Conceptualizing Surprise with the Jensen–Shannon distance*

A Bayesian Information-Theoretic approach for the Social Sciences

Emil Niclas Meyer-Hansen†

7 February 2026

Abstract

Generally in the social sciences, results are informally deemed *surprising* if their associated p-value is sufficiently small. This implicit interpretation, however, is both conceptually and mathematically inappropriate, and such misuse of the p-value can lead to erroneous conclusions about the novelty of results. To solve that issue, this paper builds on Bayesian inference, Information theory, and considerations specific to the social sciences, to argue for the adoption of a more appropriate conceptualization of surprise as the *relative entropy between prior and posterior knowledge*. Novel to the social sciences, this formal conceptualization enables researchers to appropriately measure surprise as the Jensen-Shannon distance, for which the paper contributes with easily implementable software and a demonstration of its use in relation to empirical data.©[1]

Keywords: *Surprise; Novelty; Jensen-Shannon distance; JS distance; Jensen-Shannon divergence; JS divergence; Relative entropy; Kullback-Leibler divergence; KL divergence; Distance; Divergence; Dissimilarity; Differential entropy; Entropy; p-value; S-value; Information theory; Bayesian inference*

***Version**: 2026-02-07-15-20. **Citation**: Meyer-Hansen, E. N. (2026): 'Conceptualizing Surprise with the Jensen–Shannon distance: A Bayesian Information-Theoretic Approach for the Social Sciences', *Open Science Framework*, Working paper (v2026-02-07-15-20). DOI: 10.17605/OSF.IO/GQ6C8

†**Author**: Independent Researcher, Aarhus, Denmark. Holds a Master of Science (MSc.) and Bachelor of Science (BSc.) in Political Science from Aarhus University, Denmark (email: emil098meyerhansen@gmail.com).

1 Introduction

Assessments of how surprising the results of a study are is an important yet theoretically vague concept within the social sciences. Generally, for a research paper to be deemed worthy of publication in a peer-reviewed scientific journal, besides adhering to standards on rigor and replicability (Christensen et al., 2019; Shaughnessey et al., 2015; Toshkov, 2016), its results should ideally be *surprising* (or novel) when compared to expectations based on intuition or theory (French and Murphy, 2023; Morgan, 2005; Wang et al., 2024). This preference for surprise/novelty could be attributed to the belief that scientific studies should serve to produce knowledge relevant for solving societal problems (Buunk and van Vugt, 2013; Lugg, 1979), which implies a limitation of current problem-solving capabilities and thus the need for a study to acquire *new* and improved knowledge. Similarly, from the perspective of editors, if the goal is to increase the readership of a journal, novelty may be preferred, since surprising results are more likely to attract attention and engage readers (cf. Itti and Baldi, 2009). To thus avoid their research collecting dust in the file-drawer (cf. Rosenthal, 1979), the surprise of results becomes an important factor for researchers to assess and emphasize in order to increase its perceived relevance and chances of publication.

Recognizing this preference, the scientific community has tried to standardize and quantify novelty to make it less subjective (e.g., Modirshanechi et al., 2021; Wang et al., 2024), but despite their best efforts, there is currently no broad consensus, leaving different fields to rely on different solutions (Wang et al., 2024). Within the social sciences, this has lead to the development of informal practices for assessing novelty, with particularly the (in)famous p-value currently serving as a popular heuristic. Here, the the p-value is effectively treated as a measure of surprise through the informal assessment of checking whether it is smaller than some threshold (e.g., $p \leq 0.05$), and if it is, the results are implicitly considered to be 'surprising'.

However, using the p-value as a measure of novelty can be shown to be inappropriate given its conceptualization and mathematical properties. Specifically, it can be made evident that almost any result becomes 'surprising' given a sufficiently large sample size, irrespective of how derivative the results actually are compared to the expectations of intuition or existing theory. Misleading conclusions about the novelty of results are thus abound to incur from this informal practice, and it therefore constitutes an issue for the social sciences that may lead to incorrect decisions in relation to assessments of relevance and publication.

To solve this issue, this paper argues for the need to adopt a formal (re)conceptualization of surprise within the social sciences. Building on Bayesian inference (Gelman et al., 2014) and Information theory (Cover and Thomas, 2006), this involves synonymously conceptualizing *knowledge* and *information* as *(inter-)subjective beliefs*, where the expected results given existing theory can be characterized as *prior* knowledge, while the actual results given the data can be characterized as *posterior* knowledge. Consistent with Kullback and Leibler (1951), *relative entropy* can then be used to describe the divergence in knowledge. This enables a theoretically-coherent, formal conceptualization of surprise as *the relative entropy between prior and posterior knowledge*. As demonstrated in relation to empirical data, the degree of surprise can then be appropriately measured based on the *Jensen-Shannon distance* (Lin, 1991), which possess numerous advantages suited for the social sciences. By arguing for this conceptualization, the paper thus adds to the growing literature that calls for formally conceptualizing surprise in accordance with Bayesian inference and Information theory (e.g., Bencomo and Belaggoun, 2014; Burkhart et al., 2025; Burnham and Anderson, 2001; Itti and Baldi, 2009; Mello et al., 2025; Seehars et al., 2014).

For the purposes of achieving this solution, the paper is structured as the following: Following this introduction (Section 1), the paper considers the current problem within the social sciences by reviewing the informal conceptualization of surprise and the p-value to argue for the latter's (mis)use as an implicit indicator of surprise within the social science literature (Section 2). The paper then builds towards a solution by presenting concepts central to Bayesian inference and Information theory, incorporating them into a formal conceptualization of surprise as the relative entropy between prior and posterior knowledge, measurable as the Jensen-Shannon distance (Section 3). A practical implementation of the Jensen-Shannon distance for the popular R programming language (R Core Team, 2025) is presented, before a demonstration of its ease of use in relation to empirical data is provided (Section 4). The uses and limitations of this solution are discussed (Section 5) before the paper is concluded (Section 6).

2 Problem

To build towards a formal conceptualization of surprise as the relative entropy between prior and posterior knowledge, this section first covers the current, informal definition of surprise within the social sciences (Section 2.1), before reviewing the conceptualization and mathematical properties of the p-value that make it an inappropriate measure of surprise (Section 2.2). Consistent with related literature (e.g., Itti and Baldi, 2009), novelty and surprise are treated as synomous throughout the paper.

2.1 Surprise, informally

In layman's terms, *surprise* can generally be defined as a reaction 'caused by something unexpected happening' (Cambridge Dictionary, 2025). While this definition is is consistent with researchers' application of the term (cf. Ivanova and Vaidyanathan, 2024; Itti and Baldi, 2009), including within the social sciences (e.g., Alford et al., 2011: 362; Goren and Chapp, 2017: 124; Petersen and Laustsen, 2019: 26), it can be argued to be too vague to constitute a formal scientific conceptualization. However, usage of this term by researchers can nevertheless be analyzed as a starting point to understand why the p-value is misused as a measure of surprise and what a proper conceptualization for the social sciences would entail.

That surprise is a reaction to the unexpected can within the social sciences be understood to reflect a *discrepancy between results and predictions made about those results based in theory*. Here, predictions are generally formalized as *hypotheses* about a parameter of interest (Shaughnessey et al., 2015; Toshkov, 2016), whose theoretical basis may range from weak (e.g., intuition-based) to strong (e.g., well-established theory). That such discrepancy is (informally) recognized as being synonymous with surprise is made evident by statistically discernible differences between expectations and results almost ubiquitously being referred to as surprising and worthy of discussion. For example, Cochrane et al. (1979) studied the political values of individuals in British neighborhoods with door-to-door interviews. Based on theory by Rokeach (1973), they expected individuals with different political ideologies (e.g., liberals, communists, fascists) to differ in their prioritization of freedom. Instead, they found that individuals across all ideologies prioritized freedom similarly, and this unexpected discrepancy between the predictions by Rokeach (1973) and the results by Cochrane et al. (1979) were discussed and later explicitly referred to as 'surprising' by other researchers (e.g., Sterling et al., 2019: 2). This suggests that surprise can informally be characterized as the discrepancy between results and theory, and a formal conceptualization would thus need to reflect this property.

Another related characteristic of surprise that may be recognized from its informal usage is that it is entirely *relative* to the predictions of the theory to which results are being compared. This is made evident by researchers recognizing that results may be surprising when evaluated against one theory but also entirely unsurprising when compared to another. For example, Osmundsen et al. (2021) investigated the behavior of American individuals on the social media *X* (then known as *Twitter*). In contrast to the expectations of *ignorance theory*, where fake news sharing would expectedly be attributable to individuals being unable to discern between true and fake information (Osmundsen et al., 2021: 1001; based on Altay et al., 2020; Pennycook and Rand, 2019; Pennycook et al., 2020), the authors instead find it better explained as based in hostility toward political opponents, in line with the *negative partisanship* literature (Osmundsen et al., 2021: 1012; see Abramowitz and Webster, 2018). While this result could thus be considered surprising when compared to one theory, it was entirely unsurprising when compared to another. This points to an appropriate conceptualization of surprise needing to account for it being *relative* to the different theories that results are being compared against.

A final characteristic of surprise gained from analyzing its popular use is that it is *dynamic*, evolving as science progresses. This is made evident by researchers using surprising results to argue for revisions (if not discardment) of theory, with the aim being to bring expectations in greater alignment with results and thereby reduce future surprise. For example, in the discussion of the previously noted study by Osmundsen et al. (2021), they use their results to argue for revising future expectations about fake news sharing, stating that 'If people care primarily about a story's ability to hurt political enemies, we should *not* be *surprised* if fact checking fails to reduce sharing.' (Osmundsen et al., 2021: 1013, emphasis added) Such revisions fit the self-correcting ideal of science, whereby theories are revised or discarded in favor of theories with a higher explanatory power (i.e., *verisimilitude*, Miller, 1974; Popper, 1959; see also Dellsén, 2024; Lakatos, 1970). In the event that similar results were to appear in future studies, they should accordingly incur *less* surprise.

That example also corroborates the notion that, whether phrased in terms of 'weak' explanatory power, 'bad' fit, or a similar term, discrepancy between results and predictions are indicative of the *epistemic limitations* of the theory underlying those predictions[2] (French and Murphy, 2023: 1448-1449). The better the predictions of a theory, the less surprising the results should appear, and an appropriate conceptualization of surprise should thus enable researchers to motivate revisions or discardment of theory from its assessment.

Based on these initial considerations, it is apparent that surprise lacks a formal conceptualization within the social sciences, leaving it only vaguely defined in this field. Nonetheless, based on an analysis of its common usage, it has become clear that it can be characterized as the discrepancy between results and predictions derived from theory, with surprise being relative to, and co-evolving with, theory. Higher degrees of surprise are indicative of the limitations of a theory, which may be used to motivate its revision or discardment. These insights are important for any formal conceptualization of surprise within the social sciences, since a proper measure of surprise will need to account for these characteristics to be optimally suitable for the researches of that discipline. While some may argue that a vague conceptualization of surprise is sufficient for the social sciences, the pitfalls of this informality is demonstrated in the following section by considering the misuse of the p-value as an implicit measure of surprise.

[2]Indeed, at least in a deterministic world, a researcher making predictions using a theory with perfect explanatory power will never by surprised by the results (Itti and Baldi, 2009). As a note of reference, it may thus be stated that Laplace's (1951) *demon* should be incapable of being surprised (for arguments against a deterministic world, see, e.g., Masi, 2023).

2.2 *p*-value

In the classical statistical framework, also known as the *Frequentist* approach (for an introductory text, see, e.g., Agresti, 2018), the *probability value* (i.e., p-value) denotes *the probability of at least as extreme an estimated parameter value from the sample given that the null-hypothesis about the population parameter is true* (Clayton, 2021: 291). To understand why it is an inappropriate measure of surprise, some conceptual understanding of this dense definition is required, which can be provided in brief.

The *p*-value is conceptually part of the *Null-Hypothesis Significance Testing* procedure (NHST; Cohen, 1994; Field, 2018: 72-82; Kruschke, 2014: 297-334), which involves the researcher formulating a *null-hypothesis* (i.e., H_0) about a *parameter in a population of interest* (i.e., θ), which typically takes the form of a *nil hypothesis* by stating that the parameter is *exactly* zero (i.e., $H_0 : \theta = 0$; Cohen, 1994: 999-1000). This null-hypothesis contrasts an *alternative hypothesis* (i.e., H_a) about the population parameter that the researcher has derived from theory. By convention, however, H_a takes the form of a complementary non-nil hypothesis (e.g., $H_a : \theta \neq 0$), which seldom reflects any substantive prediction of said theory (cf. Meehl, 1967).

In line with *hypothetico-deductive reasoning* (Popper, 1959), the purpose of NHST is then to build confidence for H_a through attempts at falsifying H_0. This is accomplished by collecting a prespecified amount of data from a larger population of interest, ideally in the form of a sufficiently large and representative sample, before calculating the parameter from this data using a mathematical formula known as an *estimator* (Stock and Watson, 2019: 105; Wooldridge, 2019: 715-721), which provides an estimate of the parameter in the population based on the collected data. The estimated parameter from that sample (i.e., $\hat{\theta}_s$) is then compared to the *sampling distribution* of its estimator (i.e., $\hat{\theta}$; Field, 2018: 61-64), which is a distribution of hypothetical estimates derived from an infinite number of alternative, random samples taken from that population if H_0 is true. Since Frequentists define probability as *the long-run relative frequency* (Agresti, 2018: 79), the p-value is calculated as the area under the curve in this sampling distribution of $\hat{\theta}$ that is at least as extreme as $\hat{\theta}_s$. If the p-value is lower than some threshold (i.e., α) specified before analyzing the data (e.g., $\alpha = 0.05$, with $p \leq \alpha$) then H_0 can be rejected. This would be consistent with H_a being true, and since the estimated parameter can account for *sampling uncertainty*, it can then be referred to as being 'statistically discernible'[3] from the parameter implied by the null-hypothesis. By contrast, if the p-value is higher than α, then H_0 would fail to be rejected[4]. This is consistent with H_0 being true, and since the estimated parameter cannot account for sampling uncertainty, it is referred to as being 'statistically indiscernible' from the parameter implied by the null-hypothesis.

From this conceptualization of the *p*-value, it is evident that the intended purpose of this statistic is that of *statistical inference*, that is, inferring results from sample to population (Agresti, 2018: 16-17). Nowhere in its conceptualization is a consideration for expressing the degree of discrepancy between results and predictions based in theory (i.e., either a small or large *p*-value could be consistent with any degree of actual surprise). While the *p*-value is relative to a prediction based in the null-hypothesis, the null-hypothesis generally does *not* reflect serious predictions based in theory (i.e., what social theory would be considered substantive if its sole prediction involved

[3] Instead of the term 'statistically discernible', researchers often use the less appropriate, but more popular term 'statistically significant', or just 'significant'. This constitutes an instance of *semantic overload* (cf. Bolinger, 1971), which this wording specifically alluding to the estimate being of any substantive size, which is fundamentally incorrect and borders on the deceptive (See below in this section for a mathematical demonstration).

[4] Based on the logic of NHST, researchers should note that H_0 can never be 'accepted', it can only ever be rejected or fail to be rejected (Kruschke, 2014: 305).

a parameter-value being non-zero?), and it fails to evolve with a theory by generally remaining a nil hypothesis no matter the context, making the *p*-value too static to reflect the dynamics of surprise.

These issues can be further recognized with a simple example. Suppose one conducts a *randomized controlled trial* (RCT; Deaton and Cartwright, 2018) on the *entire* population of interest, where half of individuals are exposed either to a placebo condition or some treatment, which given one theory is expected to change these individuals, while another theory expects it *not* to change anything. Even if the treatment is found to induce a change or not, the *p*-value is entirely irrelevant to the purposes of this experiment, because there is no need for statistical inference (because the 'sample' is the entire population).[5] Since an assessment of the surprise the results remains relevant in spite of the irrelevance of the *p*-value, this means that an assessment of surprise must be separate from it.

These conceptual limitations can be further realized by considering the mathematical formula underlying the *p*-value. While it can be calculated in numerous ways, the most widely used version of the p-value involves an assumption that the sampling distribution of the estimator used to calculate the parameter of interest ($\hat{\theta}$) is *Gaussian* (i.e., Normal; Wooldridge, 2019: 723-724), which under a nil hypothesis, means that the distribution can be characterized with the following zero-mean *Normal probability distribution* ($\mathcal{N}$; Agresti, 2018: 84-91):

$$\hat{\theta} \sim \mathcal{N}(0, \sigma^2_{\hat{\theta}}) = \frac{1}{\sqrt{2\pi\sigma^2_{\hat{\theta}}}} e^{-\frac{x^2}{2\sigma^2_{\hat{\theta}}}} \tag{1}$$

where $\sigma_{\hat{\theta}}$ refers to the dispersion, or *standard deviation* (SD), of this sampling distribution of $\hat{\theta}$ (i.e., the *standard error*; Field, 2018: 61-64); $\pi = 3.141...$ is the ratio between the circumference and diameter of a circle; and $e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.718...$ is the base of the natural exponential function.

If $H_0 : \theta = 0$ and $H_a : \theta \neq 0$, the direction by which $\hat{\theta}_s$ deviates from 0 is irrelevant, and the p-value is accordingly referred to as *two-tailed* (Field, 2018: 79-81), which given the symmetry of the distribution in (1) is given by the definite integral (cf. Stock and Watson, 2019: 111):

$$p = 2 \int_{|\hat{\theta}_s|}^{\infty} \mathcal{N}(0, \sigma_{\hat{\theta}}) \, \mathrm{d}x \tag{2}$$

where $\int$ denotes the integration operator and $|\hat{\theta}_s|$ is the absolute value of $\hat{\theta}_s$.

Since an application of formula (2) requires calculating $\sigma_{\hat{\theta}}$ using infinitely many samples from a population in the possibly hypothetical scenario where H_0 is true, $\sigma_{\hat{\theta}}$ generally has to be estimated. For example, if the parameter of interest is the population mean (i.e., $\theta = \mu$) with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ being its estimator (Field, 2018: 27), where x_i denotes the ith value from an n-length vector constituting a random sample of independent and identically distributed (iid) values drawn from a Normally distributed population of interest; and the population variance is estimated as $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$ (Field, 2018: 30), then the estimator for the dispersion of the sampling distribution for the estimated mean is (Stock and Watson, 2019: 113, equation 3.9): $\hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}}$.

Now, since $\hat{\sigma}^2$ is an *unbiased* and *consistent* estimator of σ^2 (Stock and Watson, 2019: 112; Wooldridge, 2019: 716-723, 766), $\hat{\sigma}^2$ and σ^2 converge as the sample size (n) increases towards infinity, while $\sqrt{n}$ increases

[5] While one may argue that calculating the p-value in this example could help express uncertainty, it should be emphasized that the only uncertainty that can be expressed by the p-value is that of *sampling uncertainty*, which is only relevant for statistical inference. Since the data constitutes the entire population of interest in the example, there is no sampling involved, meaning that for all practical purposes (See below in this section), the p-value is *a priori* zero, and therefore an irrelevant concern.

without bound. A mathematical analysis of $\hat{\sigma}_{\hat{\mu}}$ thus makes it evident that the estimated dispersion of the sampling distribution approaches zero as n increases towards infinity:

$$\lim_{n \to \infty} \frac{\hat{\sigma}}{\sqrt{n}} = 0 \tag{3}$$

The implications of equation (3) pose a detrimental issue for the ability of the p-value to reflect surprise. This can be made more clear by first converting the estimated parameter into a standardized format called a *test statistic* (Field, 2018: 78-79). In relation to an assumption of a Normal sampling distribution, this test statistic is referred to as a z-value, which involves the rescaling formula (Field, 2018: 37): $z = \frac{\hat{\mu} - \mu_{H_0}}{\hat{\sigma}_{\hat{\mu}}}$, where μ_{H_0} denotes the mean under the null-hypothesis. If the difference between the population parameter and the population parameter under the null-hypothesis is non-zero (i.e., $\mu - \mu_{H_0} \neq 0$), this z-value can be recognized to approach infinity as $\hat{\sigma}_{\hat{\mu}}$ decreases towards zero, since $\hat{\mu} - \mu_{H_0}$ converges to $\mu - \mu_{H_0}$ as n increases towards infinity due to $\hat{\mu}$ being unbiased and consistent, while μ_{H_0} is a constant (Stock and Watson, 2019: 104-107; Wooldridge, 2019: 716-723). Similar to (3), this is expressed as:

$$\lim_{\hat{\sigma}_{\hat{\mu}} \to 0} \frac{\hat{\mu} - \mu_{H_0}}{\hat{\sigma}_{\hat{\mu}}} = \infty \tag{4}$$

which has been established to occur in (3) when $n \to \infty$. Due to the z-standardization, the Normal distribution in (2) has become a standard Normal, and substituting $\hat{\theta}_s$ for z, it can finally be realized that the limit of the p-value as z increases towards infinity is zero:

$$\lim_{z \to \infty} 2 \int_{|z|}^{\infty} \mathcal{N}(0,1)\, \mathrm{d}x = 0 \tag{5}$$

as demonstrated to occur in (4) when $\hat{\sigma}_{\hat{\mu}} \to 0$, which happens when $\mu - \mu_{H_0} \neq 0$ and $n \to \infty$ as shown in (3).

The implication of these limits are critical. For any instance where the null-hypothesis is not *exactly* true — which likely applies to most of the phenomena studied in the social sciences[6] — anything becomes 'statistically discernible' with a sufficiently large sample size. In the event that one equates a low p-value with surprise, this unfortunately means that *almost anything becomes surprising with a large enough sample size*, irrespective of how expected the results actually are under the given theory.[7] Using the p-value as a measure of surprise is thus conceptually flawed and runs counter to the characteristics of actual surprise recognized by social science researchers as previously identified in section (2.1).

While one may argue that the extent of this informal practice of using the p-value as a measure of surprise is too negligible for consideration, it may be counter-argued that such misuse is best addressed as early as possible, so as to prevent it from becoming a formal practice. Such a development may seem improbable, but this happened recently within the field of epidemiology, where Cole et al. (2021) introduced the 'S-value' as a formal measure of surprise. Readers of this paper, however, could quickly realize that this 'S'-value unfurtunately had a one-

[6] That the null-hypothesis is generally false within the social sciences can be realized by considering that hypotheses generally concern point values (e.g., $H_0 : \theta = 0$) despite most studied phenomena arguably being continuous, and the assumptions for the data-generating processes of stochastic continuous phenomena involve distributions where a point hypothesis has an *a priori* zero probability of being true (cf. Cohen, 1994: 1000; Loftus, 1991, 2001; Steiger and Fouladi, 1997; Thompson, 1996; Tukey, 1991: 100; Vicente and Torenvliet, 2000: 260). In psychology, the notion that the nil hypothesis is always false is referred to as Meehl's (1978) conjecture (Gigerenzer, 2004: 601).

[7] The issue cannot be addressed by specifying the null-hypothesis to fit the point prediction of a given theory, since the most negligible difference between the result and the prediction can produce a vanishingly small p-value given a sufficiently large sample.

to-one correspondence with the p-value through the formula (Cole et al., 2021: 192): $S \equiv -\log_2(p)$. Similar to criticisms made against other measures derived solely from the p-value[8], this means that the S-value fails to incorporate, nor communicate, any new information besides that already expressed by the p-value. The S-value is thus as susceptible to the exact same issues identified here for the p-value in its ability to express surprise[9], making it just as *inappropriate*.

Now, while this paper serves to solve this misue of the p-value, the aim is *not* to portray researchers negatively for partaking in this inappropriate practice. Indeed, the p-value is infamous for being particularly susceptible to misinterpretation (Clayton, 2021: 291; Gigerenzer, 2018: 206), with examples including it supposedly being an index of scientific significance, reliability, replicability, effectsize, and the probability of H_0, H_a, or data having been generated by chance, resulting in various criticisms and cautions against its (mis)use (American Statistical Association, 2016; Clayton, 2021; Field, 2018: 97-110; Gelman and Stern, 2006; Kruschke, 2014: 297-329; McShane et al., 2019). Instead, as similarly achieved by other researchers identifying questionable practices within the scientific community (e.g., Bishop, 2006; Flake and Fried, 2020; Gelman, 2024; Ioannidis et al., 2017; Lenz and Sahn, 2021; Ritchie, 2020; Simmons et al., 2011), the aim of this paper is to improve the social scientific practice by informing researchers of this inappropriate practice and to propose a possible solution.

3 Solution

While the previous section sought to describe the current, informal conceptualization of surprise and explain why the p-value is an inappropriate measure of surprise, whose use constitutes an issue for the social sciences, this section serves to provide a solution to that problem with a formal (re)conceptualization of surprise. This solution is partly founded in similar (re)conceptualizations proposed for other disciplines (e.g., Bencomo and Belaggoun, 2014; Burkhart et al., 2025; Burnham and Anderson, 2001; Itti and Baldi, 2009; Mello et al., 2025; Seehars et al., 2014). To that end, central concepts of Bayesian inference (Section 3.1) and Information theory (Section 3.2) are presented and used to build that conceptualization and identify an appropriate measure for it.

3.1 Bayesian inference

The *Bayesian statistical framework*, also known as Bayesian inference, is an approach to statistical inference that differs from its Frequentist counterpart by defining probability in relation to (inter-)subjective beliefs (for an introduction, see, e.g., McElreath, 2019). Understanding this framework presupposes some familiarity with *Bayes' theorem* (Bayes and Price, 1763; Laplace, 1951), which can be expressed in relation to the probability of two discrete events as (McElreath, 2019: 49):

[8] Somewhat unsurprisingly, this is *not* the first time that the p-value has been inappropriately repurposed to express phenomena outside its intended scope. As an infamous example, Killeen (2005a) similarly sought to address the fact that the p-value cannot express the probability that the results of a study will replicate in another study (cf. Gigerenzer, 2018). The mathematical formula underlying the proposed solution by Killeen (2005a) was widely criticized (e.g., Iverson et al., 2009; Macdonald, 2005; Maraun and Gabriel, 2010; Serlin, 2010; for the response, see, Killeen, 2005b), since it was evident that the formula had a one-to-one correspondence with the p-value, and it thus failed to incorporate any information about the probability of replication not already communicated by the p-value. Instead, as similarly argued here in relation to surprise, it was argued that a (re)conceptualization of replicability was better founded in Bayesian inference (e.g., Macdonald, 2005).

[9] This can be made evident by considering that $\lim_{p \to 0} -\log_2(p) = \infty$, which was established by equations (3) through (5) to occur as $n \to \infty$, which means that, based on the logic of the S-value, almost anything will inevitably become surprising with a sufficiently large sample size, no matter how expected the results were in advance. Despite these issues, other researchers have worryingly advocated for the adoption of the S-value (e.g., Greenland et al., 2022; Rafi and Greenland, 2020), further revealing the need for a *proper* formal (re)conceptualization of surprise.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{6}$$

where $P(A)$ denotes the probability of the event A, which is referred to as the *prior* probability; $P(B)$ denotes the probability of event B, sometimes called the *marginal likelihood*; $P(B|A)$ is the probability of event B given that event A has occurred, denoted the *likelihood*; and $P(A|B)$ is the probability of event A given that event B has occurred, which is referred to as the *posterior* probability.

The components of (6) are generally reconceptualized for the purposes of statistical inference (Gelman et al., 2014; Kruschke, 2014; Levy and Mislevy, 2020) so that the parameter of interest (θ) substitutes A, while the collected data ($\mathcal{D}$) substitutes B. This means that $P(\theta)$ denotes the (prior) probability of θ, $P(\mathcal{D}|\theta)$ the probability of the data given the parameter, and $P(\theta|\mathcal{D})$ the (posterior) probability of θ given $\mathcal{D}$, with this latter component being of primary interest in Bayesian inference.

For practical purposes $P(\mathcal{D})$ can often be ignored due to relying on *Markov-Chain Monte Carlo methods* (MCMC; Brooks et al., 2011), which leaves $P(\theta)$ and $P(\mathcal{D}|\theta)$ relevant for calculating $P(\theta|\mathcal{D})$. While those familiar with *likelihood-based inference* (see, e.g., King, 1998) will know that $P(\mathcal{D}|\theta)$ is often easy to derive from the statistical model used in the analysis, $P(\theta)$ is by comparison rarely known. To solve that issue, this prior is specified as a probability distribution using *a priori* knowledge about θ. While the discretion involved in specifying this prior distribution makes Bayesian inference (inter-)subjective, results generally approximate Frequentist results with a sufficiently large sample size (Cover and Thomas, 2006: 388; Hastie et al., 2017: 272; King, 1998: 30; Kruschke, 2014: 113), and it can be considered a generalization of the arguably less subjective[10] likelihood-based approach (cf. King, 1998: 28). The difference between the two approaches is thus primarily in the definition, and thus interpretation, of probability.

Unlike other approaches, this conceptualization Bayesian inference provides a coherent statistical framework for characterizing uncertainty (cf. Cox, 1946; Gelman et al., 2014; Jaynes, 2003; Savage, 2017) and enables one to answer questions about the probability of parameters given the data, $P(\theta|\mathcal{D})$. This is impossible in the Frequentist framework, since that approach solely relies on the likelihood and fails to consider prior information in relation to θ (Kruschke, 2014; Wagenmakers et al., 2010). It also means that results from a Bayesian analysis are in the form of probability distributions, which conceptually express how credible different estimated values of θ are given the data. While these distributions could be summarized with point estimates (e.g., mean, dispersion) or intervals (e.g., 95% credibility interval) for practical purposes, as done in the Frequentist approach, this is theoretically unnecessary, since the posterior distribution comprehensively details one's knowledge about θ based on the prior knowledge and data (McElreath, 2019: 58). Though Bayesian inference can involve p-values and hypothesis tests (see, e.g., Makowski et al., 2019a; Wagenmakers et al., 2010), such procedures related to NHST are *not* native to this framework, which instead of the hypothetico-deductive approach, is more naturally based in *abductive reasoning* (see, e.g., Dellsén, 2024).

Now, consistent with work in other disciplines (e.g., Itti and Baldi, 2009), the conceptualization of Bayes theorem used for statistical inference can be recognized to *naturally* and *coherently* capture the informally recog-

[10]Since all human processing of information and knowledge is necessarily (inter-)subjectively dependent (cf. Husserl, 1983; see also Kant, 1781), the Frequentist statistical framework, including its likelihood-based approach, is *not* 'objective' (i.e., independent of (inter-)subjective experience), despite claims to the contrary (e.g., Mayo, 1996: 89).

nized characteristics of surprise identified in section (2.1). With surprise being the discrepancy between results and expectations/predictions based in theory, it can be equivalently understood as a discrepancy between (inter-)subjective states of knowledge *prior* to seeing the results and the state of knowledge *posterior* to seeing the results. When theory is formalized, data-driven, and well-established among peers within the social sciences, the prior thus characterizes the state of knowledge as a relatively intersubjective evidence-based belief about a phenomenon, describing what the researcher reasonable expects the results to be. As established, since the prior distribution is conceptually a comprehensive statement about this former state of knowledge, it can be appropriately used to characterize the expectations based in theory.

3.1.1 The prior reflects current knowledge

Characterizing the predictions of a theory as state of knowledge expressible as a prior distribution can be difficult, especially if it only makes vague predictions about the parameters relevant to that theory. To therefore help build intuition around them, four examples of different specifications are provided here.

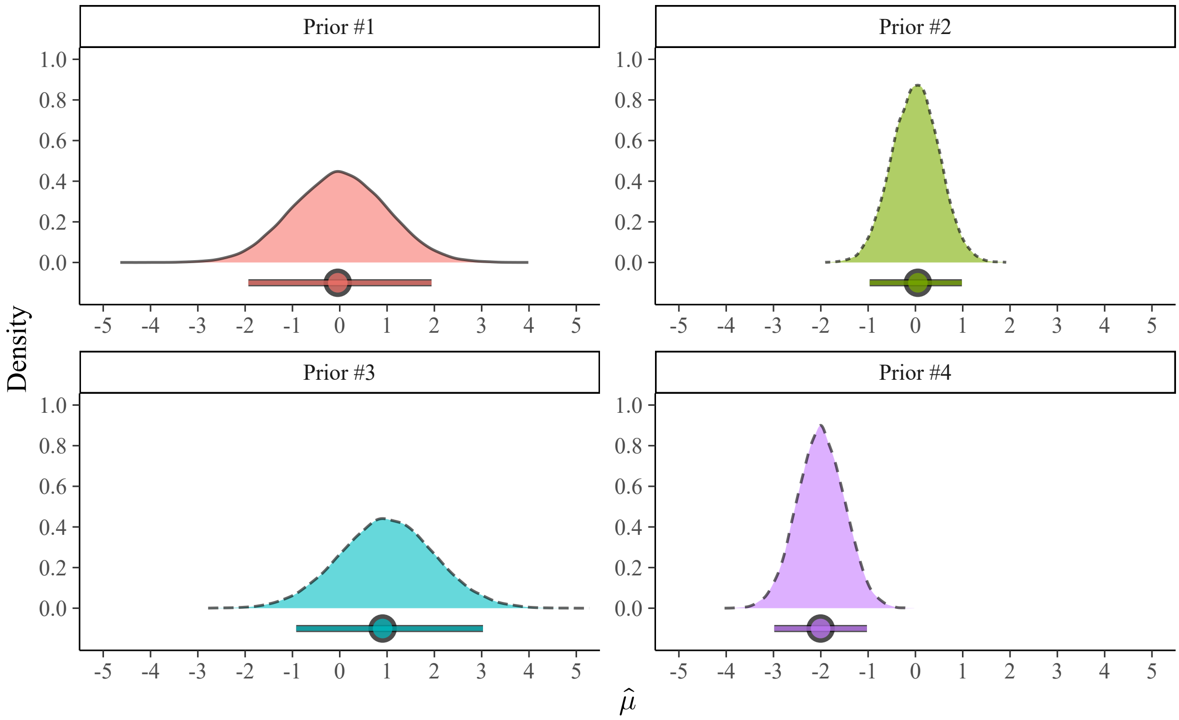

Figure 1. Examples of Prior Distributions

NOTE: Prior #1 $\sim \mathcal{N}(0, 1)$; Prior #2 $\sim \mathcal{N}(0, 0.5)$; Prior #3 $\sim \mathcal{N}(1, 1)$; Prior #4 $\sim \mathcal{N}(-2, 0.5)$. The geometrics below the distributions indicate the mode (circle) and the 95% HDCI (bar). Monte Carlo samples = 40,005.

At no loss of generality, suppose X is a Normally distributed variable in some population of interest: $X_i \sim \mathcal{N}(\mu, \sigma)$ for all i in $1, \ldots, N$, where one wishes to estimate the mean μ using a representative and sufficiently large sample. To enable Bayesian inference, it is necessary to specify its prior, $P(\mu)$, which can be achieved in numerous ways depending on one's prior information about it. As shown visually in figure (1), the first prior (Prior #1) is specified as a Normal distribution: $\mathcal{N}(0, 1)$, which in this context can be taken to reflect a diffuse state of knowledge, where there is complete theoretical ignorance about the sign of the parameter (i.e., whether the mean is positive or negative), but its relatively substantial width (i.e., size of the dispersion) implies that mean could be relatively large irrespective of its sign. This prior may be called *regularizing* (McElreath, 2019: 214-216), or 'skeptical, yet persuadable', which is a type of weakly informative prior that is critical with respect to the sign of a parameter of interest, but which remain wide enough to be influenced by the results.

The second prior (Prior #2) is similar to the first, but by comparison, it is much narrower, which implies a stronger expectation that the mean will not be that large irrespective of the sign. This contrasts the third prior (Prior #3), which has the expectation that the mean is positive, but its similar width to the first prior means it reflects a less certain expectation. The last prior (Prior #4) characterizes an opposite expectation as compared to the third prior, since it expects the mean to be negative, and its greater narrowness reflects a greater degree of certainty about this theoretical belief.

3.1.2 The posterior as updated knowledge

While the state of knowledge prior to the results could be difficult to precisely express as a probability distribution, there is no need for the researcher to specify the state of knowledge posterior to the results, since this resulting probability distribution is a function of the chosen prior distribution and the data. Following Itti and Baldi (2009), the posterior can thus be considered an *updated* state of knowledge, which is useful, since it is informative about how the expectations of that theory should change based on the data to enable better predictions in the future.

Since the posterior is the product of the prior and the data, this conceptualization helps reflect the aforementioned characteristics of surprise as being relative and dynamic in relation to the theories to which results are being compared. Since different prior specifications can reflect the varying predictions made across theories, the posterior could theoretically be compared to each to enable an assessment of how 'surprising' the results are given each theory. For example, let the posterior distribution resulting from an analysis be identical to the fourth example of the priors (Prior #4) previously shown in figure (1), meaning it is characterized as a Normal probability distribution with a mean parameter of -2 and a dispersion of 0.5: $\mathcal{N}(-2, 0.5)$. Assume that these aforementioned priors represent a set of predictions made from different theories, and that the data is large enough to make the posterior independent of these priors.

Comparing the posterior to each prior then makes it clear that the discrepancies between them are *not* equivalent. When evaluated against the the fourth prior (Prior #4), the results fit perfectly with the predictions made *a priori* about the parameter, which would be fully consistent with a complete lack of surprise, at least when based on how social science researchers use the term. While the differences between the posterior mean (i.e., -2) and the mean of priors one and two (i.e., 0) are the same, the results would intuitively appear more 'surprising' when compared against the second prior, since that distribution is narrower and thus expresses greater certainty about the parameter values as implied by its lower dispersion (i.e., 0.5 vs. 1). As such, while the third prior has the highest discrepancy with the posterior in terms of the mean, this suggests that the dispersion of the distribution should also be considered when evaluating the degree of surprise under this Bayesian conceptualization.

Using these concepts from Bayesian inference, surprise has thus far been conceptualized as a discrepancy between states of knowledge, with these states specifically being expressible as a prior and posterior probability distribution. For both the posterior and prior distributions to be valid, they should reflect a genuine state of knowledge about a phenomenon of interest, ideally based on well-established theory and quality data. The advantage of this approach has been demonstrated to arise from the conceptual properties of Bayesian inference that enable comprehensive statements to be made with relative ease about such states of knowledge. A shortcoming, however, is that Bayesian inference can be considered conceptually weak of terms of meaningfully quantifying the degree of discrepancy between states of knowledge, which is why the following section turns to the key concepts of Information theory that will build upon this Bayesian-based conceptualization to enable an appropriate measurement of surprise.

3.2 Information theory

Information theory is a relatively recent mathematical field that is primarily concerned with the efficient communication and compression of information (e.g., Hartley, 1928; Nyquist, 1924, 1928; Shannon, 1948, 1949; Wiener, 1961), but its concepts has proven useful across numerous disciplines, including statistical modeling (e.g., Akaike, 1973; Hastie et al., 2017; Mielniczuk, 2022) and Bayesian inference (e.g., Chaloner and Verdinelli, 1995; Jaynes, 1968; Vehtari et al., 2017; Watanabe, 2010). Using terminology closely related to statistical mechanics (Jaynes, 1957a; 1957b), information is generally conceptualized within this discipline as the 'entropy' of a random variable, which is the key concept in which an appropriate measure of surprise is founded. Since this and related concepts have seen relatively sparse use within the social sciences, they are briefly presented here.

3.2.1 Entropy

In Information theory, *entropy* is 'a measure of the average uncertainty in the random variable' (Cover and Thomas, 2006: 5; Shannon, 1948), making it synonymous with randomness, unpredictability, and disorganization (Berrett et al., 2019: 288; Cover and Thomas, 2006: 135; Jaynes, 1957a: 622; Rényi, 1961: 547). It can be thought of as the information needed, on average, to describe a random variable (Berrett et al., 2019: 288; Cover and Thomas, 2006: 19), which means that the more random the variable, the higher its entropy.

For a finite discrete random variable X, the *Shannon-Boltzmann entropy*, $H(X)$, is defined by the mathematical formula (Back et al., 2018: 1; Cover and Thomas, 2006: 5, equation 1.1; Rényi, 1961: 547):

$$H(X) \equiv \sum_{x \in X} p(x) \log \frac{1}{p(x)} \tag{7}$$

$$= -\sum_{x \in X} p(x) \log p(x) \tag{8}$$

where $\sum$ denotes the summation operator applied over the unique elements of X; $p(x) = P(X = x)$ is the probability mass function for X; and log is the logarithmic function, where it is taken by convention that $0\log 0 \equiv 0$ (Cover and Thomas, 2006: 14). Since the base of the logarithm is generally taken to be 2, this means that the scale of entropy is in *binary units*[11] (i.e., *bits*, which are *not* to be confused with the related concept of binary digits; Cover and Thomas, 2006; Delgado-Bonal and Marshak, 2019; Shannon, 1948). Being measured in bits means that the entropy can be meaningfully interpreted as the expected number of binary questions (i.e., Yes/No questions) needed to know a value of X with certainty (Cover and Thomas, 2006: 16).

With this formula, the entropy can generally be stated to be a non-negative real number, whose upper bound is the logarithm of the number of unique elements in X (i.e., $H \in \mathbb{R}^{[0;\log(|X|)]}$; Cover and Thomas, 2006: 29). For example, a uniform *Bernoulli* ($\mathcal{B}$) variable X (i.e., $X \sim \mathcal{B}(p = 0.5)$) has an entropy of 1 bit. This is an intuitive

[11] Popular alternatives to the base 2 for the logarithmic function are the base of the exponential e and 10, which means that the natural logarithm (i.e., ln) would be used for the former. Instead of bits, this produces entropies on the scale of *natural units* (i.e., *nats*) and *decimal units* (i.e., *dits*), respectively (Back et al., 2018; Cover and Thomas, 2006: 14; Delgado-Bonal and Marshak, 2019).

upper bound for Bernoulli variables, since it reflects the *equiprobability* of the elements, making any element of X maximally unpredictable. This also means that if p either increases or decreases, the entropy decreases, becoming 0 if $p \in \{0; 1\}$. This is similarly an intuitive lower bound, since either value makes X constant, and an entropy of 0 thus reflects that we *a priori* know the values of X and can predict them with absolute certainty (Cover and Thomas, 2006: 14-15). As such, the more uniform a distribution, the higher its entropy (Delgado-Bonal and Marshak, 2019).

While this conceptualization of entropy possess numerous desirable mathematical properties for measuring information, for example, being insensitive to the actual values of X by depending solely upon the their probabilities (Cover and Thomas, 2006: 14), it is limited to *discrete* random variables, which is unfortunate, since most phenomena of interest within the social sciences are arguably *continuous*.[12]

In the event that X instead denotes a continuous random variable, its uncertainty is instead captured by the concept of *differential entropy*, which is 'the entropy of a continuous random variable' and is denoted $h(x)$ (Cover and Thomas, 2006: 243). If the distribution of X can be characterized by the probability density function $f(x)$, its differential entropy is calculated with the formula (Ahmad and Lin, 1975: 373; Beirlant et al., 2001: 1; Berrett et al., 2019: 288; Cover and Thomas, 2006: 243, equation 8.1; Hastie et al., 2017: 561, equation 14.82; Shannon, 1948):

$$h(X) \equiv - \int_S f(x) \log f(x) \, \mathrm{d}x \tag{9}$$

where $\int_S$ denote the integration operator applied over the *support set* of X with respect to x (i.e., where $f(x) > 0$).[13] Unlike entropy, the differential entropy can be negative (Cover and Thomas, 2006: 244), generally making it a real-valued scalar (i.e., $h(X) \in \mathbb{R}$), which can be considered harder to interpret. Similarly, by involving an integral, it is harder to calculate, though closed-form expressions do exist for some known probability distributions, such as the Normal, whose differential entropy can easily be calculated by merely knowing its dispersion (Cover and Thomas, 2006: 244):

$$h(X) = \frac{1}{2} \log 2\pi e \sigma^2 \tag{10}$$

For example, if X is a standard Normal variable (i.e., $X \sim \mathcal{N}(0, 1)$), this means that its differential entropy is $\frac{1}{2}\log 2\pi e \approx 2.047$ bits.

Following Delgado-Bonal and Marshak (2019), the concept of entropy can be found useful in relation to the states of knowledge previously conceptualized using Bayesian inference in section (3.1). Since that tentative conceptualization used probability distributions to characterize prior and posterior knowledge, which can be considered 'random' in an epistemological sense of the word (cf. McElreath, 2019), the approaches can easily be combined, leading to the notion that the different states of knowledge implied by the prior and posterior distributions simply express different *states of information*. Specifically, the prior distribution has an entropy, which express the degree of uncertainty in one's information about the parameter of interest θ; the wider that distribution, the greater the uncertainty about θ. The posterior distribution, by comparison, differs from the prior to the extent

[12] In statistical analysis, rather than studying a phenomenon of interest by itself, researchers generally study the parameters of a statistical model theorized to generate said phenomenon. This often involves *generalized linear models* (GLMs; Field, 2018; Kruschke, 2014) that assume *linearity* (and thus *continuity*) in its parameters (Wooldridge, 2019: 763).

[13] Since (9) involves an integral, the calculation of differential entropy presupposes that the density function and integral for X exists (Cover and Thomas, 2006: 243-244).

that the analyzed data has increased or reduced the uncertainty around θ. For example, the Normal prior distributions with a dispersion (σ) of 1 previously specified (see Section 3.1.1) have a differential entropy of approximately 2.047 bits, while those with a dispersion of 0.5 have a differential entropy of approximately 1.047 bits.

The example demonstrates that the concept of (differential) entropy can be imported to Bayesian inference and used to characterize the information contained within a state of knowledge. However, the aforementioned example also makes it clear that a mere comparison of (differential) entropy between the prior and posterior could fail to consider changes in the plethora of ways by which the prior and posterior distributions could differ, since the discrepancy between the differential entropy of the posterior previously mentioned in section (3.1.2) and the first and third prior (i.e., $\Delta h = 1$ bits), and the second and fourth prior (i.e., $\Delta h = 0$ bits), are the same, despite having different means. This contradicts the established intuition suggesting that the degrees of discrepancy between them, and thus the induced surprise, should differ.

While the concept of (differential) entropy has thus shown to be a useful foundation to link concepts from Bayesian inference and Information theory, it is clearly insufficient for appropriately conceptualizing and measuring surprise. This can be mitigated by introducing the final concept of relative entropy.

3.2.2 Relative entropy

The *relative entropy*, also known as the Kullback-Leibler (1951) divergence (KL divergence; e.g., Pérez-Cruz, 2019: 1) quantifies information divergence by being 'a measure of the distance between two distributions.' (Cover and Thomas, 2006: 19). Unlike entropy, which quantifies the uncertainty of a distribution, the relative entropy quantifies the divergence between distributions, with 'divergence' simply referring to the 'statistical distance' between them (which should *not* be mistaken for a *geometric* distance; Nielsen, 2019: 4). It has proven 'central to information theory and statistics' (Pérez-Cruz, 2019: 1084; see also Cover and Thomas, 2006: 347-408) and seen application across numerous disciplines (e.g., Burnham and Anderson, 2001; Chaloner and Verdinelli, 1995).

For two discrete random variables p and q, the relative entropy is denoted $D(p||q)$ and mathematically defined by formula (11) (Cover and Thomas, 2006: 19; Delgado-Bonal and Marshak, 2019: equation 5); and if p and q are continuous random variables with probability density functions f and g, it is given by formula (12) (Cover and Thomas, 2006: 251; Pérez-Cruz, 2019: 1).

$$D(p||q) \equiv \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{11}$$

$$\equiv \int f \log \frac{f}{g} \tag{12}$$

The properties of relative entropy for the discrete and continous case are mostly the same (Cover and Thomas, 2006: 251), for example, they satisfy the property of *information monotonicity* by being invariant to reparameterizations (cf. Nielsen, 2019: 3), the relative entropy is zero *if and only if* (iff) $p = q$ for the discrete case, and $f = g$ (almost) everywhere for the continuous case (Cover and Thomas, 2006: 252-253; see also Pérez-Cruz, 2019: 1). Since it is taken by convention that $0\,\frac{0}{0} \equiv 0$, $0\,\frac{0}{q} \equiv 0$, and $p\,\frac{p}{0} \equiv \infty$ (Cover and Thomas, 2006: 19), the relative entropy can for both (11) and (12) be infinite, with finite values only occurring for the continuous case when the support set of f is contained in the support set of g (i.e., if f is *absolutely continuous* with respect to g). Thus, the relative entropy is generally a non-negative scalar on the *extended real number line* (i.e., $D(p||q) \in \bar{R}^{[0;\infty]}$), and if a base of 2 is used for the logarithm, the relative entropy expresses the information divergence in bits, with a unit of surprise being denoted one 'wow' (cf. Itti and Baldi, 2009).

While the relative entropy was developed to quantify how easy it is to discriminate between two distributions (Kullback and Leibler, 1951: 79), it is generally interpreted as a quantification of how inefficient it is to assume that a random variable is distributed as q when it is actually distributed as p (Cover and Thomas, 2006: 19). This has gained widespread appeal, since researchers may interpret the relative entropy as the mean of the *log-likelihood ratio* (Pérez-Cruz, 2019: 1). This makes it useful for statistical hypothesis testing and model selection (Cover and Thomas, 2006: 375-380; see, e.g., King, 1998: 84-86; McElreath, 2019: 202-214), since it can be used to justify preferring a statistical model with the lowest relative entropy on the basis of *parsimony* (i.e., Occam's razor; Cover and Thomas, 2006: 1).

That interpretation can be recognized as similarly useful for the current purposes of conceptualizing surprise, since it may be meaningfully rephrased so that the relative entropy 'reflects the gain of information resulting from adding the knowledge in p relative to q, or the gain of information when we learn that the real probability function was p instead of q as we thought initially.' (Delgado-Bonal and Marshak, 2019: 7) Such an interpretation fits the ongoing Bayesian (re)conceptualization of surprise by enabling the relative entropy between the prior and posterior distribution to be meaningfully interpreted as the gain in information resulting from adding the knowledge of the data to the prior. This means that divergence, by merit of being conceptually based in Bayesian inference and Information theory, coherently reflects how unexpected the results are compared to one's expectations (cf. Itti and Baldi, 2009), with zero divergence/surprise occurring iff the results exactly match one's expectations.

This leads to the completed, formal conceptualization of surprise as *the relative entropy between prior and posterior knowledge*. Within this conceptualization, knowledge and information are thus equated, with the prior and posterior being taken to reflect two states of knowledge, and the discrepancy between them is the degree of surprise. Higher degrees of discrepancy thus reflect greater surprise, measured in units of the *base* used for the logarithmic function, which here is specified as 2 so that the units of surprise are in bits.

This conceptualization fits perfectly with the informal notion of surprise identified among social science researchers (Section 2.1). This is evident, because the informal conceptualization of surprise involved a discrepancy between results and expectations/predictions about those results based in theory, which is directly reflected in this formal conceptualization, since the relative entropy measures the divergence (i.e., discrepancy) between the posterior knowledge (i.e., results given the data) and the prior knowledge (i.e., predictions based in theory). It also reflects the *relative* and *dynamic* characteristics of surprise, since the degree of surprise within this formal conceptualization is always relative to the prior, with different degrees of surprise being possible if the posterior is compared to different prior specifications representing different theoretical expectations, and as the prior is continuously updated based on new data to better reflect reality, the relative entropy, and thus the surprise, will lessen. This formal (re)conceptualization of surprise thus offers a theoretically coherent understanding of surprise that fits with social science researchers' intuitive use of the word, making it particularly suited for that field.

Now, perhaps attributable to its coherent conceptualization and intuitive interpretation, this is not the first time that relative entropy has been proposed as part of a formal conceptualization of surprise across different scientific fields. For example, it has previously been proposed within the disciplines of clinical care (e.g., Burkhart et al., 2025), ecology (e.g., Burnham and Anderson, 2001), software engineering (e.g., Bencomo and Belaggoun, 2014), and computer science (e.g., Itti and Baldi, 2009). However, while these proposals are undoubtedly helpful in reducing the mispractice of using the p-value as an implicit measure of surprise, their choice to not just conceptualize surprise as the relative entropy, but also as its measure, can be considered suboptimal given its limitations.

A limitation of relative entropy is that it is *not* a 'proper' metric in the geometric sence of the word. This is because it is *asymmetric*, which generally means that $D(p||q) \neq D(q||p)$, and it fails to satisfy the *triangle inequality* (Cover and Thomas, 2006: 19; Delgado-Bonal and Marshak, 2019: 7), whereby it lacks a series of intuitive properties that one may generally attribute to a distance. For example, suppose A, B, and C represent points in Euclidean space. Let the measure of the distance travelled be asymmetric and in violation of the triangle inequality. Then, the measured distance directly from A to C may *not* be the same as the distance travelled directly from C to A, and either of these may be longer than the distance travelled from A to B to C. This is counter-intuitive (cf. Cover and Thomas, 2006: 367; for an alternative view, see, e.g., McElreath, 2019: 208-209), since from real-world experience, one would expect distances to be symmetric, with stops and detours either adding or not influencing the distance travelled. These limitations thus make the relative entropy harder to interpret and invites researcher discretion by being sensitive to the arguably arbitrary choice of whether to calculate it between p and q or between q and p.

The relative entropy/KL divergence is by no means the only measure of divergence, with numerous symmetric alternatives (e.g., Jeffreys's (1946) divergence; Cha, 2007), some of which can be considered proper geometric measures of distance. For example, the Bhattacharyya (1943; 1946) distances, the Earth mover's distance (Kantorovich, 1939; Monge, 1781; Vaseršteĭn, 1969), and the Mahalanobis' (1936) generalized distance (Kullback and Leibler, 1951: 79). While each of these measures of distance possess advantages and disadvantages, the one proposed here to measure surprise within the social sciences builds on the Jensen-Shannon divergence, that, while itself based on the relative entropy, possesses none of the limitations mentioned here.

3.2.3 Jensen-Shannon divergence

Formally introduced by Lin (1991; see also Wong and You, 1985), the *Jensen-Shannon divergence* (JS divergence) is a popular alternative measure of divergence (Cha, 2007; Nielsen, 2019). Based on the KL divergence, it is mathematically defined with the formula (Lin, 1991: 146; Nielsen, 2019: 3):

$$\mathrm{JS}(p||q) \equiv \frac{\mathrm{D}(p||M) + \mathrm{D}(q||M)}{2} \tag{13}$$

where $M = \frac{p+q}{2}$ is the *mixture distribution* of p and q. From this formula, it can be recognized that the JS divergence is related to the KL divergence by being the arithmetic mean of the relative entropies between p and q and their mixture distribution (Lin, 1991: 146). JS divergence shares numerous properties of the KL divergence useful to conceptualizing and measuring information divergence (Lin, 1991; Nielsen, 2019: 3), for example, being zero iff $p = q$ (Lin, 1991: 146-147).

While the JS divergence is closely related to the KL divergence, it possesses numerous advantages. For example, the JS divergence is always finite due to involving the mixture distribution, since this means that p and q need not have overlapping support sets (i.e., they need *not* be absolutely continuous with respect to each other; Lin, 1991: 146-147).

Another advantage is that the JS divergence generally is a non-negative, real-valued scalar with an upper-bound of log 2 (i.e., JS $\in \mathbb{R}^{[0;\log 2]}$; Lin, 1991: 147; Nielsen, 2019: 3), which means that the upper bound is 1 if using base 2 for the logarithm. These bounds can be considered common for measures within the social sciences, for example, explained variance (R^2, Gelman et al., 2019), congeneric reliability (ρ_c; Jöreskog, 1971), and Sklar's ω (Hughes, 2023), making this measure more intuitively interpretable within that discipline.

Perhaps more importantly, unlike relative entropy, the JS divergence is symmetric, and by taking its square-root (i.e., $\sqrt{\text{JS}}$), one obtains the *Jensen-Shannon distance* (*JS distance*), which satisfies the triangle inequality (Nielsen, 2019: 3), making it a proper geometric measure of distance. Hence, these Jensen-Shannon-based measures may be preferred to the KL divergence as the actual measure of the 'relative entropy' between prior and posterior knowledge, and with its approximately linear properties as opposed to the more quadratic behavior of the JS divergence, the JS distance may ultimately be preferred in terms of quantifying surprise. That the JS distance is conceptually appropriate is further corroborated by the notion that it was originally conceptualized in relation to Bayesian decision-making (Lin, 1991: 147-148), making it a natural choice for use in relation to this Bayesian Information-theoretic conceptualization of surprise.

These considerations thus serve this formal conceptualization of surprise by the JS divergence enabling a symmetric measurement of the discrepancy between the prior and posterior distributions, with the JS distance providing an intuitively interpretable quantification of the distance between prior and posterior knowledge, and its $0 - 1$ bounds (when using base 2) further permits an interpretation of this distance as the (percentage) degree of surprise. This means that, unlike Itti and Baldi's (2009) conceptualization of 'wow' units, the JS distance thus provides a measure of the *degree* to which one is 'wow'ed'.

If the prior and posterior distribution are Normal, then the JS distance is relatively easily computed using numerical integration. For example, if one again considers the Normal prior distributions in figure (1) in section (3.1.1) with the posterior being equivalent to Prior #4, then the JS divergence between the posterior and each prior can be computed as approximately .684, .913, .910, and .000 bits, respectively. These values respectively correspond to JS distances of .827, .955, .954, and .000, with, for example, the distance between the posterior and the second prior here being meaningfully interpretable as a degree of surprise of 95.5%.

These values fit the conceptual expectations about surprise, since the degree of surprise implied by the discrepancy between the posterior and Prior #4 is effectively zero (i.e., 0%), consistent with the complete absence of surprise when results match the theoretical expectations; the degree of surprise is higher when the theoretical expectation is more certain, as revealed by the higher distance between the posterior and the narrower Prior #2 (i.e., 91.3%) than when compared to the wider Prior #1 (i.e., 68.4%); and the degree of surprise is also higher if the theoretical expectation incorrectly predicts the magnitude and direction of the results, with the posterior-prior distance being higher for Prior #3 (i.e., 91%) than Prior #1 (i.e., 68.4%), since the former expects the parameter of interest to be positive when it instead is negative.

While the JS divergence/distance is thus useful, these examples have the fundamental limitation of presupposing knowledge of the probability density functions of the prior and posterior. In practice, the prior and posterior need not necessarily be univariate Normal as these examples are, but can take on multiple forms, some of which have no closed-form expression. However, this is *not* a critical issue due to having built this conceptualization on Bayesian inference, whose reliance on MCMC procedures theoretically can be used to estimate almost any arbitrary distribution. While this solves the problem of identifying the distributions of the prior and posterior, this does not solve the problem of calculating the JS distance. This is a common problem in Information theory (Beirlant et al., 2001: 11), where, for example, (relative) entropy must be estimated from the data, which can require large and representative samples (cf. Back et al., 2018: 1; Schürmann and Grassberger, 1996; though see, e.g., Bonachela et al., 2008; Grassberger, 1988; Lesne et al., 2009). For the continuous case, this is further troubled by the need to estimate the density of the particular distribution, which is needed in relation to differential entropy, KL divergence, and the JS divergence. Fortunately, parametric assumptions can sometimes be justified to ease estimation (e.g., Jaynes, 1968: 236-237), and a multitude of density estimators otherwise exist (e.g., Ahmad and Lin, 1975; Berrett et al., 2019; Delgado-Bonal and Marshak, 2019; Hall and Morton, 1993; Pérez-Cruz, 2019), which may be utilized to compute the JS distance.

4 Demonstration

Having thus far provided a formal (re)conceptualization of surprise specifically suited for the social sciences, this section provides a practical demonstration in relation to empirical data from that field to show how these researchers can use this to measure surprise. In line with the principles of *open science* (Christensen et al., 2019; Xie, 2014), all data processing and analysis was done with the R programming language (R Core Team, 2025) in the RStudio IDE (Posit Team, 2025).[14]

This demonstration is based on the research article by Todorov et al. (2005) published in the prestigious journal *Science*. Their article details a series of studies conducted on (under)graduate students at Princeton University in the United States (US; see Todorov et al., 2005, and their supplementary materials), who in one study were individually exposed for 1-second to pairs of standardized black-and-white head-shot photographs of real-world political candidates from the Democratic party or the Republican party, who were either the winner or the runner-up to elections for the US Senate or House of Representatives. Here, they used an RCT design to randomize the sequence of pairs and the left-right placement of photos, with participants being instructed to choose the more competent-looking candidate. Based on this arguably *exogenous* assignment of photos, the authors could credible infer *causation* (cf. Holland, 1986; Imbens and Rubin, 2015), stating that these 'inferences of competence based solely on facial appearance predicted the outcomes of U.S. congressional elections better than chance' (Todorov et al., 2005: 1623). This result was arguably rather surprising, since the dominant theory at the time assumed voting choices to be 'based primarily on rational and deliberative considerations' (Todorov et al., 2005: 1623). Despite the recognized novelty of these results, the degree of surprise was never quantified, making this study suitable for a demonstration of this paper's conceptualization of surprise.

For this demonstration, a publicly available subset of the data by Todorov et al. (2005), provided in Imai (2017), is used. This data constitutes a 119 by 10 matrix, where each row is a pair of political candidates, while the columns are variables indicating the election year, US state, the name of the winner and runner-up, the political party affiliation of each candidate, their absolute vote share, and average competence rating by the participants. With no missing values, the data provides a statistical power of 80% to detect a *product-moment correlation coefficient* (r; Bravais, 1844; Pearson, 1895) as small as $|r| = 0.473$ at the 5% statistical discernibility level.

In accordance with Todorov et al. (2005; see also Imai, 2017: 139-148), data is processed so that the only variables are a mean competence score (i.e., C) and a difference in relative vote share (i.e., V) for each candidate.

[14] For software specifications and replication materials, see section 8.1 of the Appendix.

V is then modeled as a conditionally linear-Normal function of C, with the coefficient of C (i.e., β_C) thus being the parameter of interest (i.e., the *estimand*), while an intercept and residual dispersion (σ) serve as nuisance parameters.

For ease of specifying a prior for β_C and σ, both of these variables are rescaled using *Fisher z-scoring* (henceforth denoted *standardization*; Field, 2018: 37), meaning that values greater than zero indicate a competence score in favor of the democratic candidate for C, while values values greater than zero indicate a vote share in favor of the democratic candidate for V. By standardizing the variables, this eliminates the need to estimate the intercept and introduces the following bounds for β_C and σ: $0 \leq \sigma \leq 1$ and $-1 \leq \beta_C \leq 1$ (cf. Agresti, 2018: 273; Wooldridge, 2019: 184:185), which need to be accounted for in their respective prior specifications.

Using this paper's formal conceptualization of surprise, the priors for β_C and σ should ideally be specified so that they accurately reflect the expectations/predictions of theory. Otherwise, the discrepancy between the posterior and prior will otherwise *not* reflect the genuine surprise of the results. While Todorov et al. (2005: 1623) make numerous arguments that can be used to justify a theoretical expectation of facial competence as having a zero influence on the vote share, the certainty around this expectation is unclear. This is partly because the authors refer to numerous existing studies on facial evaluations, competence, and political candidates (e.g., Hassin and Trope, 2000; Haxby et al., 2000; Kinder et al., 1980; Todorov and Uleman, 2003; Winston et al., 2002), and while this helps the authors' argument, it also has the effect of reducing the certainty of opposing theories, thus making eventual positive results comparatively less surprising.

$$
\begin{gathered}
V_i|\mu_i, \sigma \sim \mathcal{N}(\mu_i, \sigma^2) \text{ for } i \text{ in } 1, \ldots, n \\
\mu_i = \beta_C C_i \text{ for } i \text{ in } 1, \ldots, n \\
\beta_C = \mathcal{N}^{[-1;1]}(0, 0.1) \\
\sigma = \mathcal{N}^{[0;1]}(1, 0.1)
\end{gathered} \tag{14}
$$

Nevertheless, for the purposes of this demonstration, the prior for β_C is specified as a Normal distribution, with a mean of 0 and a dispersion of 0.1, which in the context of Todorov et al.'s (2005) study can be taken to reflect a state of knowledge, where one would be comfortably skeptical about the influence of facial competence on vote outcomes. Stated differently, the expectation implied by this prior is that one would believe with 95% credibility that the absolute effect would maximally correspond to a 'small' or 'medium' effect (i.e., $|r| \approx 0.2$; cf. Cohen, 1988; Funder and Ozer, 2019; Gignac and Szodorai, 2016; Lovakov and Agadullina, 2021). Using a similar logic for the prior specification of σ, the statistical model used to estimate β_C is formalized using common notation (e.g., Gelman et al., 2014; Levy and Mislevy, 2020) and shown in model (14).

This model is then fit to the data using the `brms` R package (Bürkner, 2017, 2018), which exploits the *No-U-Turn sampler* (NUTS; Hoffman and Gelman, 2014) of the *STAN probabilistic programming language* (Stan Development Team, 2025b) to enable an efficient Hamiltonian MCMC sampling (Kruschke, 2014: 400-416). For the purposes of assessing the convergence of this sampler, 15 independent sampling chains are used, each with 2,000 warm-up samples to help them identify the posterior in advance of actual sampling, with 40,005 samples being drawn in total afterwards. This applies to the sampling of both the prior and the posterior distributions, since both are needed for computing the JS distance.

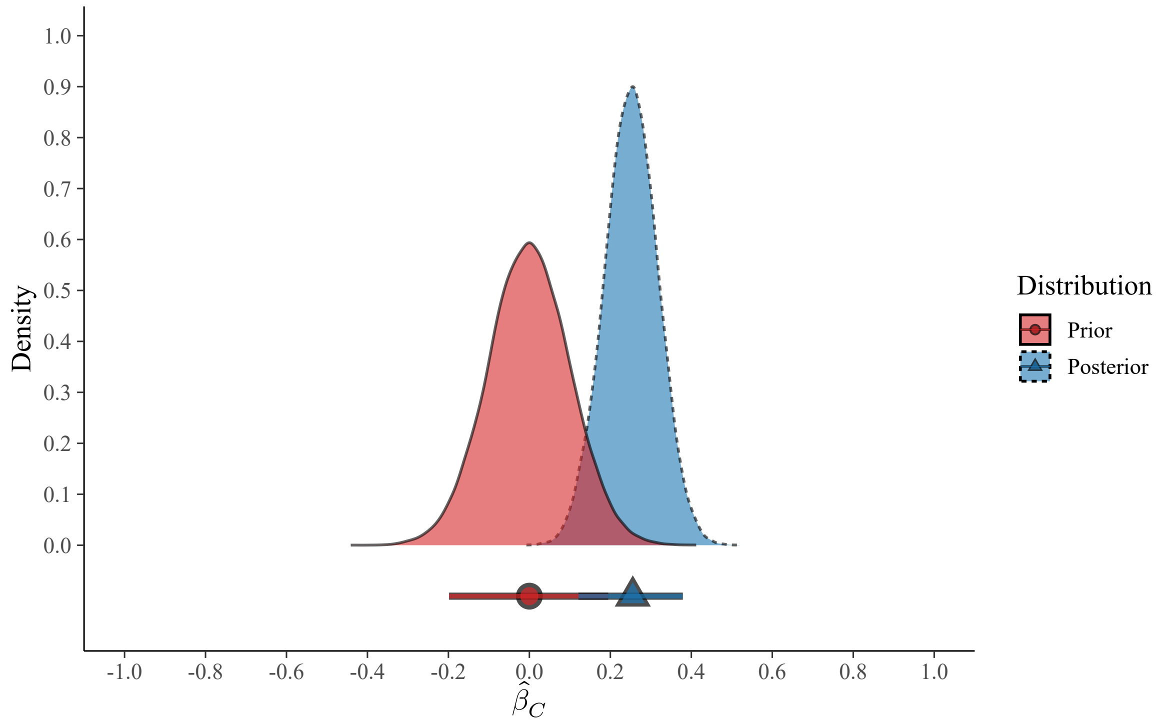

Figure 2. Estimated Effect of Facial Competence on Relative Vote Share

NOTE: Prior and posterior distributions of the estimated effect of facial competence on relative vote share. Prior/posterior draws = 40,005. The geometrics below the distributions indicate the mode (circle) and the 95% HDCI (bar). Jensen-Shannon divergence = .760 bits. Results based on data ($n = 119$) by Todorov et al. (2005; provided in Imai, 2017).

Visual inspections of *trace plots* (Kruschke, 2014: 179-180) and metric assessments of the *Gelman-Rubin convergence metric* ($\hat{R}$) and the *Monte-Carlo standard error* (MCSE) found results consistent with convergence (i.e., $\hat{R} \leq 0.01$, Vehtari et al., 2021; MCSE ≤ 0.01, Kruschke, 2014: 186-187). Accounting for autocorrelation in the MCMC samples reveals *effective sample sizes* (ESS) that could be considered sufficiently high (ESS $\geq 10,000$; Kruschke, 2014: 182-184). The explanatory power of this model (i.e., its R^2; Gelman et al., 2019) is .073 (SD = .035; 95% HDI[.010; .140]), which can be interpreted as 'weak' (cf. Cohen, 1988) with 89.3% credibility. As revealed by a *posterior predictive p-value* (PP_P; Gelman et al., 1996) of .970, this shortcoming is likely attributable to the model's tendency to make over-exaggerated predictions.

The sampled prior and posterior are displayed in figure (2), with a metric summary being provided in table (1). Inspecting the prior distribution reveals an expected estimated effect of facial competence on vote share before seeing the data of -.001 (SD = .100; 95% HDI[-.199; .194]). By comparison, the posterior distribution reveals an expected estimated effect of facial competence on vote share after seeing the data of .251 (SD = .066; 95% HDI[.123; .380]). Following King et al. (2000), this can be meaningfully interpreted as the following: *Ceteris paribus*, as the evaluated facial competence of a political candidate increases by one SD, their relative vote share increases, on average, by .251 SDs, plus or minus .066 SDs, and between .123 and .380 SDs with 95% credibility. This can be considered a 'medium'-sized effect (cf. Funder and Ozer, 2019) with 55.4% credibility.

The regularizing effect of the theoretically-motivated 'skeptical, yet persuadable' prior on the results can be made evident by comparing the posterior to the results from a Frequentist version of model (14), estimated with *ordinary least squares* (OLS; Wooldridge, 2019: 70-73) and the *non-parametric bootstrap* (Efron, 1979, 2003; Efron and Tibshirani, 1994), with 40,005 bootstrapped samples being used to compute the standard error (SE) and a 95% *bias corrected and accelerated interval* (BCaI; Efron, 1987; DiCiccio and Efron, 1996; Makowski et al., 2019b). This yields an expected estimated effect of facial competence on vote share of .432 (SE = .074; 95% BCaI[.275; .565]). The relatively substantial and credible difference between the Bayesian and Frequentist approach (i.e., $\Delta\hat{\beta}_C$ = -.181; SD = .098; 95% HDI[-.026; -.318]) is likely attributable to the relatively small sample used in this demonstration (i.e., $n = 119$), and it indicates that the prior will have greater influence on an assessment of surprise in relation to this data, since the posterior is relatively sensitive to its specification.

Table 1. Relationship between Facial Competence & Relative Vote Share

$\hat{\theta}$	**Parameter**	**Posterior**	$\hat{R}$	**MCSE**	**ESS**	**Prior**	**Surprise**
$\hat{\beta}_C$	Facial Competence	$.251\ (.066)\ [.123; .380]$	1.001	.001	14,865	$-.001\ (.100)\ [-.199; .194]$	.872
$\hat{\sigma}$	RMSE	$.932\ (.041)\ [.857; 1.000]$	1.002	.000	11,603	$.921\ (.060)\ [.805; 1.000]$	.194

NOTE: Mean estimated standardized parameters, with posterior SDs in parentheses and 95% HDIs in brackets. $n = 119$. MCMC samples = 40,005. RMSE = Root mean squared error. $\hat{R}$ = Gelman-Rubin convergence metric (Vehtari et al., 2021). MCSE = Monte-Carlo standard error (Kruschke, 2014: 186-187). ESS = Effective sample size (Kruschke, 2014: 182-184). Surprise = Jensen-Shannon distance. Bayesian R^2 (Gelman et al., 2019) = .073 (SD = .035; 95% HDI[.010; .140]). Posterior predictive p-value (PP_P; Gelman et al., 1996) = 0.97.

Based on this paper's conceptualization, the degree of surprise resulting from this study can then be measured as the Jensen-Shannon distance. While the posterior and prior could be considered approximately Normal, enabling a parametric estimation of their discrepancy (e.g., Jaynes, 1968: 236-237), this will *not* always be the case, which is why the computation of the JS distance involved in this demonstration instead considers a *non-parametric kernel density plug-in estimator* (KDE), and while such an estimator can be considered imperfect (cf. Berrett et al., 2019; Pérez-Cruz, 2019), it is generally adequate for estimating densities of univariate distributions (e.g., Ahmad and Lin, 1975; Hall and Morton, 1993; Parzen, 1962; Rosenblatt, 1956), especially considering the relatively large number of samples (i.e., $40,000+$). Provided in the Appendix (see Section 8.2), this function used for this example also mimics similar estimators (e.g., Drost, 2018), for example, by applying common adjustments to avoid a division by zero. While the function uses an efficient KDE (Wand, 2025), similarly applied for density estimation in relation to *Bayes factors* (Makowski et al., 2019b; Wagenmakers et al., 2010), to make it scalable for the large number of MCMC samples commonly drawn in relation to Bayesian inference, the author makes no claim that the function is optimal, and it merely serves as a functional demonstration for the purposes of this paper.

Computing the JS divergence in this way yields an informational divergence between the posterior and prior of .760 bits. This can be translated into the JS distance to arrive at an estimated degree of surprise of 87.2%. In line with an intuitive/qualitative assessment of these results, this degree of surprise can be considered relatively large. By comparison, a parametric approach based in numerical integration yields a JS divergence of 0.767, similarly interpretable as a degree of surprise of 87.6%, indicating that the estimated surprise is relatively robust and that a parametric approach could be considered appropriate.

5 Discussion

[Work-in-progress]

6 Conclusion

[Work-in-progress]

7 Acknowledgements

[Work-in-progress]

References

Abramowitz, A. I. and Webster, S. W. (2018). Negative partisanship: Why Americans dislike parties but behave like rabid partisans. *Political Psychology*, 39:119–135.

Agresti, A. (2018). *Statistical Methods for the Social Sciences*. Pearson, 5 edition. Global Edition.

Ahmad, I. A. and Lin, P.-E. (1975). A nonparametric estimation of the entropy for absolutely continuous distributions.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics*, volume 1, page 610–624. Springer-Verlag. Republished version of B. N. Petrov and F. Csáki (eds.): 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akadémiai Kiadó, pp. 267–281.

Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G., and Eaves, L. J. (2011). The politics of mate choice. *Journal of Politics*, 73:362–379.

Allaire, J. J., Francois, R., Ushey, K., Vandenbrouck, G., Geelnard, M., and Intel (2025). *RcppParallel: Parallel Programming Tools for 'Rcpp'*. R package version 5.1.10.

Allaire, J. J., Xie, Y., Dervieux, C., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., and Iannone, R. (2024). *rmarkdown: Dynamic Documents for R*. R package version 2.29.

Altay, S., Hacquin, A.-S., and Mercier, H. (2020). Why do so few people share fake news? it hurts their reputation. *New Media and Society*.

American Statistical Association (2016). American statistical association releases statement on statistical significance and p-values: Provides principles to improve the conduct and interpretation of quantitative science. Accessed March 18th, 2025.

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3.

Bache, S. M. and Wickham, H. (2022). *magrittr: A Forward-Pipe Operator for R*. R package version 2.0.3.

Back, A., Angus, D., and Wiles, J. (2018). Determining the number of samples required to estimate entropy in natural sequences. *arXiv*, 1805.08929v1:1–6.

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., Schwendinger, B., and Krylov, I. (2025). *data.table: Extension of 'data.frame'*. R package version 1.17.8.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bates, D., Maechler, M., and Jagan, M. (2025). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.7-2.

Bates, D., Mullen, K. M., Nash, J. C., and Varadhan, R. (2024). *minqa: Derivative-Free Optimization Algorithms by Quadratic Approximation*. R package version 1.2.8.

Bayes, T. and Price, R. (1763). An essay towards solving a problem in the doctrine of chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, A.M.F.R.S. *Philosophical Transactions of the Royal Society of London*, 53:370–417.

Beirlant, J., Dudewicz, E. J., Györfi, L., and van der Meulen, E. C. (2001). Nonparametric entropy estimation: an overview. Unpublished manuscript.

Beleites, C. (2020). *arrayhelpers: Convenience Functions for Arrays*. R package version 1.1-0.

Ben-Shachar, M. S., Lüdeck, D., and Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56):2815.

Bencomo, N. and Belaggoun, A. (2014). A world full of surprises: Bayesian theory of surprise to quantify degrees of uncertainty. ICSE Companion 2014, pages 460–463, New York, NY, USA. Association for Computing Machinery.

Bengtsson, H. (2025). *matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors)*. R package version 1.5.0.

Berrett, T. B., Samworth, R. J., and Yuan, M. (2019). Efficient multivariate entropy estimation via k-nearest neighbour distances. *The Annals of Statistics*, 47(1):288–318.

Bhattacharyya, A. K. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109.

Bhattacharyya, A. K. (1946). On a measure of divergence between two multinomial populations. *Sankya*, 7(4):401–406.

Bishop, D. V. M. (2006). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett lecture. *Quarterly Journal of Experimental Psychology*, 73(1):1–19.

Bolinger, D. (1971). Semantic overloading: A restudy of the verb remind. *Language*, 47(3):522–547.

Bolker, B. (2025). *reformulas: Machinery for Processing Random Effect Formulas*. R package version 0.4.1.

Bonachela, J. A., Hinrichsen, H., and Muñoz, M. A. (2008). Entropy estimates of small data sets. *Journal of Physics A: Mathematical and Theoretical*, 41(202001):1–9.

Boshnakov, G. N. (2025). *Rdpack: Update and Manipulate Rd Documentation Objects*. R package version 2.6.4.

Boshnakov, G. N. and Putman, C. (2024). *rbibutils: Read 'Bibtex' Files and Convert Between Bibliography Formats*. R package version 2.3.

Bravais, A. (1844). *Analyse mathematique sur les probabilités des erreurs de situation d'un point*. Impr. royale. In French.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall, 1 edition.

Burkhart, M. C., Ramadan, B., Solo, L., Parker, W. F., and Beaulieu-Jones, B. K. (2025). Quantifying surprise in clinical care: Detecting highly informative events in electronic health records with foundation models.

Burnham, K. P. and Anderson, D. R. (2001). Kullback–leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, 28(2):111–119.

Buunk, A. P. and van Vugt, M. (2013). *Applying Social Psychology. From Problems to Solutions*. Sage, 2 edition.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411.

Bürkner, P.-C., Gabry, J., Kay, M., and Vehtari, A. (2025). posterior: Tools for working with posterior distributions. R package version 1.6.1.

Cambridge Dictionary (2025). surprise.

Canty, A. and Ripley, B. D. (2024). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-31.

Cha, S.-H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 4(1):300–307.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304.

Champely, S. (2020). *pwr: Basic Functions for Power Analysis*. R package version 1.3-0.

Chang, W. (2024a). *cachem: Cache R Objects with Automatic Pruning*. R package version 1.1.0.

Chang, W. (2024b). *fastmap: Fast Data Structures*. R package version 1.2.0.

Chang, W. (2025). *R6: Encapsulated Classes with Reference Semantics*. R package version 2.6.1.

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2025a). *shiny: Web Application Framework for R*. R package version 1.11.1.

Chang, W., Cheng, J., and Gao, C. (2025b). *later: Utilities for Scheduling Functions to Execute Later with Event Loops*. R package version 1.4.2.

Cheng, J. (2025a). *miniUI: Shiny UI Widgets for Small Screens*. R package version 0.1.2.

Cheng, J. (2025b). *promises: Abstractions for Promise-Based Asynchronous Programming*. R package version 1.3.3.

Cheng, J., Chang, W., Reid, S., Brown, J., Trower, B., and Peslyak, A. (2025). *httpuv: HTTP and WebSocket Server Library*. R package version 1.6.16.

Cheng, J., Sievert, C., Schloerke, B., Chang, W., Xie, Y., and Allen, J. (2024). *htmltools: Tools for HTML*. R package version 0.5.8.1.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and Reproducible Social Science Research: How to do Open Science*. University of California Press, Oakland, California.

Clayton, A. (2021). *Bernoulli's Fallacy: Statistical Illogic and the Crisis of Modern Science*. Columbia University.

Cochrane, R., Billig, M., and Hogg, M. (1979). Politics and values in Britain: A test of Rokeach's two-value model. *British Journal of Clinical Psychology*, 18:159–167.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge, New York, 2nd edition.

Cohen, J. (1994). The Earth is round ($p < .05$). *The American Psychologist*, 49(21):997–1003.

Cole, S. R., Edwards, J. K., and Greenland, S. (2021). Surprise! *American Journal of Epidemiology*, 190(2):191–193.

Coombes, K. R., Brock, G., Abrams, Z. B., and Abruzzo, L. V. (2019). Polychrome: Creating and assessing qualitative palettes with many colors. *Journal of Statistical Software, Code Snippets*, 90(1):1–23.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. John Wiley and Sons, 2 edition.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, 14(1):1–13.

Csárdi, G. (2019). *pkgconfig: Private Configuration for R Packages*. R package version 2.0.3.

Csárdi, G. (2025). *cli: Helpers for Developing Command Line Interfaces*. R package version 3.6.5.

Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan, M., and Tenenbaum, D. (2024). *remotes: R Package Installation from Remote Repositories, Including 'GitHub'*. R package version 2.5.0.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge. ISBN 0-521-57391-2.

Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science and Medicine*, 210:2–21.

Delgado-Bonal, A. and Marshak, A. (2019). Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy*, 21:541.

Dellsén, F. (2024). *Abductive Reasoning in Science*. Cambridge University Press.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, 11(3):189–212.

Drost, H.-G. (2018). Philentropy: Information theory and distance quantification with R. *Journal of Open Source Software*, 3(26):765.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. (2017). *random: True Random Numbers using RANDOM.ORG*. R package version 0.2.6.

Eddelbuettel, D. (2024). *digest: Create Compact Hash Digests of R Objects*. R package version 0.6.37.

Eddelbuettel, D. and Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician*, 72(1):28–36.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Eddelbuettel, D., Francois, R., Allaire, J. J., Ushey, K., Kou, Q., Russell, N., Ucar, I., Bates, D., and Chambers, J. (2025). *Rcpp: Seamless R and C++ Integration*. R package version 1.1.0.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397):171–185.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science*, 18(2):135–140.

Efron, B. and Tibshirani, R. (1994). *An introduction to the bootstrap*. Chapman and Hall.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. SAGE, 5 edition.

Flake, J. K. and Fried, E. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4):456–465.

French, S. and Murphy, A. (2023). The value of surprise in science. *Erkenntnis*, 88:1447–1466.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Funder, D. C. and Ozer, D. J. (2019). Evaluating effect size in psychological research: sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2):155–168.

Gabry, J., Goodrich, B., Lysy, M., and Johnson, A. (2024). *rstantools: Tools for Developing R Packages Interfacing with 'Stan'*. R package version 2.4.0.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A. (2019). Visualization in Bayesian workflow. *The Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 182:389–402.

Gagolewski, M. (2022). stringi: Fast and portable character string processing in R. *Journal of Statistical Software*, 103(2):1–59.

Gelman, A. (2024). A chain as strong as its strongest link? understanding the causes and consequences of biases arising from selective analysis and reporting of research results. *Journal of Research on Educational Effectiveness*, 17(3):459–461.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman and Hall, 3 edition.

Gelman, A., Goodrich, B., Gabry, J., and Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3):307–309.

Gelman, A., Meng, X.-L. M., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6(4):733–760.

Gelman, A. and Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, 60(4):328–331.

Genz, A. and Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33(5):587–606.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2):198–218.

Gignac, G. E. and Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual differences*, 102:74–78.

Gordon, M., Gragg, S., and Konings, P. (2024). *htmlTable: Advanced Tables for Markdown/HTML*. R package version 2.4.3.

Goren, P. and Chapp, C. (2017). Moral power: How public opinion on culture war issues shapes partisan predispositions and religious orientations. *The American Political Science Review*, 111(1):110–128.

Grassberger, P. (1988). Finite sample corrections to entropy and dimension estimates. *Physics Letters A*, 128(67):369–373.

Greenland, S., Mansournia, M. A., and Joffe, M. (2022). To curb research misreporting, replace significance and confidence by compatibility. *Preventive Medicine*, 164:107127.

Grolemund, G. and Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25.

Gronau, Q. F., Singmann, H., and Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, 92(10):1–29.

Grosjean, P. (2025). *SciViews-R*. UMONS, MONS, Belgium.

Gruber, J. (2014). *The Markdown File Extension*. The Daring Fireball Company, LLC.

Grund, S., Robitzsch, A., and Luedtke, O. (2023). *mitml: Tools for Multiple Imputation in Multilevel Modeling*. R package version 0.4-5.

Hall, P. and Morton, S. C. (1993). On the estimation of entropy. *Ann. inst. Statist. Math.*, 45(1):69–88.

Hankin, R. K. S. (2007). Very large numbers in R: Introducing package Brobdingnag. *R News*, 7.

Harrell Jr, F. E. (2025). *Hmisc: Harrell Miscellaneous*. R package version 5.2-3.

Hartley, R. V. L. (1928). Transmission of information. *The Bell System Technical Journal*, page 535.

Hassin, R. and Trope, Y. (2000). Facing faces: Studies on the cognitive aspects of physiognomy. *Journal of Personality and Social Psychology*, 78(5):837–852.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition.

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6):223–233.

Henry, L. and Wickham, H. (2023). *lifecycle: Manage the Life Cycle of your Package Functions*. R package version 1.0.4.

Henry, L. and Wickham, H. (2024). *tidyselect: Select from a Set of Strings*. R package version 1.2.1.

Henry, L. and Wickham, H. (2025). *rlang: Functions for Base Types and Core R and 'Tidyverse' Features*. R package version 1.1.6.

Hester, J. and Bryan, J. (2024). *glue: Interpreted String Literals*. R package version 1.8.0.

Hester, J., Henry, L., Müller, K., Ushey, K., Wickham, H., and Chang, W. (2024). *withr: Run Code 'With' Temporarily Modified Global State*. R package version 3.0.2.

Hester, J., Wickham, H., and Csárdi, G. (2025). *fs: Cross-Platform File System Operations Based on 'libuv'*. R package version 1.6.6.

Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15:1593–1623.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.

Hughes, J. (2023). *sklarsomega: Measuring Agreement Using Sklar's Omega Coefficient*. R package version 3.0-2.

Hurault, G. (2021). *HuraultMisc: Guillem Hurault Functions' Library*. R package version 1.1.1.

Husserl, E. (1983). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy*. Springer. First book.

Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton University Press.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Ioannidis, J. P. A., Stanley, T. D., and Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal (London)*, 127(605):F236–F265.

Itti, L. and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research (Oxford)*, 49(10):1295–1306.

Ivanova, M. and Vaidyanathan, B. (2024). Surprise in science: A qualitative study. *Erkenntnis*.

Iverson, G. J., Lee, M. D., and Wagenmakers, E.-J. (2009). P(rep) misestimates the probability of replication. *Psychonomic Bulletin and Review*, 16(2):424.

Jaynes, E. T. (1957a). Information theory and statistical mechanics. i. *The Physical Review*, 106(4):620–630.

Jaynes, E. T. (1957b). Information theory and statistical mechanics. ii. *The Physical Review*, 108(2):171–190.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, ssc-4(3):227–241.

Jaynes, E. T. (2003). *Probability Theory. The Logic of Science*. Cambridge University Press.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society, Series A*, 186:453–461.

Johnson, A. R. (2025). *QuickJSR: Interface for the 'QuickJS-NG' Lightweight 'JavaScript' Engine*. R package version 1.8.0.

Johnson, S. G. (2008). *The NLopt nonlinear-optimization package*.

Jung, L. (2024). *scrutiny: Error Detection in Science*. R package version 0.5.0.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2):109–133.

Kant, I. (2018[1781]). *The Critique of Pure Reason*. CreateSpace Independent Publishing Platform. English translation by J. M. D. Meiklejohn.

Kantorovich, L. V. (1939). Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422.

Kay, M. (2024a). ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):414–424.

Kay, M. (2024b). *tidybayes: Tidy Data and Geoms for Bayesian Models*. R package version 3.0.7.

Kay, M. (2025). *ggdist: Visualizations of Distributions and Uncertainty*. R package version 3.3.3.

Killeen, P. R. (2005a). An alternative to null-hypothesis significance tests. *Psychological Science*, 16(5):345–353.

Killeen, P. R. (2005b). Replicability, confidence, and priors. *Psychological Science*, 16:1009–1012.

Kinder, D. R., Peters, M. D., Abelson, R. P., and Fiske, S. T. (1980). Presidential prototypes. *Political Behavior*, 2(4):315–337.

King, G. (1998). *Unifying Political Methodology - The Likelihood Theory of Statistical Inference*. The University Of Michigan Press.

King, G., Tomz, M., and Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 44(2):347–361.

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2 edition.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Lakatos, I. (1970). Falsification and the methodology of scientific research. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowledge*, pages 91–196. Cambridge University Press.

Lamport, L. (1986). *LATEX: a document preparation system*. Addison-Wesley Pub. Co.

Lang, M. (2017). checkmate: Fast argument checks for defensive R programming. *The R Journal*, 9(1):437–445.

Lang, M., Murdoch, D., and R Core Team (2024). *backports: Reimplementations of Functions Introduced Since R-3.0.0*. R package version 1.5.0.

Laplace, P. S. (1951). *A Philosophical Essay on Probabilities*. Dover Publications, 6 edition. English translation by F. W. Truscott and F. L. Emory.

Lenz, G. and Sahn, A. (2021). Achieving statistical significance with covariates and without transparency. *Political Analysis*, 29(3):356–369.

Lesne, A., Blanc, J.-L., and Pezard, L. (2009). Entropy estimation of very short symbolic sequences. *Phys. Rev. E*, 79:046208.

Levy, R. and Mislevy, R. J. (2020). *Bayesian Psychometric Modeling*. Chapman and Hall.

Ligges, U. and Mächler, M. (2003). Scatterplot3d - an R package for visualizing multivariate data. *Journal of Statistical Software*, 8(11):1–20.

Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145 – 151.

Loftus, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology*, 36:102–105.

Loftus, G. R. (2001). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*.

Lovakov, A. and Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*.

Lugg, A. (1979). Laudan and the problem-solving approach to scientific progress and rationality. *Phil. Soc. Sci*, 9:466–474.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., and Makowski, D. (2020). Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, 5(53):2445.

Lüdecke, D., Waggoner, P., and Makowski, D. (2019). insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, 4(38):1412.

Macdonald, R. R. (2005). Why replication probabilities depend on prior probability distributions. *Psychological Science*, 16:1006–1008.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2024). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.8 — For new features, see the 'NEWS' and the 'Changelog' file in the package source).

Mahalanobis, P. C. (2018[1936]). On the generalized distance in statistics. *Sankhya, Series A*, 80:S1–S7.

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., and Lüdecke, D. (2019a). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, 10:2767.

Makowski, D., Ben-Shachar, M. S., and Lüdecke, D. (2019b). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, 4(40):1541.

Maraun, M. and Gabriel, S. (2010). Killeen's (2005) p rep coefficient: logical and mathematical problems. *Psychological Methods*, 15(2):182–191.

Masi, M. (2023). Quantum indeterminism, free will, and self-causation. *Journal of Consciousness Studies*, 30(5):32–56.

Mayo, D. G. (1996). *Error and the Growth of Experimental Knowledge*. The University of Chicago Press.

McElreath, R. (2019). *Statistical Rethinking: A Bayesian Course with Examples in R*. Chapman and Hall, 2 edition.

McShane, B. B., Gal, D., Gelman, A., Robert, C., and Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, 73(sup1):235–245.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2):103–115.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46:806–834.

Mello, P. R., Quartin, M., Schaefer, B. M., and Schosser, B. (2025). On the full non-Gaussian surprise statistic and the cosmological concordance between DESI, SDSS and Pantheon. *The Open Journal of Astrophysics*, 8.

Microsoft (2024). *Windows 11*. x64, build 26100.

Microsoft and Weston, S. (2022). *foreach: Provides Foreach Looping Construct*. R package version 1.5.2.

Mielniczuk, J. (2022). Information theoretic methods for variable selection—a review. *Entropy (Basel, Switzerland)*, 24(8):1079.

Miller, D. (1974). Popper's qualitative theory of verisimilitude. *The British Journal for the Philosophy of Science*, 25(2):166–177.

Modirshanechi, A., Brea, J., and Gerstner, W. (2021). Surprise: a unified theory and experimental predictions.

Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. Imprimerie Royale.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology*, 12(2):317–329.

Müller, K. (2023). *hms: Pretty Time of Day*. R package version 1.1.3.

Müller, K. and Wickham, H. (2025a). *pillar: Coloured Formatting for Columns*. R package version 1.11.0.

Müller, K. and Wickham, H. (2025b). *tibble: Simple Data Frames*. R package version 3.3.0.

Neuwirth, E. (2022). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-3.

Nielsen, F. (2019). On the Jensen–Shannon symmetrization of distances relying on abstract means. *entropy*, 21:485.

Nyquist, H. (1924). Certain factors affecting telegraph speed. *The Bell System Technical Journal*, page 324.

Nyquist, H. (1928). Certain topics in telegraph transmission theory. *A.I.E.E. Trans.*, 47:617.

O'Hara-Wild, M., Kay, M., Hayes, A., and Hyndman, R. (2024). *distributional: Vectorised Probability Distributions*. R package version 0.5.0.

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between JSON data and R objects. *arXiv:1403.2805 [stat.CO]*.

Ooms, J. (2025a). *curl: A Modern and Flexible Web Client for R*. R package version 6.4.0.

Ooms, J. (2025b). *V8: Embedded JavaScript and WebAssembly Engine for R*. R package version 7.0.0.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *The American Political Science Review*, 115(3):999–1015.

Parzen, E. (1962). On estimation of a probability density and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076.

Pasek, J. (2025). *weights: Weighting and Weighted Statistics*. R package version 1.1.2.

Patil, I., Makowski, D., Ben-Shachar, M. S., Wiernik, B. M., Bacher, E., and Lüdecke, D. (2022). datawizard: An R package for easy data preparation and statistical transformations. *Journal of Open Source Software*, 7(78):4684.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347):240–242.

Pedersen, T. L., Nicolae, B., and François, R. (2024). *farver: High Performance Colour Space Manipulation*. R package version 2.1.2.

Peng, R. D. (2006). Interacting with data using the filehash package. *R News*, 6(4):19–24.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., and Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7):770–780.

Pennycook, G. and Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188:39–50.

Petersen, M. B. and Laustsen, L. (2019). Upper-body strength and political egalitarianism: Twelve conceptual replications. *Political Psychology*, 40(2):375–394.

Pinheiro, J., Bates, D., and R Core Team (2025). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-167.

Plate, T. and Heiberger, R. (2024). *abind: Combine Multidimensional Arrays*. R package version 1.4-8.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.

Popper, K. (2002[1959]). *The Logic of Scientific Discovery*. Routledge, New York.

Posit Team (2025). *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA.

Pérez-Cruz, F. (2019). Kullback-Leibler divergence estimation of continuous distributions. *IEEE International Symposium on Information Theory*.

Quartagno, M. and Carpenter, J. (2023). *jomo: A package for Multilevel Joint Modelling Multiple Imputation*.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Core team, Hester, J., and Csárdi, G. (2021). *urlchecker: Run CRAN URL Checks from Older R Versions*. R package version 1.0.1.

Rafi, Z. and Greenland, S. (2020). Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Medical Research Methodology*, 20(1):244.

Revolution Analytics and Weston, S. (2022). *iterators: Provides Iterator Construct*. R package version 1.0.14.

Rinker, T. W. and Kurkiewicz, D. (2018). *pacman: Package Management for R*. Buffalo, New York. version 0.5.0.

Ritchie, S. (2020). *Science Fictions: How Fraud, Bias, Negligence, and Hype Undermine the Search for Truth*. Metropolitan Books.

Robinson, D., Hayes, A., and Couch, S. (2025). *broom: Convert Statistical Objects into Tidy Tibbles*. R package version 1.0.8.

Rokeach, M. (1973). *The nature of human values*. The Free Press.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates for a probability density function. *Annals of Mathematical Statistics*, 27(3):832–837.

Rosenthal, R. (1979). File drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3):638–641.

Rényi, A. (1961). On measures of entropy and information. *Berkeley Symposium on Mathematical Statistics and Probability*, 4.1:547–561.

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York.

Savage, L. J. (2017). *The Foundations of Statistics*. Dover Publications, 2 edition.

Schürmann, T. and Grassberger, P. (1996). Entropy estimation of symbol sequences. *Chaos*, 6(3):414–427.

Seehars, S., Amara, A., Refregier, A., Paranjape, A., and Akaret, J. (2014). Information gains from cosmic microwave background experiments. *Physical Review. D, Particles, fields, gravitation, and cosmology*, 90(2):023533.

Serlin, R. C. (2010). Regarding p rep: comment prompted by Iverson, Wagenmakers, and Lee (2010); Lecoutre, Lecoutre, and Poitevineau (2010); and Maraun and Gabriel (2010). *Psychological Methods*, 15(2):203–208.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 623–656.

Shannon, C. E. (1964[1949]). The mathematical theory of communication. In Shannon, C. E. and Weaver, W., editors, *The Mathematical Theory of Communication*, pages 29–125. University of Illinois Press, Urbana.

Sharpsteen, C. and Bracken, C. (2023). *tikzDevice: R Graphics Output in LaTeX Format*. R package version 0.12.6.

Shaughnessey, J. J., Zechmeister, E. B., and Zechmeister, J. S. (2015). *Research Methods in Psychology*. McGraw-Hill, 10 edition. International Edition.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11):1359–1366.

Sklyar, O., Eddelbuettel, D., and Francois, R. (2025). *inline: Functions to Inline C, C++, Fortran Function Calls from R*. R package version 0.3.21.

Soetaert, K. (2024). *shape: Functions for Plotting Graphical Shapes, Colors*. R package version 1.4.6.1.

Spinu, V. (2024). *timechange: Efficient Manipulation of Date-Times*. R package version 0.3.0.

Stan Development Team (2020). StanHeaders: Headers for the R interface to Stan. R package version 2.21.0-6.

Stan Development Team (2025a). RStan: the R interface to Stan. R package version 2.32.7.

Stan Development Team (2025b). *Stan Reference Manual*. Version 2.37.0 (2 September 2025).

Steiger, J. H. and Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A., and Steiger, J. H., editors, *What if there were no signiWcance tests?*, pages 221–257. Lawrence Erlbaum Associates.

Sterling, J., Jost, J. T., and Hardin, C. D. (2019). Liberal and conservative representations of the good society: A (social) structural topic modeling approach. *SAGE Open*, pages 1–13.

Stock, J. and Watson, M. W. (2019). *Introduction to Econometrics*. Pearson Education Limited, 4 edition. Global Edition.

The LaTeX Project (2025). *The LaTeX Project*. The LaTeX Project.

Therneau, T. and Atkinson, B. (2025). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1.24.

Therneau, T. M. (2024). *A Package for Survival Analysis in R*. R package version 3.8-3.

Thompson, B. (1996). Aera editorial policies regarding statistical significance testing: three suggested reforms. *Educational Researcher*, 25:26–30.

Thành, H. T. and Knuth, D. E. (2025). *MikTeX*. MiKTeX-pdfTeX version 4.21 (MiKTeX version 25.4).

Tierney, L. (2024). *codetools: Code Analysis Tools for R*. R package version 0.2-20.

Todorov, A., Mandisodza, A. N., Goren, A., and Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(10):1623–1626.

Todorov, A. and Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology*, 39(6):549–562.

Toshkov, D. (2016). *Research Design in Political Science*. Palgrave.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science*, 6:100–116.

Urbanek, S. (2015). *base64enc: Tools for base64 encoding*. R package version 0.1-3.

Ushey, K., Allaire, J. J., Wickham, H., and Ritchie, G. (2024). *rstudioapi: Safely Access the RStudio API*. R package version 0.17.1.

Vaidyanathan, R., Xie, Y., Allaire, J., Cheng, J., Sievert, C., and Russell, K. (2023). *htmlwidgets: HTML Widgets for R*. R package version 1.6.4.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3):1–67.

van den Boogaart, K. G. (2023). *tensorA: Advanced Tensor Arithmetic with Named Indices*. R package version 0.36.2.1.

Vaseršteĭn, L. N. (1969). Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredači Informacii*, 5(3):64–72.

Vaughan, D. (2025). *tzdb: Time Zone Database Information*. R package version 0.5.0.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., and Gelman, A. (2024). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.8.0.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis*, 16(2):667–718.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, 4 edition. ISBN 0-387-95457-0.

Vicente, K. J. and Torenvliet, G. L. (2000). The earth is spherical ($p < 0.05$): alternative methods of statistical inference. *Theoretical Issues in Ergonomics Science*, 1(3):248–271.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60(3):158–189.

Wand, M. (2025). *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23-26.

Wang, Z., Zhang, H., Chen, J., and Chen, H. (2024). An effective framework for measuring the novelty of scientific articles through integrated topic modeling and cloud model. *Journal of Informetrics*, 18(4):101587.

Warnes, G. R., Bolker, B., Lumley, T., Magnusson, A., Venables, B., Ryodan, G., and Moeller, S. (2023). *gtools: Various R Programming Tools*. R package version 3.9.5.

Warnes, G. R., Gorjanc, G., Magnusson, A., Andronic, L., Rogers, J., MacQueen, D., and Korosec, A. (2024). *gdata: Various R Programming Tools for Data Manipulation*. R package version 3.0.1.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. (2021). *ellipsis: Tools for Working with ...* R package version 0.3.2.

Wickham, H. (2023a). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0.

Wickham, H. (2023b). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wickham, H., Bryan, J., Barrett, M., and Teucher, A. (2024a). *usethis: Automate Package and Project Setup*. R package version 3.1.0.

Wickham, H., Chang, W., Flight, R., Müller, K., and Hester, J. (2025a). *sessioninfo: R Session Information*. R package version 1.2.3.

Wickham, H., Chang, W., Hester, J., and Henry, L. (2024b). *pkgload: Simulate Package Installation and Attach*. R package version 1.4.0.

Wickham, H., Chang, W., Luraschi, J., and Mastny, T. (2024c). *profvis: Interactive Visualizations for Profiling R Code*. R package version 0.4.0.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023a). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4.

Wickham, H. and Henry, L. (2025). *purrr: Functional Programming Tools*. R package version 1.1.0.

Wickham, H., Henry, L., and Vaughan, D. (2023b). *vctrs: Vector Helpers*. R package version 0.6.5.

Wickham, H., Hester, J., and Bryan, J. (2024d). *readr: Read Rectangular Text Data*. R package version 2.1.5.

Wickham, H., Hester, J., Chang, W., and Bryan, J. (2022). *devtools: Tools to Make Developing R Packages Easier*. R package version 2.4.5.

Wickham, H., Hester, J., Chang, W., Müller, K., and Cook, D. (2021). *memoise: 'Memoisation' of Functions*. R package version 2.0.1.

Wickham, H., Hester, J., and Csárdi, G. (2025b). *pkgbuild: Find Tools Needed to Build R Packages*. R package version 1.4.8.

Wickham, H., Kuhn, M., and Vaughan, D. (2025c). *generics: Common S3 Generics not Provided by Base R Methods Related to Model Fitting*. R package version 0.1.4.

Wickham, H. and Pedersen, T. L. (2024). *gtable: Arrange 'Grobs' in Tables*. R package version 0.3.6.

Wickham, H., Pedersen, T. L., and Seidel, D. (2025d). *scales: Scale Functions for Visualization*. R package version 1.4.0.

Wickham, H., Vaughan, D., and Girlich, M. (2024e). *tidyr: Tidy Messy Data*. R package version 1.3.1.

Wickham, H. and Xie, Y. (2025). *evaluate: Parsing and Evaluation Tools that Provide More Details than the Default*. R package version 1.0.4.

Wiener, N. (2019[1961]). *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press, 2 edition. Reissue of the 1961 second edition.

Winston, J. S., Strange, B., O'Doherty, J., and Dolan, R. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5(3):277–283.

Wong, A. K. C. and You, M. (1985). Entropy and distance of random graphs with application to structural pattern recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-7(5):599–609.

Wooldridge, J. M. (2019). *Introductory Econometrics*. Cengage, 7 edition.

Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. In Stodden, V., Leisch, F., and Peng, R. D., editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2 edition. ISBN 978-1498716963.

Xie, Y. (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida.

Xie, Y. (2025a). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.43.

Xie, Y. (2025b). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.50.

Xie, Y. (2025c). *mime: Map Filenames to MIME Types*. R package version 0.13.

Xie, Y. (2025d). *xfun: Supporting Functions for Packages Maintained by 'Yihui Xie'*. R package version 0.52.

Zeileis, A. and Croissant, Y. (2010). Extended model formulas in R: Multiple parts and multiple responses. *Journal of Statistical Software*, 34(1):1–13.

Zeileis, A., Fisher, J. C., Hornik, K., Ihaka, R., McWhite, C. D., Murrell, P., Stauffer, R., and Wilke, C. O. (2020). colorspace: A toolbox for manipulating and assessing colors and palettes. *Journal of Statistical Software*, 96(1):1–49.

Zeileis, A., Hornik, K., and Murrell, P. (2009). Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics & Data Analysis*, 53(9):3259–3270.

Zhao, J. H. and Schafer, J. L. (2023). *pan: Multiple imputation for multivariate panel or clustered data*. R package version 1.9.

Zhu, H. (2024). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.4.0.

8 Appendix

This appendix contains the specifications for the software used in this paper (Section 8.1) and an R function implementing the KDE plug-in estimator that was used to compute the Jensen-Shannon distance from MCMC samples (Section 8.2) as demonstrated in section (4). Besides this appendix, supplementary materials, including replication files, are made freely available online on the OSF project page dedicated to this paper (DOI: 10.17605/OSF.IO/GQ6C8).

8.1 Software Specifications

This document was written with *markdown* (Gruber, 2014) and LaTeX(Lamport, 1986; The LaTeX Project, 2025), using pdfTeX(v3.141592653-2.6-1.40.2; MiKTeX[v25.4]; Thành and Knuth, 2025). Compiled on a Windows 11 x64 (build 26200) operating system (OS; Microsoft, 2024) on the 2026-02-07 15:20 CET (Timezone: Europe/Copenhagen), the machine has 16 cores available, and the range of doubles is 2.23×10^{-308} - $1.8 \times 10^{+308}$.

All data processing and analyses were conducted on the aforementioned software using the R programming language (v4.4.3 [2025-02-28 ucrt]; R Core Team, 2025) in the RStudio IDE (2025.5.1.513; Posit Team, 2025), which besides the base R packages by R Core Team (2025) used the following packages (including their dependencies, which were *not* necessarily used): `abind` (v1.4.8; Plate and Heiberger, 2024), `arrayhelpers` (v1.1.0; Beleites, 2020), `backports` (v1.5.0; Lang et al., 2024), `base64enc` (v0.1.3; Urbanek, 2015), `bayestestR` (v0.16.1; Makowski et al., 2019b), `bayesplot` (v1.13.0; Gabry et al., 2019), `bookdown` (v0.43; Xie, 2016, 2025a), `boot` (v1.3.31; Canty and Ripley, 2024; Davison and Hinkley, 1997), `bridgesampling` (v1.1.2; Gronau et al., 2020), `brms` (v2.22.0; Bürkner, 2017, 2018), `Brobdingnag` (v1.2.9; Hankin, 2007), `broom` (v1.0.8; Robinson et al., 2025), `cachem` (v1.1.0; Chang, 2024a), `checkmate` (v2.3.3; Lang, 2017), `cli` (v3.6.5; Csárdi, 2025), `cluster` (v2.1.8; Maechler et al., 2024), `coda` (v0.19.4.1; Plummer et al., 2006), `codetools` (v0.2.20; Tierney, 2024), `colorspace` (v2.1.1; Zeileis et al., 2009, 2020), `curl` (v6.4.0; Ooms, 2025a), `data.table` (v1.17.8; Barrett et al., 2025), `datawizard` (v1.1.0; Patil et al., 2022), `devtools` (v2.4.5; Wickham et al., 2022), `digest` (v0.6.37; Eddelbuettel, 2024), `distributional` (v0.5.0; O'Hara-Wild et al., 2024), `dplyr` (v1.1.4; Wickham et al., 2023a), `effectsize` (v1.0.1; Ben-Shachar et al., 2020), `ellipsis` (v0.3.2; Wickham, 2021), `evaluate` (v1.0.4; Wickham and Xie, 2025), `farver` (v2.1.2; Pedersen et al., 2024), `fastmap` (v1.2.0; Chang, 2024b), `filehash` (v2.4.6; Peng, 2006), `forcats` (v1.0.0; Wickham, 2023a), `foreach` (v1.5.2; Microsoft and Weston, 2022), `foreign` (v1.5.2; Microsoft and Weston, 2022), `Formula` (v1.2.5; Zeileis and Croissant, 2010), `fs` (v1.6.6; Hester et al., 2025), `gdata` (v3.0.1; Warnes et al., 2024), `generics` (v0.1.4; Wickham et al., 2025c), `ggdist` (v3.3.3; Kay, 2024a, 2025), `ggplot2` (v3.5.2; Wickham, 2016), `glmnet` (v4.1.9; Friedman et al., 2010), `glue` (v1.8.0; Hester and Bryan, 2024), `gridExtra` (v2.3; Auguie, 2017), `gtable` (v0.3.6; Wickham and Pedersen, 2024), `gtools` (v3.9.5; Warnes et al., 2023), `Hmisc` (v5.2.3; Harrell Jr, 2025), `hms` (v1.1.3; Müller, 2023), `htmlTable` (v2.4.3; Gordon et al., 2024), `htmltools` (v0.5.8.1; Cheng et al., 2024), `htmlwidgets` (v1.6.4; Vaidyanathan et al., 2023), `httpuv` (v1.6.16; Cheng et al., 2025), `HuraultMisc` (v1.1.1; Hurault, 2021), `inline` (v0.3.21; Sklyar et al., 2025), `insight` (v1.3.1; Lüdecke et al., 2019), `iterators` (v1.0.14; Revolution Analytics and Weston, 2022), `jomo` (v2.7.6; Quartagno and Carpenter, 2023), `jsonlite` (v2.0.0; Ooms, 2014), `kableExtra` (v1.4.0; Zhu, 2024), `KernSmooth` (v2.23.26; Wand, 2025), `knitr` (v1.50; Xie, 2014, 2015, 2025b), `later` (v1.4.2; Chang et al., 2025b), `lattice` (v0.22.6; Sarkar, 2008), `lifecycle` (v1.0.4; Henry and Wickham, 2023), `lme4` (v1.1.37; Bates et al., 2015), loo (v2.8.0; Vehtari et al., 2024), lubridate (v1.9.4; Grolemund and Wickham, 2011), magrittr (v2.0.3; Bache and Wickham, 2022), MASS (v7.3.64; Venables and Ripley, 2002), Matrix (v1.7.2; Bates et al., 2025), matrixStats (v1.5.0; Bengtsson, 2025), memoise (v2.0.1; Wickham et al., 2021), mice (v3.18.0; van Buuren and Groothuis-Oudshoorn, 2011), mime (v0.13; Xie, 2025c), miniUI (v0.1.2; Cheng, 2025a), minqa (v1.2.8; Bates et al., 2024), mitml (v0.4.5; Grund et al., 2023), mvtnorm (v1.3.3; Genz and Bretz, 2009), nlme (v3.1.167; Pinheiro et al., 2025), nloptr (v2.2.1; Johnson, 2008), nnet (v7.3.20; Venables and Ripley, 2002), pacman (v0.5.1; Rinker and Kurkiewicz, 2018), pan (v1.9; Zhao and Schafer, 2023), parameters (v0.27.0; Lüdecke et al., 2020), pillar (v1.11.0; Müller and Wickham, 2025a), pkgbuild (v1.4.8; Wickham et al., 2025b), pkgconfig (v2.0.3; Csárdi, 2019), pkgload (v1.4.0; Wickham et al., 2024b), Polychrome (v1.5.4; Coombes et al., 2019), posterior (v1.6.1; Bürkner et al., 2025), profvis (v0.4.0; Wickham et al., 2024c), promises (v1.3.3; Cheng, 2025b), purrr (v1.1.0; Wickham and Henry, 2025), pwr (v1.3.0; Champely, 2020), QuickJSR (v1.8.0; Johnson, 2025), R6 (v2.6.1; Chang, 2025), rbibutils (v2.3; Boshnakov and Putman, 2024), RColorBrewer (v1.1.3; Neuwirth, 2022), Rcpp (v1.1.0; Eddelbuettel, 2013; Eddelbuettel and Balamuta, 2018; Eddelbuettel and François, 2011; Eddelbuettel et al., 2025), RcppParallel (v5.1.10; Allaire et al., 2025), Rdpack (v2.6.4; Boshnakov, 2025), readr (v2.1.5; Wickham et al., 2024d), reformulas (v0.4.1; Bolker, 2025), remotes (v2.5.0; Csárdi et al., 2024), rlang (v1.1.6; Henry and Wickham, 2025), rmarkdown (v2.29; Allaire et al., 2024), rpart (v4.1.24; Therneau and Atkinson, 2025), rstan (v2.32.7; Stan Development Team, 2025a), rstantools (v2.4.0; Gabry et al., 2024), rstudioapi (v0.17.1; Ushey et al., 2024), scales (v1.4.0; Wickham et al., 2025d), scatterplot3d (v0.3.44; Ligges and Mächler, 2003), scrutiny (v0.5.0; Jung, 2024), sessioninfo (v1.2.3; Wickham et al., 2025a), shape (v1.4.6.1; Soetaert, 2024), shiny (v1.11.1; Chang et al., 2025a), StanHeaders (v2.32.10; Stan Development Team, 2020), stringi (v1.8.7; Gagolewski, 2022), stringr (v1.5.1; Wickham, 2023b), survival (v3.8.3; Therneau, 2024), svUnit (v1.0.6; Grosjean, 2025), tensorA (v0.36.2.1; van den Boogaart, 2023), tibble (v3.3.0; Müller and Wickham, 2025b), tidybayes (v3.0.7; Kay, 2024b), tidyr (v1.3.1; Wickham et al., 2024e), tidyselect (v1.2.1; Henry and Wickham, 2024), tidyverse (v2.0.0; Wickham et al., 2019), tikzDevice (v0.12.6; Sharpsteen and Bracken, 2023), timechange (v0.3.0; Spinu, 2024), tzdb (v0.5.0; Vaughan, 2025), urlchecker (v1.0.1; R Core team et al., 2021), usethis (v3.1.0; Wickham et al., 2024a), V8 (v7.0.0; Ooms, 2025b), vctrs (v0.6.5; Wickham et al., 2023b), weights (v1.1.2; Pasek, 2025), withr (v3.0.2; Hester et al., 2024), xfun (v0.52; Xie, 2025d), and xtable (v1.8.4; Dahl et al., 2019). For the random number generator used by MCMC, the seed (i.e., 896663432) was exogenously predetermined by the random R package (v0.2.6; Eddelbuettel, 2017).

8.2 Jensen-Shannon distance function

An implementation of the Jensen-Shannon distance function using R (v4.4.3 [2025-02-28 ucrt]; R Core Team, 2025), and an example of its application in relation to the posterior and prior from Bayesian inference in relation to empirical data using the brms R package (v2.22.0; Bürkner, 2017, 2018), is provided in the code below:

```
# Function for computing the Jensen-Shannon (JS) distance/divergence based on MCMC samples from the posterior and prior distributions using default parameter values
#
# @param p An n-dimensional vector of distributional samples
```

```r
# @param q An n-dimensional vector of distributional samples
# @param n An integer indicating the length of the grid used for estimating the densities
# @param epsilon A small, positive real-valued scalar used to avoid instances of division by zero
# @param base A small, positive real-valued scalar used as the base of the logarithmic function
# @param type A character that determines whether the output is the JS distance or divergence
# @return A non-negative real-valued scalar reflecting the estimated discrepancy between the distributions

# install.packages("KernSmooth")
JSD <- function(p = NULL, q = NULL, n = 10000, epsilon = 1e-05, base = 2, type = c("Distance", "Divergence")){
  # Parameter checks
  stopifnot(length(p)==length(q))
  stopifnot(is.numeric(p))
  stopifnot(is.numeric(q))
  stopifnot(is.numeric(n))
  stopifnot(n > 1)
  stopifnot(is.numeric(epsilon))
  stopifnot(epsilon > 0)
  stopifnot(is.numeric(base))
  stopifnot(base > 0)
  type <- match.arg(type)

  # Define a common grid
  grid <- seq(min(c(p, q)), max(c(p, q)), length.out = n)

  # Use Kernsmooth::bkde to estimate densities on the same grid if available, otherwise use stats::density
  if(requireNamespace("KernSmooth", quietly = TRUE)){
    require(KernSmooth)
    p_density <- KernSmooth::bkde(p, gridsize = n, range.x = c(min(c(p, q)), max(c(p, q))))$y
    q_density <- KernSmooth::bkde(q, gridsize = n, range.x = c(min(c(p, q)), max(c(p, q))))$y
  }
  else{
    p_density <- stats::density(p, bw = "SJ", kernel = "gaussian", from = min(grid), to = max(grid), n = length(grid))$y
```

```
    q_density <- stats::density(q, bw = "SJ", kernel = "gaussian", from = min(grid)
        , to = max(grid), n = length(grid))$y
  }

  # Avoid any zeros
  p_density <- pmax(p_density, epsilon)
  q_density <- pmax(q_density, epsilon)

  # Ensure normalized densities
  dx <- diff(grid)[1]
  p_density <- p_density / sum(p_density * dx)
  q_density <- q_density / sum(q_density * dx)

  # Compute mixture
  m <- 0.5 * (p_density + q_density)

  # Compute JSD
  jensen_shannon_divergence <- 0.5 * sum(p_density * log(p_density / m, base = base
      ) * dx) +
        0.5 * sum(q_density * log(q_density / m, base = base) * dx)

  # Return JSD
  if(type == "Distance"){
    jensen_shannon_distance <- sqrt(jensen_shannon_divergence)
    return(jensen_shannon_distance)
  }
  else{
    return(jensen_shannon_divergence)
  }
}
```

```
# Example
# install.packages("tidyverse")
library("tidyverse")

# Load data
# install.packages("devtools")
# library("devtools")
# devtools::install_github("kosukeimai/qss-package", build_vignettes = TRUE)
data(face, package = "qss")
df <- face; rm(face)

# Recode data
```

```r
df <- df %>%
  mutate(
  d.share = d.votes / (d.votes + r.votes),
  r.share = r.votes / (d.votes + r.votes),
  d.diff.share = d.share - r.share,
  d.diff.share = standardize(d.diff.share),
  d.comp = standardize(d.comp)
) %>%
  select(d.comp, d.diff.share) %>%
  rename(
    democratic_competence = d.comp,
    democratic_relative_support = d.diff.share
  )

# Standardize data matrix
# install.packages("datawizard")
library("datawizard")
df <- apply(df, 2, datawizard::standardize) %>% as.data.frame()

# Specify MCMC
# install.packages("parallel")
library("parallel")
posterior_samples <- 4000
chains <- cores <- parallel::detectCores()-1
warmup <- 1000
iter <- ceiling((posterior_samples / chains) + warmup)
posterior_samples <- (iter - warmup)*chains

# Fit BGLM
# install.packages("brms")
library("brms")
democratic_relative_support_competence_fit <- brms::brm(
  democratic_relative_support ~ 0 + democratic_competence,
  data = df,
  prior = c(
    brms::set_prior("normal(0, 0.1)", class = "b", lb = -1, ub = 1),
    brms::set_prior("normal(1, 0.1)", class = "sigma", lb = 0, ub = 1)
  ),
  family = gaussian(
    link = "identity"
  ),
  sample_prior = "yes",
```

```
  cores = cores,
  chains = chains,
  seed = 896663432,
  iter = iter,
  warmup = warmup
)
# summary(democratic_relative_support_competence_fit)

# Extract draws
democratic_relative_support_competence_draws <- brms::as_draws_df(
    ↪ democratic_relative_support_competence_fit)
b_democratic_competence_posterior <-
    ↪ democratic_relative_support_competence_draws$b_democratic_competence
b_democratic_competence_prior <-
    ↪ democratic_relative_support_competence_draws$prior_b

# Compute JSKL divergence
JSD(b_democratic_competence_prior, b_democratic_competence_posterior)
```