g. Shultz, 1982; Shultz, Fisher, Pratt & Rulf, 1986; Fugelsang & Thompson, 2000). Experiment 3 also evaluated judged valence, that is how good or bad each coincidence was for the person who experienced it.

5.1 Methods

5.1.1 Participants

Fifty participants completed this experiment in exchange for partial course credit as part of a psychology participant panel at Cardiff University.

5.1.2 Materials, Design and Procedure

As with Experiment 2, Experiment 3 included judgments for coincidentality, the causal and chance hypothesis posteriors, and the causal and chance likelihoods. In addition, Experiment 3 included a rating of surprise, “Please specify how surprising the events in each description are where 1 is not at all surprising and 100 is extremely surprising.” Also, there was a valence rating task, “On a scale from -100 to +100, please specify how bad or good the person who experienced the events in each description found them where -100 is extremely bad, 0 is neutral and +100 is extremely good.”

All participants in Experiment 3 completed four separate judgment tasks on the diary set of 97 coincidences. In particular, all participants did the coincidentality rating task and a separate surprise and valence task which asked for both surprise and valence ratings for each coincidence. Half of the participants did the two posteriors judgments as two separate tasks while the other half of the participants did the two likelihoods judgments as separate tasks. Task order was counterbalanced.

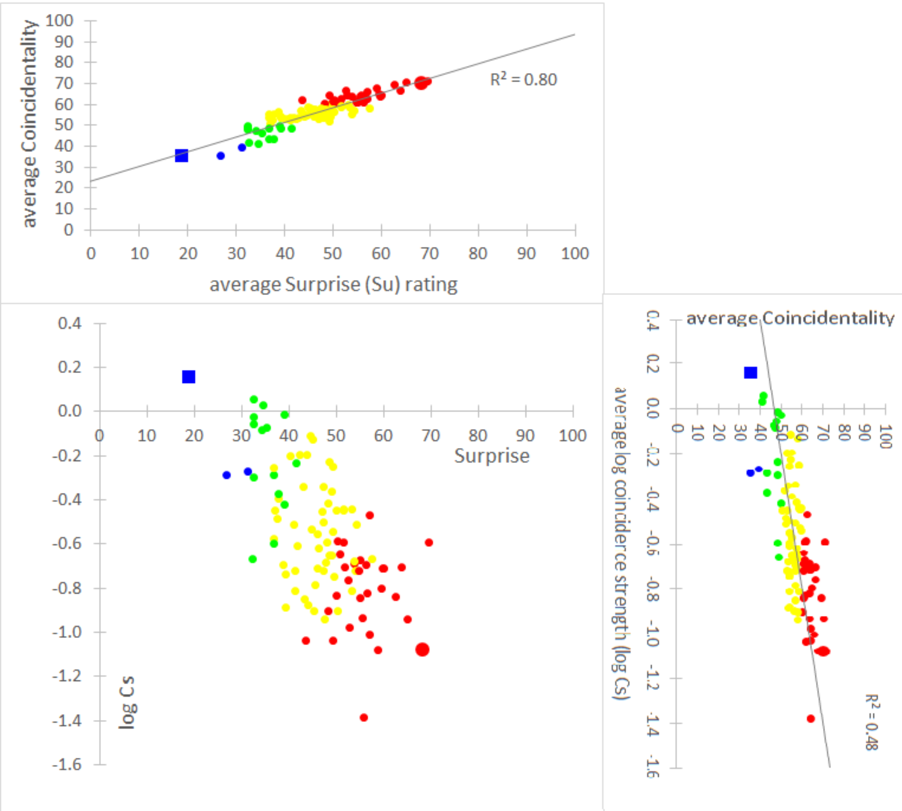
5.2 Results and Discussion

Experiment 3 replicated the key relationships between coincidentality, the two posteriors hypothesis judgments and the two coincidence likelihoods judgments; consistent with Experiment 2's results, Figure 3, average coincidentality here was negatively related to the causal posterior, $r = -0.62$, positively related to the chance posterior, $r = 0.68$, negatively related to the causal likelihood, $r = -0.76$, and negatively related to the chance likelihood, $r = -0.69$. More importantly, coincidentality was positively related to judged surprise, $r = 0.90$ for averaged data and $r = 0.52$ for combined (nonreversed) data. However, coincidentality and judged valence had no apparent relationship, $r = 0.04$ for averaged data. While the valence results are somewhat puzzling, a plausible reason for the lack of any apparent relationship is that the valence judgments were for someone else's rather than personally experienced events. Since they are not predictive of coincidentality here and not explicitly in our definition of coincidence, we don't consider them further.

Having measures of surprise and coincidence strength means that color-coded coincidentality can be represented in an analogue of the rotated Bayes' space from Experiment 2's Figure 6 as shown here in Figure 7. Notably, coincidentality increases toward the bottom right where surprise is high and the log of the posteriors belief ratio is negative, favoring the chance hypothesis. The combination of surprise and coincidence strength predicted coincidentality reasonably well, $R^2 = 0.84$ for averaged data and an average of $R^2 = 0.19$ for individual participants (model details in Appendix B). In this context, surprise by itself was a better predictor of coincidentality ($R^2 = 0.80$) than coincidence strength ($R^2 = 0.48$) for averaged data and for individual (nonreversed) participants, average $R^2 = 0.18$ versus 0.11 (Appendix B). However, about a quarter of the individual participants who did all the relevant judgments had coincidence strength as a better predictor of coincidentality than surprise, suggesting that both are relevant, consistent with the combined space (Figure 7).

As anticipated in the results of Experiment 2 (Figure 6), Bayesian surprise (the log of the likelihood ratio divided by the priors ratio) was positively related to judged surprise in this experiment, $R^2 = 0.12$. The underlying reason for this relatively poor predictiveness was that the (log) likelihoods ratio was only moderately (negatively) related to judged surprise, $R^2 = 0.11$, whereas the sum of the two coincidence likelihood judgments was strongly (negatively) related to surprise, $R^2 = 0.67$. And this suggests an alternative measure of surprise, data surprise.

Figure 7. Coincidentality in a space composed of surprise, central panel x-axis, and the log of the posteriors ratio, y-axis, for averaged data from Experiment 3, color coded by coincidentality. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to surprise on the x-axis in the top panel and in relation to the log of coincidence strength (the posterior hypothesis belief ratio) on the y-axis in the right panel.



Conceptually, a key aspect of surprising events is that they are unexpected/unanticipated from mechanisms known to the individual and, as such, improbable from these perspectives. When surprise was high, both coincidence likelihood judgments tended to be low. So a plausible reason that the sum of the likelihoods was a good (negatively related) predictor of surprise, and the their likelihood ratio was not, is that the two data likelihoods combine additively, after being multiplied by respective priors, to give an estimate of the overall probability of the data due to both known hypotheses: cause, $\mathrm{p}(\mathbf{d_k}|\mathbf{C_a}) \cdot \mathrm{p}(\mathrm{C_a})$, and chance, $\mathrm{p}(\mathbf{d_k}|\mathbf{C_h}) \cdot \mathrm{p}(\mathrm{C_h})$, in Equation 5. (Note that this combined probability of the data is in the denominator of the basic form of Bayes' theorem, Equation 1, but gets cancelled out in the process of taking Bayesian ratios, e.g., Equation 2.) Thus, events that are unlikely in the context of known causal hypotheses but also unlikely by chance have an overall combined probability that is small, and this corresponds to surprise being large. In contrast, if the combined probability of the events is near one, surprise should be minimal. This suggests that surprise should correspond to 1 divided by the combined probability of the data, Equation 6 left side, as this has the desired inverse relationship. In these equations, the k subscripts indicate "known" mechanisms for cause and change, that is, mechanisms that are sufficiently well understand and applied cognitive mechanisms in the mind of the individual as to generate expectations about what will happen in reality. For example, the occurrence of rain in the UK winter generates no surprise because of the experienced derived expectation that it's likely to rain on any given day. This formalization has the desired inverse relationship, but when the combined probability of the data due to the known hypotheses is one, 1 divided by that probability gives a minimum value of surprise as 1. But a more intuitive minimum value of surprise would be 0, the complete absence of surprise, so Equation 6 expresses surprise as 1 divided by the overall probability of the data given known causal and chance hypotheses *minus* 1. And this is equivalent to one minus the overall probability of the data divided by the probability of the data, as shown in the right-hand side of Equation 6. A useful consequence of this is that it puts surprise into an odds ratio form like the odds ratio form for coincidence strength (the right-hand side of Equation 3). The judgement scale responses (1 to 100) were converted to probabilities by dividing by 101 when calculating data surprise (rather than by 100 so as to avoid 0's and 1's symmetrically). And data surprise was a good predictor of averaged surprise, $R^2 = 0.67$. Data surprise can then replace Bayesian surprise in Equation 4 resulting in

Equation 5. $\mathrm{p}(\mathrm{d_k}) = \mathrm{p}(\mathbf{d_k}|\mathbf{C_a}) \cdot \mathrm{p}(\mathrm{C_a}) + \mathrm{p}(\mathbf{d_k}|\mathbf{C_h}) \cdot \mathrm{p}(\mathrm{C_h})$

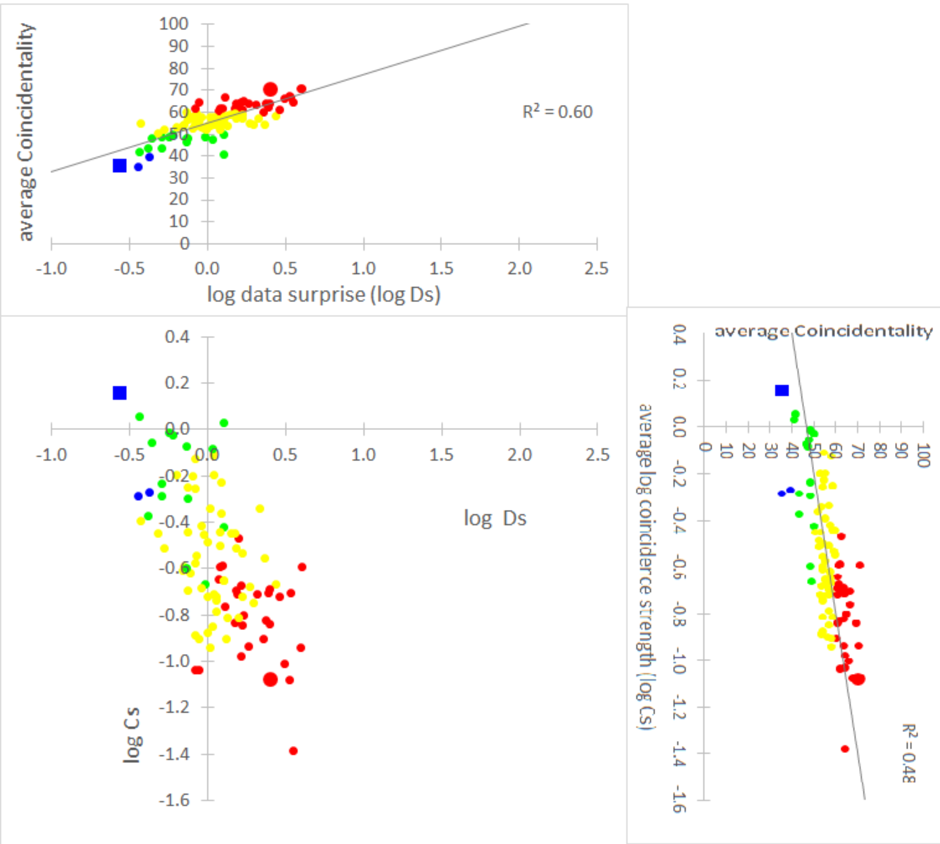
Equation 6. $\mathbf{S_u} \sim \frac{1}{\mathrm{p}(\mathrm{d_k})} - 1 = \frac{1-\mathrm{p}(\mathrm{d_k})}{\mathrm{p}(\mathrm{d_k})} = \mathrm{D_S}$

Equation 7; coincidentality is related to the ratio of two odds ratios, data surprise divided by coincidence strength. This corresponds to an updated coincidentality space, in Figure 8, with dimensions for data surprise and coincidence strength and coincidentality increasing toward the bottom right in the space.

Equation 7. $\mathbf{C_{\emptyset}} \sim \mathrm{C_{o_{scale}}} \cdot \log_e \left[\frac{\mathrm{D_S}}{\mathrm{C_S}}\right] + \mathrm{C_{o_{mid}}}, \qquad \mathrm{D_S} \approx \frac{1 - p(d_k)}{p(d_k)}, \qquad \mathrm{C_S} \approx \frac{p(\mathbf{C_a}|\mathbf{d})}{p(\mathbf{C_h}|\mathbf{d})} = \frac{1 - p(\mathbf{C_h}|\mathbf{d})}{p(\mathbf{C_h}|\mathbf{d})}$

Overall, surprise and coincidentality were strongly related, consistent with our definition. In addition, coincidentality is predicted by surprise and coincidence strength. However, data surprise is a better predictor of judged surprise than Bayesian surprise, thus providing more support for our

Figure 8. Coincidentality space, central panel, composed of log coincidence strength on the y-axis and log data surprise on the x-axis for averaged data from Experiment 3. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to log data surprise on the x-axis in the top panel and in relation to the log of coincidence strength (the posterior hypothesis belief ratio) on the y-axis in the right panel.



definition of coincidence than that of Griffiths and Tenenbaum (2007). To replicate and extend the findings from Experiment 3, Experiment 4 used the same hypothesis probability judgments but as a fully within-subjects design.

6. Experiment 4. Coincidentality, Coincidence Strength and Data Surprise

Coincidence strength (Griffiths & Tenenbaum, 2007), as the balance of support for the causal versus chance hypotheses, occurs in our definition as "[coincidences are] ultimately ascribed to change since the search for causal mechanisms has not produced anything more plausible than mere chance." Similarly, surprise occurs in our definition as "surprising pattern repetitions ... unlikely by chance...." So the practical purpose of this experiment was to measure coincidentality, the two posterior and the two likelihood probabilities for the causal and chance hypotheses, as combinations of these are sufficient to predict both coincidence strength and surprise within participants. Further, having all four terms allows assessment of whether Bayesian surprise (based on Griffiths and Tenenbaum, 2007) or data surprise is a better predictor of coincidentality at the level of individual participants. Thus, we pit these two formulations of surprise against each other in this experiment.

6.1 Methods

6.1.1 Participants

Twenty-two participants were recruited from a psychology participant panel at Cardiff University in exchange for partial course credit.

6.1.2 Materials, Design and Procedure

Coincidences, instructions and judgment scales for coincidentality, the causal and chance hypothesis posteriors, and the causal and chance likelihoods were the same as Experiment 3. The two posteriors were evaluated in a single, combined judgment task where participants provided the causal posterior judgment and then the chance posterior judgment for each coincidence in turn. Coincidentality and each of the two likelihoods were evaluated in fully separate judgment tasks.

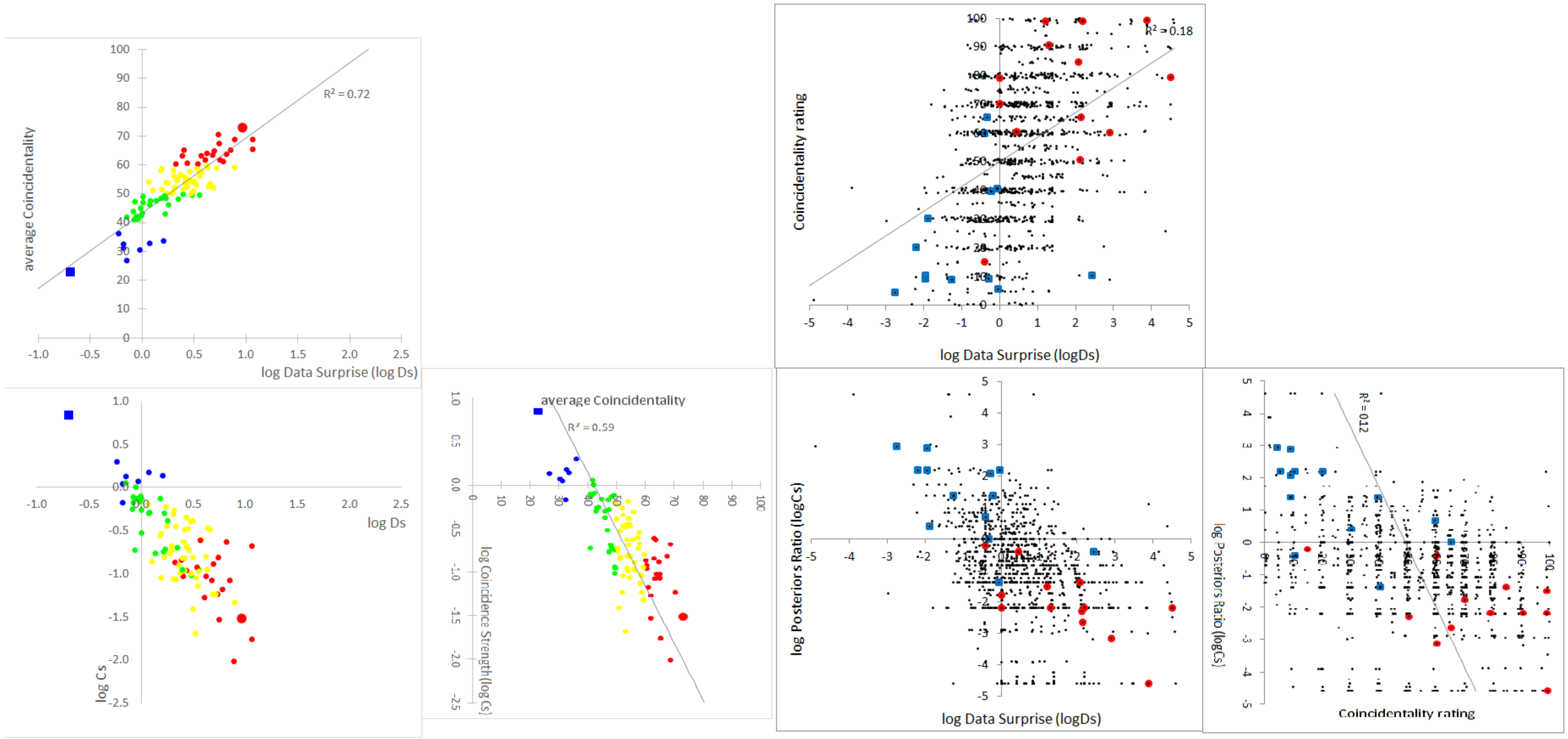
6.2 Results and Discussion

As in prior experiments, coincidence strength was a good predictor of coincidentality for averaged data, $R^2 = 0.59$, and for individual (nonreversed) participants, $R^2 = 0.20$ (model details in Appendix B). Data surprise was a substantially better predictor of coincidentality than Bayesian surprise for averaged data, $R^2 = 0.72$ in Figure 9 versus $R^2 = 0.40$, and a significantly better predictor for individual (nonreversed) participants--sign test $M(15) = 2$, $p = 0.007$, two-tailed (model fit details in Appendix B). The combination of Bayesian surprise and coincidence strength did a reasonable job of predicting coincidentality, $R^2 = 0.59$ for averaged data and an average $R^2 =0.23$ for individual (nonreversed) participants (Appendix B). However, coincidentality space, as the combination of data surprise and coincidence strength in Figure 9, did a substantially better job of predicting coincidentality for averaged data, $R^2 = 0.75$, and was a significantly better predictor for individual (nonreversed) participants, average $R^2 = 0.35$, sign test $M(15) = 2$, $p = 0.007$, two-tailed (model fit details in Appendix B). The combination of data surprise with coincidence strength is only a little bit better predictor than data surprise by itself, $R^2 = 0.72$ for averaged data and $R^2 = 0.31$ for individual (nonreversed) participants; however, there were individual participants for whom coincidence strength was a better predictor than data surprise. This along with the previous support for coincidence strength (Griffiths & Tenenbaum, 2007) implies that both coincidence strength and data surprise are important predictors of coincidentality. Further, this combination is the basis for a new, coincidentality space (Figures 8 and 9), the conceptual implications of which are developed below.

7. General Discussion

At their core, causal conclusions imply a decision between cause and coincidence, and thus both concepts are deeply implicated in learning, reasoning and goal directed behavior (Osman, 2014; Johansen & Osman, 2015). While a great deal of prior research has asked, “how causal is this?”, the present research has evaluated the nonredundant construct of coincidentality, “how coincidental is this?” (The nonequivalence of causality and coincidentality is spelled out below.) The present

Figure 9. Coincidentality space, central panels (e.g. bottom left panel), with color coded coincidentality (blue = low coincidentality, red = high), log data surprise on the x-axis and log coincidence strength on the y-axis. The color coding of coincidentality is shown versus data surprise, x-axis in the adjoining top left panels and versus coincidence strength, y-axis, in the adjoining bottom right panels. Left panels are averaged data and right are combined (nonreversed) individual participant data from Experiment 4. Combined data have a small amount of noise jitter to show data density.

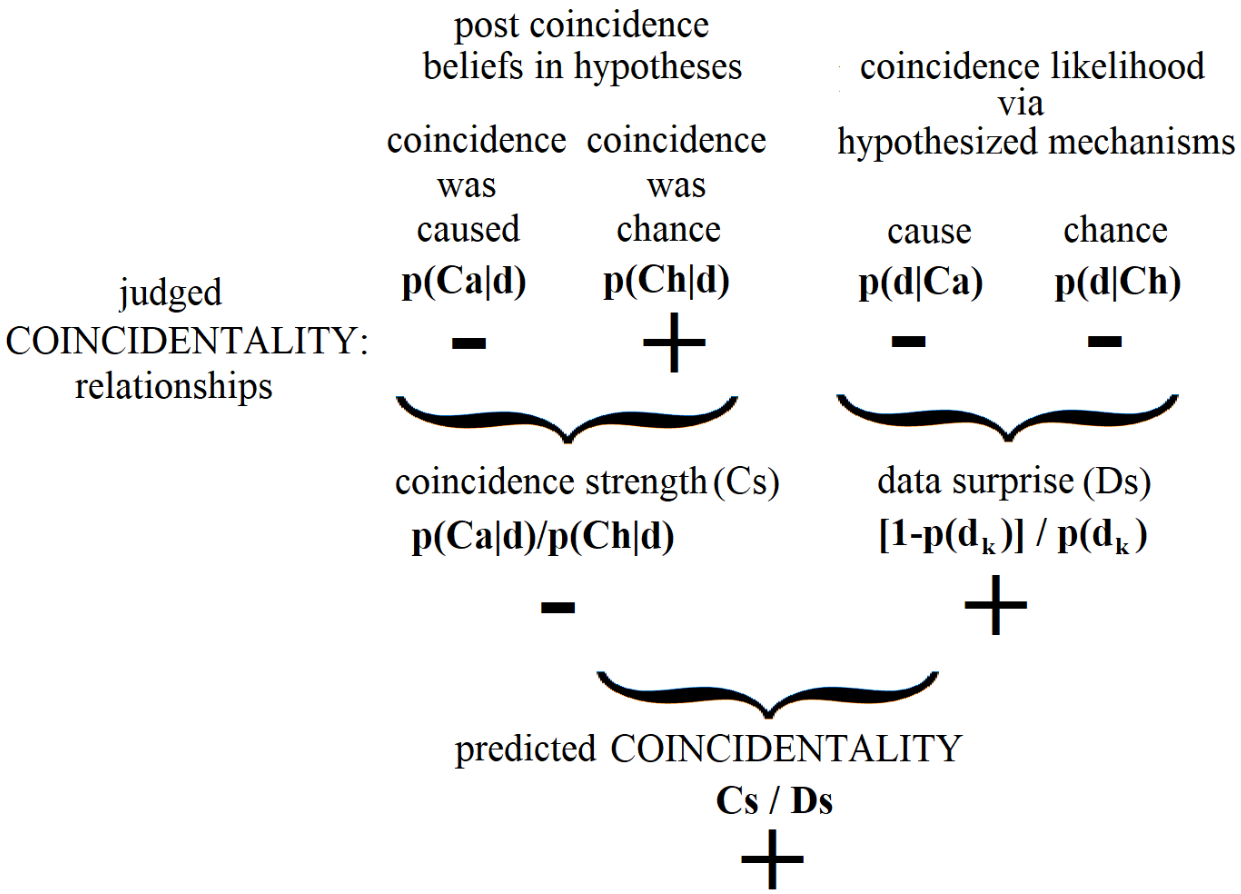


research also evaluated an array of possible predictors of coincidentality informed by components in our definition of coincidence and supporting that definition: “coincidences are surprising pattern [co-occurrence] repetitions that are observed to be unlikely by chance but are nonetheless ascribed to chance since the search for causal mechanisms has not produced anything more plausible than mere chance” (Johansen & Osman, 2015, p. 34). As such, the evaluation of co-occurrence events that are ultimately attributed to coincidence is important for understanding causal reasoning because, we argue, the same mechanisms underlie both prior to the categorical decision: cause versus coincidence. We also argue that the continuous construct of coincidentality, when perceived as high, is even more important as a suspicion flag that new causal knowledge may be available.

Four experiments measured coincidentality for a set of about a 100 naturally occurring coincidences that were self-reported personal experiences from an initial diary study. The experiments demonstrated that coincidentality is a stable psychological construct with substantial levels of agreement between participants and experiments in terms of correspondence with other psychological judgment variables. Methodologically it is worth noting that while the diary study prospectively collected “coincidences”, fortuitously these included some events which most people judge to be causal and hence not coincidences (e.g. the weekend-alarm coincidence, Table 1).

Across experiments, coincidentality was related to an array of judgments about event uncertainty, degree of belief and surprise as psychological constructs, summarized in Figure 10, that conceptually align with Bayesian probabilities for the truth of hypotheses given data (posteriors) and the probability of data given hypotheses (likelihoods) as discussed in the introduction: Coincidentality was negatively related to causal likeliness and positively related to chance likeliness, consistent with these being estimates of the posterior belief probabilities for causal and chance hypotheses, $p(Ca|d)$ and $p(Ch|d)$; in other words, given the events in the coincidence, what is the psychological degree of belief in the causal and chance hypotheses respectively? And this corresponds to the implied contrast in support for causal and chance hypotheses in the definition, “ascribed to chance [p(Ch|d) is large] since the search for causal mechanisms has not produced anything [p(Ca|d) is small]….” Further the construct of coincidence strength (Griffiths & Tenenbaum, 2007)—as relative belief in the causal and chance hypotheses, the ratio of these posteriors beliefs, p(Ca|d)/p(Ch|d)—was negatively related to coincidentality, the higher the relative belief in the causal over the chance hypothesis the *lower* coincidentality. However, unlike the posterior hypothesis beliefs which had opposing relationships with coincidentality, judgments of coincidence likelihood given causal and chance hypotheses, p(d|Ca) and p(d|Ch), were *both* negatively related to coincidentality. So the difference in the positive relationship between

Figure 10. Summary of key relationships found between judged coincidentality and its predictors. See main text for details.



coincidentality and the chance posterior belief, p(Ch|d), versus the negative relationship between coincidentality and the chance likelihood estimate, p(d|Ch), is consistent with the definition of coincidences as events that are “unlikely by chance [p(d|Ch) is small] but nonetheless ascribed to chance [p(Ch|d) is large]….” Furthermore, these two hypothesis beliefs and two data likelihoods can be used to specify all three Bayesian ratios contrasting causal and chance hypotheses (Equation 2 in the Introduction); the posterior belief ratio, p(Ca|d)/p(Ch|d); the data likelihoods ratio, p(d|Ca)/p(d|Ch), and the inferred prior belief ratio, p(Ca)/p(Ch). Moreover, this allows coincidentality to be represented in Griffiths & Tenenbaum’s (2007) coincidence strength space (Figure 1) composed of the prior beliefs and data likelihoods ratios (Experiment 2) as well as in a rotated coincidence strength space composed of coincidence strength, p(Ca|d)/p(Ch|d), and Bayesian surprise (Experiments 2, 3 and 4). Coincidentality was positively related to Bayesian surprise as the contrast between the data likelihoods ratio and prior beliefs ratio, that is as a ratio of these two ratios, [p(d|Ca)/p(d|Ch)]/[p(Ca)/p(Ch)]. In addition, coincidentality was, consistent with the definition, strongly positively related to judged surprise (Experiment 3), and judged surprise was positively related to Bayesian surprise (Experiment 3); however, surprise was better predicted by data surprise than Bayesian surprise (Experiment 3) where data surprise is an odds ratio, $[1 – p(d_k)]/p(d_k)$, based on the overall probability of the data, $p(d_k)$, in the context of the known cause and chance hypotheses. Finally, consistent with both surprise and relative support for the causal and chance hypotheses in the definition, coincidentality is well predicted by Equation 7 as conceptualized in a space composed of data surprise and coincidence strength (Experiments 3 and 4). Taken together, these predictors support our definition of coincidence and provide a way of answering the key question, how coincidental is this?

Not only are data surprise and coincidence strength predictors of coincidentality, the ratio D_s/C_s in Equation 7, they represent two conceptual constraints on coincidentality as reflected in where judgment data points do and don’t tend to occur: Although averaged data obscure this (because they represent only central tendency), the combined data plots in Figures 2 and 9 show it; coincidence strength, in the bottom of the ratio, corresponds to a kind of maximum conceptual boundary on coincidentality, Figure 9 far right panel, such that when coincidence strength is sufficiently high coincidentality must be low; put simply, co-occurrences that are clearly the result of a known causal connection are *not* coincidences. Data surprise, in the top of the ratio, corresponds to essentially the opposite constraint in terms of a kind of minimum bound on coincidentality such that when data surprise is sufficiently high, coincidentality *must* be at least somewhat high; i.e. surprising co-occurrence events are unexpected, that is not well anticipated from known causal and chance hypotheses and are unlikely in the context of both. This corresponds to surprising co-occurrences being at least somewhat coincidental because if either a causal hypothesis or chance were perceived to be likely true the events wouldn't be surprising anymore; events with clear, known explanations don't usually remain surprising (Foster & Keane, 2015; 2019; Maguire, et al., 2019; Schmidhuber, 2009). Importantly, these two constraints and the corresponding ratio of data surprise divided coincidence strength support a broader conceptual interpretation of the answer to the question, how coincidental is this?

The final empirical form of the model, Equation 7 composed of a ratio of the odds ratios for data surprise and coincidence strength, has a broader conceptual implication as shown in Equation 8; it indicates the extent of the support for the *possibility* that some unknown, as yet unconsidered, causal mechanism was responsible for the surprising co-incidence. The basic intuition here for an experienced co-incidence is that if the probability of the known causal mechanism under consideration is sufficiently small, then that might be enough to reject that causal mechanism and favor chance if the probability due to chance is reasonably large, i.e. "just a chance co-occurrence". This corresponds to relative posterior belief in the chance hypothesis over the known cause hypothesis, the posterior belief ratio being close to zero and the log ratio being negative. But what if the probability of the co-incidence is also extremely unlikely due to chance and the posterior chance hypothesis belief probability is also small? For some co-incidences of this kind, the posterior belief ratio could still have a value reasonably close to zero indicating that there was *more* support for the chance hypothesis than the known cause hypothesis. However, that might not correspond to compelling support for the chance hypothesis if the overall probability of the data in the context of both known cause and chance hypotheses, $p(d_k)$ Equation 5, is sufficiently small because that starts to lead to the *suspicion* that some as yet unknown/unconsidered causal mechanism was responsible for the surprising co-incidence. Essentially this corresponds to extending the hypothesis space from hypotheses for known things—known causal mechanisms and chance, the first two terms in equation 9—to include an unknown cause hypothesis with a probability nontrivially greater than zero, the third term in Equation 9.

Equation 8. $\log_e \left[\frac{\left(\frac{p(c_{\neg k} | \mathbf{d_k})}{(1 - p(c_{\neg k} | \mathbf{d_k}))} \right)}{\left(\frac{p(\mathbf{C_a} | \mathbf{d_k})}{(1 - p(\mathbf{C_a} | \mathbf{d_k}))} \right)} \right] \sim \mathbf{C}_{\emptyset} \sim \log_e \left[\frac{\max(\mathbf{S_u})}{\min(\mathbf{C_s})} \right] = \log_e \left[\frac{\left(\frac{1 - p(\mathbf{d_k})}{p(\mathbf{d_k})} \right)}{\left(\frac{p(\mathbf{C_a} | \mathbf{d_k})}{p(\mathbf{C_h} | \mathbf{d_k})} \right)} \right] = \log_e \left[\frac{\left(\frac{1 - \left(\mathbf{p(d_k | C_a) \cdot p(C_a) + p(d_k | C_h) \cdot p(C_h)} \right)}{\mathbf{p(d_k | C_a) \cdot p(C_a) + p(d_k | C_h) \cdot p(C_h)}} \right)}{\left(\frac{\mathbf{p(d_k | C_a) \cdot p(C_a)}}{\mathbf{p(d_k | C_h) \cdot p(C_h)}} \right)} \right]$

Equation 9. $\{p(\mathbf{C_a} | \mathbf{d_k}) + p(\mathbf{C_h} | \mathbf{d_k})\} + p(C_{\neg k} | \mathbf{d_k}) = 1$

Equation 10. $p(d) \; = p(\mathbf{d_k} | \mathbf{C_a}) \cdot p(C_a) + p(\mathbf{d_k} | \mathbf{C_h}) \cdot p(C_h) + p(d_{\neg k} | C_{\neg a}) \cdot p(C_{\neg a})$

Equation 11. $1 - p(d_k) = p(d_{\neg k} | C_{\neg a}) \cdot p(C_{\neg a})$

Put another way, this implied discrepancy (as a kind of randomness deficiency, Li & Vitányi, 2019; Maguire, et al., 2019) between the actually occurring co-incidence and the probability of that co-incidence in the context of the known cause and chance hypotheses is just too great for either one to be plausible when that probability of the data is very small. But the discrepancy can be reduced by the operation of an unknown causal mechanism which at least conceptually makes the observed co-occurrence more *likely*. To emphasize, this is not about jumping to definite conclusions about the existence of some new causal mechanism, like a believer might do, or dismissing the possibility of a new mechanism, like a skeptic might do, but rather having sufficient suspicion to seek out more *evidence*, like a scientist would do.

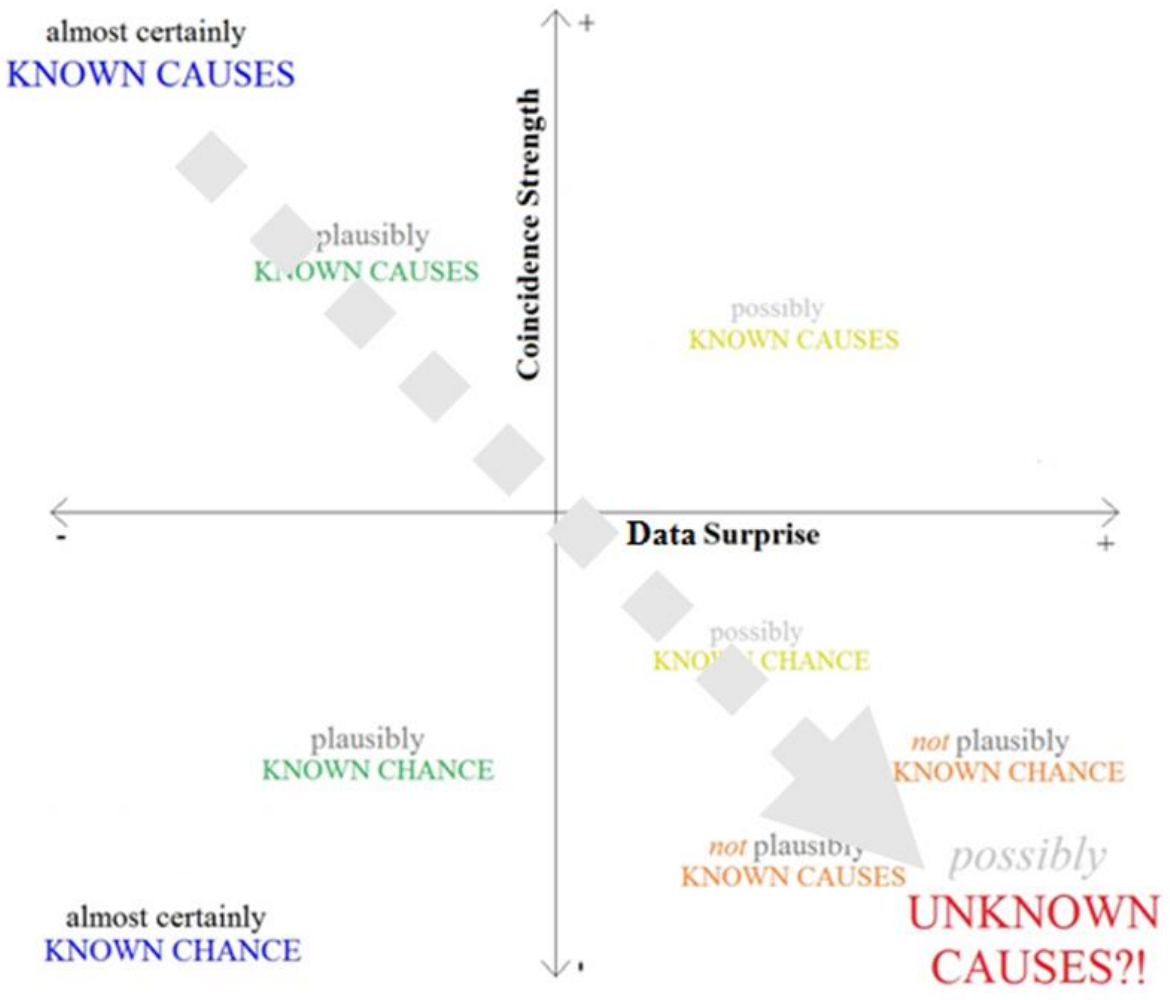
As discussed above, co-occurrence surprise corresponds to a co-incidence being substantially unexpected by known cause and chance. So surprise is related to the discrepancy between 1 and the probability of the data from known cause and chance: the higher the known probability of the data, the smaller the numerator of data surprise in Equation 6, the bigger the denominator and the smaller the data odds ratio and the smaller the surprise. But if the probability of the data for known mechanisms is small, the discrepancy between that probability and 1 in the numerator is large, the odds ratio is large and surprise is high. Conceptually then the 1 in the numerator of data surprise corresponds to the overall probability of the data occurring as a result of both known mechanisms (cause and chance) and *unknown* mechanisms, p(d) in Equation 10, being as (hypothetically) high as it possibly can be, i.e. one, and this corresponds to surprise being as high as it can be. However, there's no reason why the probability of the data via all possible mechanisms has to be a certainty, one. The true overall probability of the data even with the possibility of the unknown cause operating might very well be less than 1. If the 1 in the numerator of data surprise is replaced by the unknown p(d), the overall probability of the data for both known (cause and chance) and unknown mechanisms, then data surprise on the right hand side of Equation 8, is equivalent to the odds ratio supporting the probability of an unknown cause hypothesis, $p(c_{\neg k}|d_k)$, on the left hand side. And as the coincidence strength ratio composed of the known causal hypothesis divided by the chance hypothesis is also equivalent to the odds ratio for the known cause, the model composed of data surprise and coincidence strength is equivalent to a conceptual model of the odds ratio for the unknown cause divided the odds ratio for the known cause.

Unlike coincidence strength based on known mechanisms, this ratio of odds ratios is not an estimate of the probability that an unknown causal mechanism is operating but an indication of the maximal amount of support there *might* be for such a mechanism, Equation 11. The essential reason for this is that the overall probably of the data, p(d), is unknown because it includes the unknown probability of the data due to an unknown causal mechanism. But still the highest the probability of the data can be is one. So data surprise is an estimate of maximum suspicion, Equation 11, that is an indicate of what the highest estimated probability of the unknown cause hypothesis can reasonably be given the data, the third term in equation 9. On the other hand, coincidence strength based solely on known mechanisms represents a kind of minimum suspicion that an unknown mechanism might be operating. So coincidentality represents, not the true probability of an unknown causal hypothesis, but rather a ratio of the maximum to minimum support values that a new mechanism *might* be operating and hence a good suspicion trigger to gather more evidence. It is also worth noting that this final conceptualization, while not identical to the definition of coincidence proposed by Griffiths and Tenenbaum (2007), is similar in terms of a contrast between standard and nonstandard accounts.

In a more intuitive spatial form, coincidentality as related to causal discovery can be represented in a space composed of data surprise and coincidence strength, Figure 11. The left-hand side of the space corresponds to data surprise being low with the conclusion determined by coincidence strength, known causes in the top left quadrant or known chance in the bottom left quadrant. If coincidence strength favors cause over chance but data surprise is still high, the top right-hand quadrant, then it's *possible* that that known causes apply, though not necessarily probable. But if coincidence strength is low and data surprise is high, the bottom right hand quadrant, then this invokes the possibility that a new mechanism *might* be operating, and more strongly so toward the bottom right hand of the space. This is where great scientific discoveries exist along with very weird, improbable coincidences (for examples of both see Johansen & Osman, 2015). So when faced with a new co-incidence which of these is the right interpretation for a given context is only likely to be decided by subsequent evaluation and evidence. But this coincidentality ratio of ratios provides a quantitative estimate of the potential for a new causal mechanism and as such can be a guide for which merit further evaluation and which don't. So while we are proposing a psychological theory of coincidences, central to this perspective is the assessment of uncertain belief and data surprise probabilities, and these, we argue from the Bayesian perspective, cannot be too inaccurate or fundamental mechanisms for co-incidence perception in causal discovery wouldn't work.

Despite the conventional wisdom in judgment and decision making that people are bad at probabilities, Bayesian perspectives have argued the opposite. Griffiths and Tenenbaum (2006) showed that peoples' judgments about a range of real phenomena are reasonably accurate and consistent with underlying Bayesian probabilities. Lewandowsky, Griffiths and Kalish (2009) used an iterated design to argue that people update their judgments in an optimal Bayesian way, even for impoverished data. And the close correspondence between probability estimates and Bayesian models has been shown in children as young 4-5 years old (e.g. Denison, Bonawitz, Gopnik, & Griffiths, 2013). Not surprisingly, the Bayesian optimality perspective has been critiqued as for example Griffiths, & Tenenbaum, (2014) argue that people trade off decision accuracy against the

Figure 11. Causal discovery in coincidentality space composed of (log) coincidence strength, y-axis, and (log) data surprise, x-axis, as in Figure 9. The large gray arrow corresponds to coincidentality increasing toward the bottom right-hand side of the space, as per equation 7. Details in main text.



effort of obtaining data, so on an individual basis, judgments can seem more noisy and sub-optimal but globally approximate Bayesian optimality. In the present context, whether or not probability estimates are optimal in a Bayesian sense, we argue for a less strong conclusion for coincidences: people have some ability to discriminate the merely improbable from the extremely unlikely. Coincidence perception and causal discovery are related psychological phenomena associated with at least somewhat accurate assessments of improbability, particularly in the context of chance.

Probability judgment can't be that defective or causal reasoning would be impossible; to draw a causal conclusion from data, a scientist needs to be able to distinguish between extremities probabilities of the data being due to chance, the unlikely from the extremely unlikely; one or two co-occurrences may be easily attributed to chance, while 100 such co-occurrences may have such a vanishingly small probability by chance as to compellingly support the causal hypothesis. The ability to assess something as "unlikely by chance" can't be too bad or its difficult to see how scientists could ever be entitled to draw causal conclusions from empirical data, at least in the absence of formal statistical calculations, as the chance hypothesis could never be compellingly eliminated. Collecting more data has to drive the implied probability of the data by chance to such a low value that the chance hypothesis can be discarded in favor of the causal hypothesis as an addition to causal knowledge or when would enough data actually be enough data? So not only is the ability to assess improbability-by-chance at the heart of causal search, it is at the heart of causal knowledge as well.

As an example, one of the greatest empirical observations of all time is Fleming's discovery of the antibiotic properties of penicillin mold. Fleming was a biologist who was specifically looking for substances that would kill disease causing bacteria; so he was very much in targeted causal search mode. Penicillin mold is fairly common in human environments because it grows easily on bread, but needless to say, bacteriologists have always gone to some lengths to keep their cultures from being contaminated by things other than the bacteria they want to study. The basic idea of culturing bacteria is that to study a particular kind of bacteria they need to be "grown" in pure form. So a bacterial culture essentially consists of a support medium and chemicals that either help the bacteria multiply or don't and they die off. The problem for bacteriologists is that cultures get contaminated by undesirable things all the time, and such contaminated cultures get discarded. The interesting aspect of this discovery isn't that a culture serendipitously happened to get contaminated by penicillin mold (you may have some growing in your kitchen) but rather that Fleming noticed. And in the context of co-incidences it is worth considering what he noticed: In simplistic terms bacteria in cultures do two basic things, they grow (multiply) and they die. Because most bacteria are extremely small, a relatively young bacterial culture as seen by the naked eye includes some spatial regions where bacteria are clearly growing and some where they aren't. There are two basic explanations for why bacteria aren't clearly growing in a specific region: One reason is just chance. That is, when a culture is "seeded" with bacteria, exactly where in space growths appear has a random component to it resulting from the underlying random positions of individual bacteria. A known causal reason for why bacteria aren't clearly growing at a particular spatial position can be that the nutrients in that location have already been used up or possibly never existed. Essentially what Fleming must have noticed was the spatial co-incidence of contaminant growth and culture death. Because of the random component for exactly where things grow in a culture for both the contaminant and the bacteria being cultured, it is mathematically possibly that the spatial co-incidence of the two occurred by chance. While it is challenging to specify this probability exactly, what is clear is that it was fairly small. In fact, it seems incredibly implausible that Fleming even considered the possibility that the surprising spatial co-incidence was due to chance. Far more plausibly, he formed the suspicion that a new causal mechanism was operating because it was the kind of mechanism he was looking for. He and other scientists then went on to amass overwhelming evidence for this mechanism. In broader terms, peoples' probability judgments may not be perfectly accurate, but they can't be too poor. More importantly, the skeptical view that most chance events will occur given enough time and people misses the point (see the summary of this perspective in Johansen & Osman, 2015): People *are* sometimes capable of recognizing the extreme improbability of some co-incidences when they see them and *sometimes* go on to make new causal discoveries as a result. The construct of coincidentality in Equation 8 provides an explicit approach to deciding when further evaluation is warranted, i.e. when known cause and chance mechanisms aren't judged plausible.

The present research has also provided partial support for Griffiths and Tenenbaum's (2007) Bayesian ratio perspective as an account of coincidentality while also flagging a limitation of this perspective: The conceptual limitation of the Bayesian ratios perspective is that it is based on Bayesian odds ratios where the unknown overall probability of the data (in the denominator of the basic form Bayes' theorem, Equation 1), occurs in both the numerator and the denominator respectively for the two hypotheses under consideration and so gets cancelled out, Equation 2. Thus, the Bayesian odds ratio form using the posterior hypothesis probabilities effectively treats the component probabilities as *relative* to each other rather than absolute. That is, the posterior odds ratio quantifies the relative extent to which one hypothesis is favored over the other, e.g. if the posterior causal hypothesis given the data has a probability of 0.8 and the chance hypothesis 0.1 then the posteriors odds ratio 0.8/0.08 = 10 says there is 10 times more support for the causal hypothesis than the chance hypothesis. In some respects, this emphasis on relative support makes perfect sense because the overall probability of the data in the denominator of Bayes' theorem is commonly unknown. There's always a chance that some unconsidered mechanism was responsible, so conceptually this should be included in the overall probability of the data except its value is unknown. This absolute uncertainty about the overall probability of the data may seem irrelevant when the goal is to evaluate the *relative* hypothesis support in terms of Bayesian posteriors ratio. But the reason the overall probability of the data matters is that the interpretation of a given Bayesian posteriors ratio, say 10, changes quite dramatically depending on this probability: If the overall probability of the data from known causal mechanisms and chance is high, then a posteriors ratio of 10 provides quite compelling support for the specified causal hypothesis (as in the example above).

But a Bayesian posteriors ratio of 10 can also arise out of absolute posterior probabilities that are extremely small, e.g. 0.0000000001/0.00000000001 = 10. In this case the appropriate conclusion is not that this posteriors ratio of 10 provides reasonably compelling support for the specified causal hypothesis over the chance hypothesis, but rather than neither of the hypotheses in hand is very strongly supported and some as yet unknown mechanism might be in operation. That is, some unknown mechanism is needed to account for the occurrence of the data and thus make the overall, conceptually expected probability of the observed pattern more consistent with the actual events.

In broader terms, the present research contributes to but also highlights the need for a better specification of a normative conceptualization of coincidence evaluation as an intrinsic part of causal reasoning and discovery. Historically causal discovery, especially via the mechanisms of science, has been an predominantly human, psychological activity; however the rise of computer automation and various kinds of artificial intelligence, especially machine learning, strongly suggest these will play an increasing role in the scientific process in terms of both large scale data acquisition and causal reasoning based on that data. But it is also very plausible that the development of this automation will be and should be informed by the psychological evaluation of how people make causal discoveries (Fenton & Neil, 2012). The specification of how people normatively *ought* to engage in causal reasoning is going to be strongly informed by how they actually do causal reasoning given the prior success of science (Pearl, 2019).

In conclusion, coincidences provide an important perspective on the mechanisms of causal reasoning and as such the concept of coincidentality is central to the mechanisms of causal learning operating in individuals, science and society. That is, coincidentality is a cognitive marker that new causal knowledge *might* be available. In short, coincidentality as “how coincidental is this?” is a cognitive analogue for an ancient map’s “here be dragons!” in regions of uncertainty; the risk averse shy away, but the bold learn.

Acknowledgements

We thank Christos Bechlivanidis, Leonard Boss, Marc Buehner, Jacques Grange, Gabriel Lins de Holanda Coelho, Paul Hanel, Rob Honey, Bjoern Meder, Emma Morgan, David Shanks, Ulrich von Hecker and Peter White for comments on prior drafts of this manuscript. We also would like to thank our many students who have forced us to clarify our thinking and have helped with this research in many ways including Beth Adams, Calum Davies, Ross Garland, Kathryn George, Billy Harris, Emma Harris, Emma Keltz, Natalie Richards and Maria Azcoitia Revuelta. Finally, for Kotapolis Kotovich, a cat who has taken his curiosity with him.

References

Baldi, P, & Itti, L. (2005). Attention: Bits versus Wows. *International Conference on Neural Networks and Brain*, 56-61.

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

Beitman, B. D. (2009). Brains seek patterns in coincidences. *Psychiatric Annals*, *39*, 255-264.

Blackmore, S. J. (1984). A postal survey of OBEs and other experiences. *Journal of the Society for Psychical Research*, *52*, 225–244.

Blackmore, S. J. (1992). Psychic experiences: Psychic illusions. *Skeptical Inquirer*, *16*, 367–376.

Blackmore, S., & Troscianko, T. (1985). Belief in the paranormal: Probability judgements, illusory control, and the ‘chance baseline shift’. *British Journal of Psychology*, *76*, 459–468.

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology*, 16, 17-34.

Brugger, P., Landis, T., & Regard, M. (1990). A ‘sheep-goat effect’ in repetition avoidance: Extra-sensory perception as an effect of subjective probability? *British Journal of Psychology* 81, 455-468.

Brugger, P., Regard, M., Landis, T., & Graves, R.E. (1995). The roots of meaningful coincidence. *The Lancet*, 345, 1306-1307.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*(2), 367-405.

Coleman, S. L., & Beitman, B. D. (2009). Characterizing high-frequency coincidence detectors. *Psychiatric Annals, 39*, 271-279.

Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., & Kleinberg, J. (2010). Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences, 107*(52), 22436-22441.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*(2), 285-300.

Dessalles, J. L. (2008). Coincidences and the encounter problem: a formal account. Proceedings of the 30th annual conference of the Cognitive Science Society, 2134-2139.

Diaconis, P., & Mosteller, F. (1989). Methods for studying coincidences. *Journal of the American Statistical Association*, *84*, 853-861.

Diaconis, P., & Mosteller, F. (2006). Methods for studying coincidences. In *Selected Papers of Frederick Mosteller* (p. 605-622). Springer, New York, NY.

Falk, R. (1989). Judgment of coincidences: Mine vs. Yours. *American Journal of Psychology*, *102,* 477-493.

Falk, R., & MacGregor, D. (1983). The surprisingness of coincidences. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analysing and aiding decision processes* (pp. 489-502). Budapest: Adademiai Kiad6.

Fenton, N., & Neil, M. (2012). *Risk assessment and decision analysis with Bayesian networks*. Crc Press.

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology, 81*, 74–116.

Foster, M. I., & Keane, M. T. (2019). The role of surprise in learning: Different surprising outcomes affect memorability differentially. *Topics in Cognitive Science, 11*, 75-87.

Fugelsang, J. A., & Thompson, V. A. (2000). Strategy selection in causal reasoning: When beliefs and covariation collide. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *54*(1), 15-32.

Gill, J. (2014). *Bayesian Methods: A Social and Behavioral Sciences Approach* (Vol. 20). CRC press.

Glicksohn, J. (1990). Belief in the paranormal and subjective paranormal experience. *Personality and Individual Differences*, *11*, 675–683.

Griffiths, T. (2017). Formalizing prior knowledge in causal induction. *The Oxford Handbook of Causal Reasoning*, 115.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767-773.

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

Hanley, J. (1984). “Lotteries and Probabilities”: Three cases reports. *Teaching Statistics*, *6*, 88-92.

Hanley, J. (1992). Jumping to coincidences: Defying odds in the realm of the preposterous. *The American Statistician*, *46*, 197-202.

Henry (1993). Coincidence experience survey. *Journal of the Society for Psychical Research*, 97-108.

Houran, J., & Lange, R. (1996). Hauntings and poltergeist-like episodes as a confluence of conventional phenomena: A general hypothesis. *Perceptual and Motor Skill*, *83*, 1307–1316.

Izard, C. E. (1991). Surprise-Astonishment. In *The Psychology of Emotion*. (pp. 177-183). Plenum, New York, NY

Jenkins, H., & Ward, W. (1965). Judgment of contingencies between responses and outcomes. *Psychological Monographs, 7,* 1-17.

Johansen, M. K., & Osman, M. (2015). Coincidences: A fundamental consequence of rational cognition. *New Ideas in Psychology, 39,* 34-44.

Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The wisdom of individuals: Exploring people's knowledge about everyday events using iterated learning. *Cognitive Science*, *33*(6), 969-998.

Li, M., & Vitányi, P. (2019). *An Introduction to Kolmogorov Complexity and Its Applications*. 4th Ed. New York: Springer.

Maguire, P., Moser, P., Maguire, R., & Keane, M. T. (2019). Seeing patterns in Randomness: A computational model of surprise. *Topics in Cognitive Science, 11,* 103-118.

Matthews, R., & Blackmore, S. (1995). Why are coincidences so impressive? *Perceptual and Motor Skills, 80*(3_suppl), 1121-1122.

Matthews, R, & Stones, F. (1989). Coincidences, the truth is out there. *Teaching Statistics*, *1*, 17-19.

Mill, J. S. (1843). *A System of Logic.* Honolulu: University Press of the Pacific.

Mock, C., & Weisberg (1992). Political innumerary: Encounters with coincidence, improbability, and chance. *American Journal of Political Sciences*, *36*, 1023 1046.

Navarro, D. J., & Kemp, C. (2017). None of the above: A Bayesian account of the detection of novel categories. *Psychological Review*, *124*(5), 643.

Osman, M. (2014). *Future-minded: The psychology of agency and control*. Macmillan International Higher Education.

Pacer, Michael D. *Mind as Theory Engine: Causation, Explanation and Time.* PhD dissertation, University of California, Berkeley, 2016.

Pearl, J. (1982). *Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach* (pp. 133-136). Cognitive Systems Laboratory, School of Engineering and Applied Science, University of California, Los Angeles.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, England: Cambridge University Press.

Pearl, J. (2014). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, *62*(3), 54-60.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II: Current theory and research* (pp. 64-99). New York: Appleton-Century Crofts.

Schmidhuber, J. (2009). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In *Workshop on Anticipatory Behavior in Adaptive Learning Systems* (pp. 48-76). Springer, Berlin, Heidelberg.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development, 47*, 1-51.

Shultz, T. R., Fisher , G. W., Pratt , C. C., & Rulf, S. (1986). Selection of causal rules. *Child Development*, 57, 143-152.

Spirtes, P., Glymour, C., and Scheines, R. (2000/1993). *Causation, Prediction and Search* (2nd edition). Boston, MA: MIT Press.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review, 124*(4), 410.

Tobacyk, J. (1995). What is the correct dimensionality of paranormal beliefs?: A reply to Lawrence's critique of the paranormal belief scale. *The Journal of Parapsychology*, *59*, 27–46.

Tobacyk, J., & Milford, G. (1983). Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. *Journal of Personality and Social Psychology*, *44*, 1029–1037.

Watt, C. (1991). Psychology and coincidences. *European Journal of Parapsychology, 8,* 66-84.

Appendix A: data tabulation and analysis

A.1. Handling of missing, ambiguous and repetitive responses

Across all experiments, the tabulation of the data for presentation in the figures involved removing ambiguous data points, e.g. 606, that were not actually on the judgment scales (1 to 100). There were only a handful of off-scale data points across all the experiments with the exception that there were quite a few 0 responses even though technically the minimum scale value was 1. As the difference between a 0 versus a 1 is minute in terms of presented data in figures, these 0's were left as 0's for data in the figures for accuracy. Similarly, across experiments, participants occasionally failed to provide individual judgments in a given task and these missing data points were simply not included in the data in figures.

The tabulated averaged data for model evaluation was identical to the averaged data presented in the figures. However, the tabulation of the data for evaluating models on individual participants handled missing and out of range values somewhat differently: 0 judgments were replaced with 1 judgments to keep various models from collapsing predictions to zero. Other missing and out of range judgments were replaced with participant-wise averages of the judgments they did supply in a given judgment task so as to be as theoretically neutral as possible while still allowing the models to be fully applied to each individual data set. In addition, a handful of participants across the entire set of experiments reported made repetitive responses to most or all items in a given judgment task, e.g. making coincidentality judgments of 100 for all items. These participants judgments are presented in the figures, but these participants were not evaluated when fitting models to individual participants as there was no variance in the judgments for the models to account for. Lastly participants occasionally failed to provide any judgments in a given judgment task, e.g. because they ran out of time. These partial data sets are included in the figures but again the models were not applied to these individual participants because none of the judgments in a given task were present.

A.2. Handling of judgment scale reversals

Scale reversals occurred for a minority of participants in all the experiments reported here. Correlation matrices provide a compact way of showing when participants probably reversed particular judgment scales. Figure A.1 shows the correlation matrices for each judgment task in every experiment. Each matrix shows the correlation between each individual participant to each other individual participant in that task as indicated by a cell in the matrix that is either black, a correlation between a pair of participants more negative than r = -0.1, or white indicating a correlation greater than r = -0.1. Thus each individual row (also each column) indicates how a given participant's judgments related to all other individual participants in that task. Systematic judgment scale reversals flag up in terms of systematic negative correlations to other individual participants, that is as horizontal (and vertical) black lines. Further, scale reversed participants tended to be positively correlated with each other corresponding to white boxes where black horizontal and vertical lines intersect. As can be seen, scale reversals clearly occurred in all judgment tasks, but only for a relatively small minority of participants in any given task. Thus the dominant interpretation of each scale was used by the substantial majority of participants, and the final experiment, Experiment 4, had especially few apparent reversals.

Figure A1. Correlation matrices with correlations between individual participants for each of the main judgment tasks (rows of panels) in each experiment (columns of panels). Each row (and each column) in a given matrix corresponds to the correlations between a given participant's judgments in a task for all 97 coincidences individually to the judgments from each other participant. Black cells correspond to a negative correlation coefficients (r) less than or equal to (i.e. more negative than) -0.1, and white cells to correlations bigger than this.


1
2
3
4
Co rating
Co ratings
Co
Co
Ca rating
Ca|d
Ca|d
Ca|d
Ch ratings
Ch|d
Ch|d
Ch|d
d|Ca
d|Ca
d|Ca
d|Ch
d|Ch
d|Ch

Appendix B: model procedures and model fit details

Model free parameters were adjusted (via hill-climbing and/or simulated annealing) to minimize a fit function of the disparity between the model predictions and the data. This fit function was either sum-squared error (SSE) or weighted sum-squared error (wSSE). Note that in the main text the SSE fits are reported in terms of percentage of variance accounted for, R^2. The weightings used in wSSE were 1 divided by the variance of the mean for a given coincidence, i.e. the sample variance of the coincidentality judgments for that coincidence divided by the number of such judgments, when fitting averaged data, and 1 divided the coincidence variance estimated from a variance model when fitting individual participant data sets.

The variance model is needed to provide variance estimates for individual participant judgments that occurred outside of the range of the averaged judgments, i.e. although individual participants gave coincidentality judgments of 100, none of the coincidentality judgments averaged across participants were this extreme, so a variance model was needed to provide variance estimates for such judgments. The variance model was a parabola, x*(100-x) multiplied by a variance scaling coefficient that was fitted to the coincidence sample variances as a function the coincidence means for a give experiment with an added minimum variance intercept of 1. When a model made a prediction beyond the range of the judgment scale, it was associated with a variance estimate corresponding to the nearest value on the judgment scale. The variance model had the effect of smoothing out differences in the coincidences variance for a given range of coincidentality judgments as well as providing smaller variance estimates for judgments toward the ends of the judgment range. Overall, while we have included wSSE fits in the interests of completeness, the differences between SSE and wSSE fits across models and data sets were generally very small, with both supporting the same conclusions.

Best fitting model parameters and fit values for all models and data sets discussed in the main text are in the tables below by experiment. The SSE fits are in the left columns and the wSSE fits are in the right columns. Note that the percentage of variance accounted for in the wSSE fits is the *unweighted* percentage of variance. Thus the generally minute decline in percentage of variance for wSSE fits compared to SSE fits shows that the weighted and unweighted fits were generally very similar.

Table B.1. Experiment 1 modeling results. Log inverse coincidence strength model (Equation 3) fit summary for averaged data, first row, and the average of individual participant fits, second row, based on sum-squared error (SSE), left columns, and weighted sum-squared error (wSSE), right columns

	SSE fits				wSSE fits			
	% var	SSE	CoScale	CoMid	% var	wSSE	CoScale	CoMid
Averaged	**0.698**	1779	16.6	45.2	0.697	70.0	17.0	45.2
Individual	**0.264**	33631	8.2	48.0	0.247	44.6	6.9	48.2

Table B.2. Experiment 2 modeling results. Log inverse coincidence strength model fit summary for averaged data, first row, and log Bayesian surprise model, second row, based on sum-squared error (SSE), left columns, and weighted sum-squared error (wSSE), right columns. The third row is the combined log coincidence strength and Bayesian surprise model (Equation 4).

	SSE	fits				wSSE	fits			
model	% var	SSE	CsScale	SuScale	CoMid	% var	wSSE	CsScale	SuScale	CoMid
Ln Cs	**0.530**	4902	17.3		44.6	0.523	286.8	18.8		44.5
Ln Bs	**0.490**	5312		11.3	40.0	0.486	330.2		12.0	39.8
Ln[Bs/Cs]	**0.545**	4739	11.8	4.3	42.4	0.538	280.4	14.3	3.4	42.7

Table B.3. Experiment 3 model results. Log inverse coincidence strength model fit summary for averaged data, first row, and individual participants, second row, based on sum-squared error (SSE), left columns, and weighted sum-squared error (wSSE), right columns. The judged surprise model (Su) is in rows three and four, and an combined surprise and log inverse coincidence strength model, rows five and six (Equation 4). The models in the top rows are all models of coincidentality. The two models in the bottom rows, log Bayesian surprise and log data surprise, are models predicting judged surprise rather than coincidentality.

Coincidentality		SSE		fits					wSSE fits		
models		%var	SSE	CsScale	SuScale	CoMid	%var	wSSE	CsScl	SuScl	CoMid
(log 1/Cs)	avg	**0.477**	2599	16.7		46.8	0.476	175.9	16.6		47.0
	ind	**0.109**	3026	4.0		51.3	0.097	48.5	3.3		51.1
Su	avg	**0.803**	976		0.700	23.6	0.803	67.2		0.705	23.4
	ind	**0.183**	31463		0.375	36.0	0.172	47.9		0.334	37.5
Su+log 1/Cs	avg	**0.836**	816	5.5	0.592	25.5	0.836	55.8	5.6	0.595	25.3
	ind	**0.187**	27184	2.8	0.224	41.4	0.172	43.7	2.4	0.229	41.7
Models of **surprise**											
Log Bs (avg)	avg	**0.119**	7179		9.2	41.7	0.119	586.1		10.4	40.1
Log Ds (avg)	ind	**0.671**	2677		30.0	45.0	0.670	221.7		30.6	44.6

Table B.4. Experiment 4 modeling results. Models of coincidentality: log inverse coincidence strength model fit summary for averaged data, first row, and the average of individual participant fits, second row, based on sum-squared error (SSE), left columns, and weighted sum-squared error (wSSE), right columns. Log Bayesian surprise (log Bs) is in rows three and four. Log Data surprise (log Ds) is in rows five and six. The combined model of log Bayesian surprise and log inverse coincidence strength (log Bs/Cs), rows seven and eight. The combined model of log data surprise and log inverse coincidence strength (Equation 7), rows nine and ten.

		SSE fits					wSSE fits				
		%var	SSE	CsScale	SuScale	CoMid	%var	wSSE	CsScl	SuScl	CoMid
(log 1/Cs)	avg	**0.588**	3676	15.3		42.3	0.580	149.4	17.0		41.5
	ind	**0.201**	31533	5.8		49.6	0.187	57.5	4.9		50.2
Log Bs	avg	**0.404**	5316		17.3	36.1	0.401	238.4		18.3	35.3
	ind	**0.157**	34818		3.3	49.7	0.142	61.9		3.0	49.9
Log Ds	avg	**0.722**	2479		26.1	43.3	0.719	109.3		27.5	43.0
	ind	**0.306**	27873		10.2	48.5	0.275	53.1		8.5	48.9
Log Bs/Cs	avg	**0.588**	3672	14.7	0.95	41.8	0.579	149.4	17.2	-0.4	41.7
	ind	**0.231**	30638	4.0	1.3	48.8	0.217	55.9	3.2	1.3	49.3
Log Ds/Cs	avg	**0.748**	2248	5.2	19.8	42.0	0.742	95.7	6.6	19.6	41.6
	ind	**0.352**	25374	2.7	8.5	46.1	0.337	47.8	2.5	7.3	47.1

Figure Captions

Figure 1. Coincidence strength, the posteriors belief ratio for the causal and chance hypotheses, in Bayes' space. The x-axis is the log of the data likelihoods ratio for the causal and chance hypotheses respectively. The y-axis is the log of the priors ratio of beliefs in the causal and chance hypotheses respectively. The blue arrow indicates increasing coincidence strength. See main text for details. Figure adapted partly from Figure 2 in Griffiths and Tenenbaum (2007).

Figure 2. Average coincidentality versus the ratio of the average causal likeliness (Ca|d) to the average chance likeliness (Ch|d), top panel, and combined plots of all nonreversed participants' judgments, bottom panels, from Experiment 1. Error bars are standard error. Note that the left bottom panel is a magnification of the right bottom panel for posterior ratios between 0 and 5. The trend line is logarithmic. Red circles are the blond-guy coincidence and blue squares are the weekend-alarm coincidence from Table 1. A small amount of noise jitter has been added to the data points in this and subsequent combined data panels to more clearly convey data density.

Figure 3. Coincidentality versus each of the four probability judgments--the causal and chance posteriors and the causal and chance likelihoods--for averaged judgments from Experiment 2. Error bars are standard error. Trendlines are linear with percentage of variance accounted for. Red circles are the blond-guy coincidence and blue squares are the weekend-alarm coincidence from Table 1.

Figure 4. Average coincidentality color coded in Bayes' space (Figure 1) composed of the log of the priors ratios of the averages, y-axis, and the log of the likelihoods ratios of the averages, the x-axis of the central panel, for data from Experiment 2. The top panel shows the color coding of coincidentality, y-axis, in relation to the log likelihoods ratio, the x-axis, red is most coincidental and blue is least. And the right panel shows a similar coding of coincidentality for the priors ratio. Large Red circles are the blond-guy coincidence, and large blue squares are the weekend-alarm coincidence from Table 1.

Figure 5. Bayesian surprise as the extent of the disagreement between the prior belief and data likelihoods ratios the y and x axes respectively in the Bayes' space from Figure 1 in the Introduction.

Figure 6. Average coincidentality from Experiment 2 color coded in rotated Bayes' space, with the log of Bayesian surprise on the x-axis and log coincidence strength on the y-axis of the central panel. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to log Bayesian surprise on the x-axis in the top panel and in relation to the log posteriors ratio on the y-axis in the right panel. Large Red circles are the blond-guy coincidence, and large blue squares are the weekend-alarm coincidence from Table 1.

Figure 7. Coincidentality in a space composed of surprise, central panel x-axis, and the log of the posteriors ratio, y-axis, for averaged data from Experiment 3, color coded by coincidentality. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to surprise on the x-axis in the top panel and in relation to the log of coincidence strength (the posterior hypothesis belief ratio) on the y-axis in the right panel.

Figure 8. Coincidentality space, central panel, composed of log coincidence strength on the y-axis and log data surprise on the x-axis for averaged data from Experiment 3. The color coding of coincidentality, blue for low coincidentality and red for high, is shown in relation to log data surprise on the x-axis in the top panel and in relation to the log of coincidence strength (the posterior hypothesis belief ratio) on the y-axis in the right panel.

Figure 9. Coincidentality space, central panels (e.g. bottom left panel), with color coded coincidentality (blue = low coincidentality, red = high), log data surprise on the x-axis and log coincidence strength on the y-axis. The color coding of coincidentality is shown versus data surprise, x-axis in the adjoining top left panels and versus coincidence strength, y-axis, in the adjoining bottom right panels. Left panels are averaged data and right are combined (nonreversed) individual participant data from Experiment 4. Combined data have a small amount of noise jitter to show data density.

Figure 10. Summary of key relationships found between judged coincidentality and its predictors. See main text for details.

Figure 11. Causal discovery in coincidentality space composed of (log) coincidence strength, y-axis, and (log) data surprise, x-axis, as in Figure 9. The large gray arrow corresponds to coincidentality increasing toward the bottom right-hand side of the space, as per equation 7. Details in main text.

Figure A1. Correlation matrices with correlations between individual participants for each of the main judgment tasks (rows of panels) in each experiment (columns of panels). Each row (and each column) in a given matrix corresponds to the correlations between a given participant’s judgments in a task for all 97 coincidences individually to the judgments from each other participant. Black cells correspond to a negative correlation coefficients (r) less than or equal to (i.e. more negative than) -0.1, and white cells to correlations bigger than this.