Zero-sum reasoning in information selection

Toby D. Pilditch[1,2], Alice Liefgreen[1], and David Lagnado[1]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, London, WC1H 0AP, UK
[2]University of Oxford, School of Geography and the Environment, South Parks Road, Oxford, OX1 3QY, UK

Abstract

Recent research (Pilditch, Fenton, & Lagnado, 2019) shows that people are susceptible to zero-sum thinking in evidence evaluation, where they dismiss or underweight the probative value of evidence that is equally predicted by multiple independent hypotheses. But such an assumption is only valid when explanations are mutually exclusive and exhaustive. The present work extends these findings by looking at the context of information selection, and the decisional consequences of the zero-sum fallacy. It uses an information metric to quantify the cost of the error in terms of overlooked information.

Keywords: zero-sum; evidential reasoning; probabilistic reasoning; Bayesian Networks; belief updating

Introduction

When reasoning under uncertainty, the search and selection of evidence is fundamental to accurate and efficient prediction and diagnosis. Whether in formal investigative domains such as medical diagnosis, forensics, or intelligence gathering, or in everyday reasoning, we often have to search out information to make inferences about a target hypothesis (e.g. which test to conduct? Which source to query? Etc.). To address these questions, reasoners must consider the prospective "value" or *information* provided by new evidence. These estimates are often fraught with biases and errors (e.g. Jones & Sugden, 2001; Nelson, McKenzie, Cottrell & Sejnowski, 2010; Slowiaczek, Klayman, Sherman & Skov, 1992) making accurate choice of what evidence to gather a non-trivial task for lay reasoners.

In the present work, we explore the question of evidence selection in the context of a novel evidential reasoning fallacy, the zero-sum error (Pilditch, Fenton, & Lagnado, 2019), where reasoners assume that evidence which is equally predicted by multiple alternative hypotheses is non-probative. We explore whether this error also drives similar errors in information choice, in particular whether it leads to people overlooking the most useful evidential tests. We explore the mechanisms that might underpin this reasoning fallacy. Furthermore, we highlight the methodological and theoretical value of incorporating information measures into our understanding of how reasoners navigate more complex reasoning structures.

The Zero-sum fallacy

When reasoning about evidence that is equally predicted by two independent explanations, lay reasoners tend to assume that this evidence offers no support to either hypothesis, because it does not discriminate between them (Pilditch, Fenton, & Lagnado, 2019). However, this assumption is only applicable when the explanations are both mutually exclusive and exhaustive (i.e. exactly one of the explanations is true). In fact, given positive evidence, *both* explanations become more probable. Across a number of experiments, reasoners judged such evidence irrelevant to a target hypothesis, even when the inappropriateness of applying the assumptions of exclusivity and exhaustiveness was highlighted.

The posited mechanism behind this error was a fallacy of considering evidential support between hypotheses to be a "zero-sum" situation: one hypothesis may only gain support (i.e. become more probable) at the detriment of another. To elucidate, reasoners were inclined to dismiss a medical test that could not distinguish between 2 diseases – failing to consider that the positive test result could in fact make the patient having *both* diseases more probable.

Work on the zero-sum fallacy has so far looked at qualitative judgments of support. In building on this work, via the incorporation of alternative evidence options and a measure of the amount of overlooked information given a preference, we seek to quantify the *cost* of this error, and further uncover the mechanism underpinning it.

A Bayesian Framework

To further elucidate the nature of the zero-sum fallacy, and outline the foundational formalism upon which information in the context of reasoning under uncertainty may be built, we briefly highlight the role of Bayesian Networks (BNs; Pearl, 1988; 2009) in evidential reasoning.

BNs are directed acyclic graphs (DAGs) that provide a computational framework for modelling the strength of inferential relationships when reasoning under uncertainty. A BN is made up of nodes that represent the variables of interest, and directed arrows capturing probabilistic dependency relations between variables, quantified by conditional probability tables. The probabilities of the unknown nodes are normatively updated given new evidence using Bayes rule (Pearl, 1988). Consequently, BNs are used as a normative comparison against which human reasoning can be compared (e.g. Pilditch, Fenton, & Lagnado, 2019).

To explain in the zero-sum case, two possible hypotheses, each with their own prior probabilities are represented by separate, *independent* nodes (see H1 and H2 in Fig. 1). This reflects the acknowledged assumptions that the two hypotheses are neither mutually exclusive (i.e. both could be true) nor exhaustive (i.e. both could be false), and there are no direct causal links between them. Critical to the fallacy, however, is the conditional probability table (CPT) of the evidence that depends on both hypotheses (E1 in Fig. 1). Table 1 below provides an example of how likely the evidence is to be observed, given the possible states of the two hypotheses.

Table 1: Example conditional probability table for "common effect" evidence, given two possible causes, H1 and H2.

E	¬H1, ¬H2	H1, ¬H2	¬H1, H2	H1, H2
E = T	0.01	0.9	0.9	0.99
E = F	0.99	0.1	0.1	0.01

The two central columns of Table 1 represent the possibilities that participants making the zero-sum fallacy arguably focus on. More precisely, if one (falsely) assumes that only one of the two hypotheses is true (i.e. they are exclusive and exhaustive), then one is only considering two possibilities: the probability of E given H1 being true (P(E|H1,¬H2); center-left column) or given H2 being true (P(E|¬H1,H2); center-right column). Consequently, by adopting this narrow focus, the evidence appears to be *equally predicted* by each possibility (P(E|H1,¬H2) = P(E|¬H1,H2) = 0.9) suggests the evidence is non-probative.

Critically, this reasoning neglects two important possibilities: first, the fact that evidence *could* still occur when neither hypothesis is true (P(E|¬H1,¬H2) > 0) – i.e. the hypotheses are *not* exhaustive explanations of the evidence. Second, that not only is there the possibility that both hypotheses are true i.e. the hypotheses are not exclusive, but that when both *are in fact true*, this results in an even greater probability of observing the evidence (i.e. P(E|H1,H2) > (P(E|H1,¬H2) | P(E|¬H1,H2))). Thus, when making the diagnostic inference from observed evidence to probable hypotheses, *both* H1 and H2 become more probable, given E.

Information Search

In the real world people are habitually required to *actively* seek and acquire information in order to make a decision, causal inference or judgement, and do not merely act as passive observers of their surroundings. Within the psychological literature, measures have been proposed to quantify the informative value of a piece of evidence and the exploration of people's information search behaviour in a variety of contexts (for an overview, see Nelson, 2008). Here we adopt the Kullback-Liebler Divergence (KL-D; Kullback & Liebler, 1951) as a quantitative measure of the expected informative value of different pieces of evidence given a defined probabilistic environment. KL-D is a form of relative entropy and assigns high informative value to evidence that reduces uncertainty the most, entailing the largest divergence between prior and posterior probability distributions (Nelson, 2008). Formally, it quantifies the subjective expected usefulness of evidence before the state of the evidence is known as:

$$KL(E_i) = \sum_{H_j} P(H_j|a_i) \log \frac{P(H_j|a_i)}{P(H_j)}$$

Where E_i is an item of evidence within a set $\{E_1, E_{2...}E_i\}$, H is a set of hypotheses, $\{H_1, H_{2...}H_j\}$ and a_i is a set of possible states of the evidence, $\{a_1, a_2, a_i\}$. This quantification enables not only the evaluation of whether people have a preference for evidence with the highest information value, but also allows for a quantitative measure of the amount of *overlooked* information (as a consequence of sub-optimal search behaviour). This approach directly addresses how violations of normative measures of the value of information relate to known violations of normative models of evidence evaluation such as the zero-sum fallacy. Or more informally, puts an explicit value on the cost of the error.

Present Work

As mentioned above, the goal of the present work is to investigate the zero-sum fallacy further, via the inclusion of information search. To do this we expand the previous zero-sum fallacy model (two hypotheses, H1 and H2, with a single, shared piece of evidence, E1) to include an alternative evidence option (E2) – only explainable by the target hypothesis.

In this way, the reasoning probe shifts from an explicit evaluation of whether E1 provides any support for H1, to a decision-making preference between two evidence items: E1, which has an alternative explanation H2 (and thus invites the zero-sum error), and E2, with no alternative cause represented in the model. To explore the possible influence of zero-sum thinking, and to quantify overlooked information costs, the general structure illustrated in Fig. 1 required populating with several different sets of parameters.

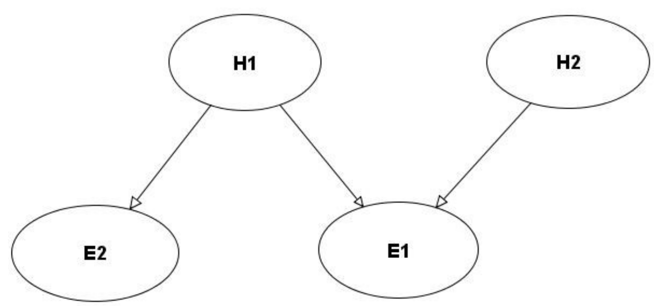

Figure 1. Graphical representation of BN Model.

Four sets of parameters were created (shown in Table 2), each incrementally differing from another, so as to determine the influence of various reasoning components. The prior probabilities of each hypothesis were manipulated as either both rare (P(H1) = P(H2) = .1), both common (P(H1) = P(H2) = .5), or unequal (P(H1) = .5, P(H2) = .1). In this way, the degree to which H2 is providing a "false positive" for E1 (i.e. another explanation for a positive, that is not the hypothesis of interest, H1) is manipulated. This is of interest to determine whether the zero-sum fallacy is based on the integration of this "false positive" probability (i.e. when H2 is more probable, the zero-sum fallacy is more prevalent), or solely on the *presence* of a possible alternative explanation. Further, the manipulation of 50/50 (or "common") priors can be used to assess whether participants will be more inclined to apply the false assumptions of mutual exclusivity and exhaustiveness that underpin the zero-sum fallacy. Lastly, if the manipulation of unequal priors (P(H1) = .5, P(H2) = .1) resulted in a *reduction* of zero-sum fallacy errors, it would be suggestive of participants using the relative rarity of H2 to *discount it* as an explanation of E1.

In addition, across these three sets, the likelihoods of E1 and E2 were held as unequal, in that E1 was more diagnostic of H1 than E2 (P(E1|H1,¬H2) = .9, vs P(E2|H1) = .6). However, one final parameter set was added in which (along with rare priors) these values were equal across E1 and E2 (P(E1|H1,¬H2) = P(E2|H1) = .8). It should be noted that in all these parameter sets, this results in E1 being the more informative evidence for determining H1, and thus selecting E2 comes at a cost of overlooked information. However, by manipulating the false positive rate of E2 as either high (P(E2|¬H1) = .2/.4), or low (P(E2|¬H1) = .01), we can manipulate between subjects a condition in which E1 is superior (the former), or inferior (the latter), to further determine sensitivity to the parameters underlying the fallacy.

This leads to several predictions: Firstly, there will be a general aversion to selecting E1 (i.e. the decision analogue of a zero-sum fallacy). Secondly, participants will be sensitive to parameter manipulations, such that when E2 is manipulated as more diagnostic (e.g. P(E2|¬H1) = .01 condition), aversion to E1 / preference for E2 will (correctly in this instance) increase. Conversely, when parameter manipulations in fact favour E1 (e.g. equal likelihoods parameter set) participants will (falsely) remain aversive to it.

Method

Participants 180 US participants were recruited and participated online through the Amazon Mechanical Turk platform. Participants were native English speakers (leading to 2 exclusions), with a mean age of 35.88 (SD = 10.5), and 90 participants identified as female. All participants gave informed consent, and were paid \$1.20 for their time (*Median* = 12.75 minutes, SD = 9.62).

Procedure & Design Participants were shown 4 scenarios in a randomized order. These scenarios all originated from the model structure of Fig. 1, to include a target hypothesis (H1), evidence that may inform on the hypothesis (E1), but may also be explainable by an alternative hypothesis (H2), and finally an alternative evidence item only dependent on H1, and not H2 (E2). The scenario contexts were an arson case (identifying an accelerant), a conservation case (tracking a target species), a medical diagnosis case (confirming a brain tumor), and a digital forensics case (identifying a cyberattack culprit).

Crucially, along with the structure of Fig. 1, contexts were also furnished within the text with sufficient parameter details to fully populate a Bayesian Network model of the scenario. These included the priors for each hypothesis (P(H1) and P(H2), the likelihoods for each evidence-hypothesis relationship (P(E1|H1,¬H2), P(E1|¬H1,H2), and P(E2|H1)), and false positives - P(E1|¬H1,¬H2) and P(E2|¬H1). The latter of these parameters (E2 false positive) was manipulated between subjects, as a method of shifting the balance of expected information between E1 and E2. The remaining parameters were deployed as 4 "sets" (see Table 1 below), each designed to test particular parameters trade-offs, and randomly allocated to scenario contexts.[1]

Table 2. Parameter sets, allocated across scenario contexts.

	Parameter Sets			
	RareP. EqL	RareP. UneqL	UneqP. UneqL	ComP. UneqL
P(H1)	.1	.1	.5	.5
P(H2)	.1	.1	.1	.5
P(E1|H1,¬H2)	.8	.9	.9	.9
P(E1|¬H1,H2)	.8	.9	.9	.9
P(E1|¬H1,¬H2)	.01	.01	.01	.01
P(E2|H1)	.8	.6	.6	.6
P(E2|¬H1)	.01 / .2	.01 / .4	.01 / .4	.01 / .4
Information				
KL(E1)*	0.12	0.135	0.27	0.06
KL(E2)	0.22/0.06	0.16/0.005	0.268/0.01	0.268/0.01
KL(E1 – E2)	-0.1/0.06	-0.026/0.13	0.002/0.25	-0.205/0.05

*Only takes into account H1

For each scenario, participants answered the following questions:

Priors: Participants were asked to provide the prior probabilities of H1 and H2 (i.e. *before* observing any evidence). Although participants had already been provided with prior probabilities for H1 and H2, by also eliciting these prior probabilities any participant-based assumptions could be incorporated into the models used for normative comparisons. More precisely, for each participant, elicited priors were used to outfit a Bayesian Network fitting the structure of Fig. 1 (and the remaining parameters drawn from the parameter set being tested), using the gRain package in R (Højsgaard, 2012). These individually fitted BNs (hereafter termed Behaviorally Informed Bayesian Networks; BIBNs) thus provided a fitted normative comparison for participant inferences on the participant by parameter set level. BIBNs were not only then used to generate predicted responses, but also to calculate the informative value (KL-D) of each item of evidence, given

[1] P(E|H1, H2), though not provided explicitly to participants, is based on an assumption of a noisyOR function (see Pearl, 1988), which is based on the reasonable assumption that causes H1 and H2 are independent.

that model – essential for calculating any forgone information.

Preference: Participants were then asked "Which test (evidence item) would you prefer, so as to best determine [H1]?" This qualitative judgment was forced choice [E1 / E2 / "They are the same."]

Confidence in preference: Following the qualitative evidence preference, participants were asked to provide a confidence in that preference ("How confident are you that your response is correct?" 0-100%).

Other DVs: Although posterior probability estimates for each evidence item ("Probability of [H1] ***only*** **given a positive [E1]**" 0 - 100%; "Probability of [H1] ***only*** **given a positive [E2]**" 0 - 100%), and open text reasoning responses were collected, for the sake of brevity, these results are not reported here.

Results

Using the JASP statistical software (JASP Team, 2018), Bayesian statistics were employed throughout[2].

Evidence Preferences

Overall, binomial tests comparing evidence preferences to chance (.33) found the evidence with a single possible cause (E2) to be preferred at a rate decisively greater than chance (.54, N = 712), $BF_{10} = 3.06 * 10^{26}$, whilst preferences for the evidence with two potential cause (E1) were no different than chance, (.35, N = 712), BF_{10} = 0.083, and preferences for "They are the same." occurred decisively less often than expected by chance (.11, N = 712), $BF_{10} = 4.59 * 10^{37}$. Further, a contingency table comparing observed to predicted preferences found decisive evidence for these preferences deviating from normative expectation (N = 1424), $BF_{10} = 1.196 * 10^{25}$. Importantly, there was a null influence of the potential confounds of scenario order (N = 712), BF_{10} = 0.109 , or scenario context (N = 712), $BF_{10} = 5.087 * 10^{-5}$.

In line with expectations, when the false positive rate of E2 was low (.01), and thus sensitivity was higher, then E2 was preferred substantially more often (and E1 less often) than when the false positive of E2 was high (N = 712), BF_{10} = 4.068.

Turning next to parameter sets (rows of Fig. 2), we break down the analysis for each set to determine a) the dominant participant preference, and b) whether this deviates from the normative predictions for that set. This split by parameter set is motivated by the potential sensitivity of participants to particular combinations of parameters (e.g. equal likelihoods, or unequal priors).

[2]All analyses assumed an uninformed prior. Bayes Factors (BFs), are interpreted as: 1 – 3 = anecdotal support; 3-10 = substantial; 10-30 = strong; 30-100 = very strong; >100 = decisive (Jeffreys, 1961). Conversely, Bayes Factors < .33 are considered substantial support for the *null* (Dienes, 2014).

Rare Priors, Equal Likelihoods. When both H1 and H2 priors were rare, and evidence likelihoods were equal, participants chose E2 at levels decisively above chance (.612, N = 178), $BF_{10} = 6.729 * 10^{11}$, and E1 significantly less than chance (.23, N = 178), BF_{10} = 5.565. This runs contrary to model predictions, where E1 is preferred decisively above chance level (.674, N = 178), $BF_{10} = 1.035 * 10^{18}$, and E2 at no different than chance (.326, N = 178), BF_{10} = 0.088. This is further corroborated by a contingency table analysis which finds decisive evidence for a deviation of participant choices from normative expectation (N = 356), $BF_{10} = 6.729 * 10^{11}$.

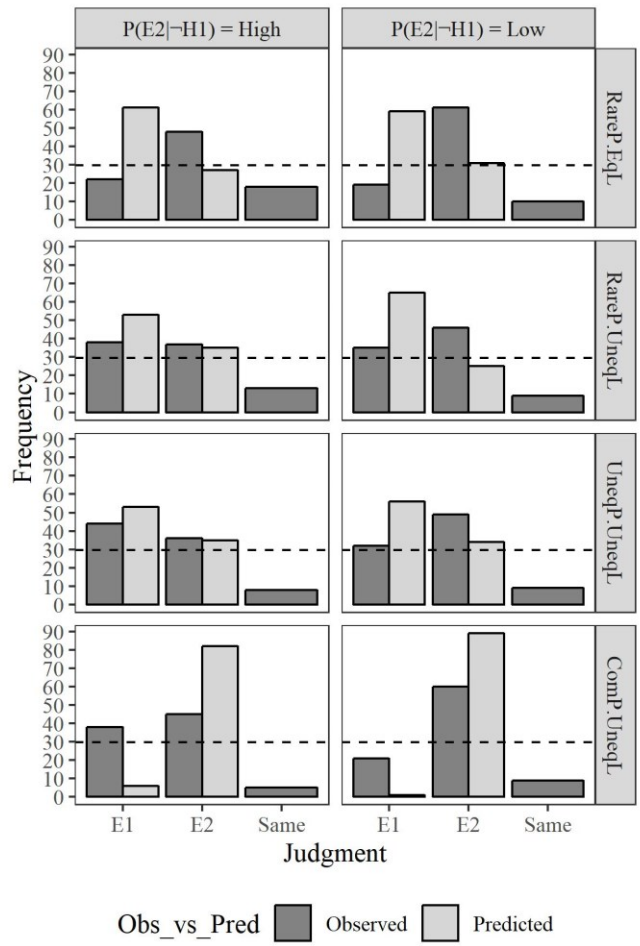

Figure 2. Evidence choice frequencies across parameter sets (rows) and condition (columns).

Rare Priors, Unequal Likelihoods. When priors are rare, and evidence likelihoods are unequal (E2 at .6, and E1 at .9), we again find the same pattern. Participants choose E2 at above chance levels (.466, N = 178), BF_{10} = 111.88, and E1 no different than chance (.41, N = 178), BF_{10} = 1.115. Once again, however, model predictions show the opposite pattern, with E1 choices above chance level (.663, N = 178), $BF_{10} = 6.223 * 10^{16}$, and E2 choices no different than chance, (.337, N = 178), BF_{10} = 0.09. This is again corroborated by the decisive deviation between participants and their model predictions found by contingency table analysis (N = 356), $BF_{10} = 1.473 * 10^7$.

Unequal Priors, Unequal Likelihoods. When both priors (P(H1) = .5; P(H2) = .1) and likelihoods are unequal, we find the same general trend, albeit to a lesser degree. More precisely, although participant choices for E2 are again greater than chance (.478, N = 178), $BF_{10} = 368.73$, choices for E1 are also just above chance level (.427, N = 178), $BF_{10} = 3.499$. However, model predictions again show a decisive preference for E1 (.612, N = 178), $BF_{10} = 6.729 * 10^{11}$, whilst E2 should be preferred no more often than chance (.388, N = 178), $BF_{10} = 0.335$. This insufficiency of E1 choices is again captured by the decisive difference in judgment proportions when comparing participants and model predictions in a contingency table (N = 356), $BF_{10} = 15735.87$.

Common Priors, Unequal Likelihoods. Turning finally to when priors are both common (.5) and likelihoods are unequal, we see the same behavioral pattern of a preference for E2 above chance level (.59, N = 178), $BF_{10} = 7.748 * 10^9$, and E1 no different than chance (.331, N = 178), $BF_{10} = 0.088$. However, unlike the preceding parameter sets, E2 is also chosen above chance level by model predictions (.961, N = 178), $BF_{10} = 1.996 * 10^{69}$, whilst E1 is in fact chosen decisively less than chance (.039, N = 178), $BF_{10} = 7.246 * 10^{18}$. Further, participant choices for E2 are shown to be insufficient compared to model predictions (N = 356), $BF_{10} = 9.44 * 10^{14}$. This is likely due to the high E1 "false positive" due to marginalization over high H2 probability, making E1 comparatively less diagnostic of H1.

Confidence in evidence preferences. Confidence was generally high across all preferences (M = 66.00, SD = 23.96). Although a Bayesian repeated measures ANOVA revealed confidence to be unaffected by preference, $BF_{Inclusion} = 0.781$, or parameters, $BF_{Inclusion} = 1.064$, but there was strong evidence for confidence being higher in the E2 false positive rate = low condition (M = 68.98, SD = 23.39), rather than high (M = 62.95, SD = 24.18), $BF_{Inclusion} = 11.377$. This finding fits with an easier E2 preference when it is a more sensitive test.

Overlooked information

To elucidate the information cost of the above deviations from normative expectation, for each BIBN model (i.e. each participant-fitted model) the expected informative value (in KL-D) was calculated for E1 and E2. In this way, if a participant selected the evidence with the highest KL-D as predicted by their model, they had not overlooked any information, and thus scored 0. However, if participants selected the less informative evidence, then the overlooked information was the difference (in KL-D) between the optimal (i.e. most informative) evidence and their selected option.[3]

As Table 3 indicates, across all break-downs of evidence choices (overall, by condition, and by parameter set), there was a decisive amount of information overlooked – calculated via Bayesian one sample t-tests (test value = 0). This significant amount of overlooked information can be attributed to the sub-optimal undervaluing of E1 (i.e. the zero-sum fallacy) in all cases barring common priors, unequal likelihoods (bottom row, Table 3). In this latter parameter set, E2 in fact yielded the most information, but was not chosen sufficiently often across participants.

Table 3. Overlooked information; overall, split by condition, and split by parameter sets.

	M	*SD*	*N*	*>0 (BF_{10})*	*δ*	*δ 95% CI*
Overall	.045	.048	712	$5.79*10^{95}$	0.934	.847, 1.021
P(E2|¬H1) = L	.042	.047	360	$1.23*10^{44}$	0.886	.766, 0.999
P(E2|¬H1) = H	.048	.049	352	$2.33*10^{50}$	0.980	.849, 1.113
RareP.EqL	.052	.045	178	$3.12*10^{31}$	1.153	.956, 1.348
RareP.UneqL	.047	.042	178	$2.543*10^{30}$	1.121	.938, 1.318
UneqP.UneqL	.046	.057	178	$6.600*10^{17}$	0.795	.623, 0.958
ComP.UneqL	.034	.045	178	$2.569*10^{16}$	0.753	.59, 0.925

Conclusions

Previous work has shown that evidence equally predicted by multiple explanations is often erroneously dismissed due to the misplaced assumption that support for one hypothesis (of interest) must come at the detriment of another (the zero-sum fallacy; Pilditch, Fenton, & Lagnado, 2019). In the present work, we show that this fallacy results in poor decisions regarding evidence selection, and that such selections come at a quantified cost of overlooked information. Crucially, we also show that participants are sensitive to priors and likelihoods parameters, with different evidence preference patterns as a consequence. However, the general pattern of overlooked information holds despite this sensitivity.

Foremost, the present work confirms the presence of zero-sum reasoning, showing that it is active in people's choice of which evidence to examine. It also highlights the potential costs of the fallacy, via the quantification of (costly) overlooked information. In this way, we argue for the inclusion of different question methods and information measures when investigating reasoning errors – whether across simple or complex structures. This would not only contribute to understanding how violations of normative frameworks of human information acquisition relate to known violations of information evaluation, such as the

[3] If evidence items were equally informative, then participants were pragmatically correct, in terms of information, with any preference (including "They are the same"), and thus scored 0. However, if participants erroneously judged the evidence items the same, the amount of overlooked information was taken from the KL-D of the most informative option.

zero-sum fallacy, but it would allow for the exploration of how consequential sub-optimal evidence selection choices are, in laboratory as well as real-world settings.

Given that information-seeking is a critical aspect of so many areas of decision making – including intelligence analysis, legal reasoning, and medical diagnosis – the use of zero-sum reasoning is a strong concern. Future work will seek ways to alleviate this bias, and shift people towards more normative information gathering.

References

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, 5, 1-17.

Højsgaard, S. (2012). Graphical independence networks with the gRain package for R. *Journal of Statistical Software*, *46*(10), 1-26.

JASP Team (2018). JASP (Version 0.9.1)[Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd Ed.). Oxford, UK: Oxford University Press.

Jones, M., & Sugden, R. (2001). Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50(1), 59-99.

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*. Advance online publication.

Nelson, J. D. (2008). Towards a rational theory of human information acquisition. In M. Oaksford & N. Chater (Eds.), *The probabilistic mind: Prospects for rational models of cognition* (pp. 143–164). Oxford, United Kingdom: Oxford University Press.

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological science,* 21(7), 960-969.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Francisco, CA: Morgan Kaufmann.

Pearl, J. (2009). *Causality. Models, reasoning, and inference*. Second edition. New York: Cambridge University Press.

Pilditch, T. D., Fenton, N., & Lagnado, D. (2019). The zero-sum fallacy in evidence evaluation. *Psychological Science*, 1-11. DOI: 10.1177/0956797618818484

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition,* 20(4), 392-405.