Title

Reasoning about (In)Dependent Evidence: A Mismatch between Perceiving and Incorporating Dependencies?

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Strittmatter, Laura Elaine

Pilditch, Toby D

Lagnado, David

Publication Date

2024

Peer reviewed

Reasoning about (In)Dependent Evidence: A Mismatch between Perceiving and Incorporating Dependencies?

Laura Elaine Strittmatter[1,2], Toby D. Pilditch[2,3], and David A. Lagnado[2]

[1]ETH Zürich, Department of Management, Technology, and Economics, Weinbergstrasse 56/58, 8092 Zürich, Switzerland
[2]University College London, Department of Experimental Psychology, London, WC1H 0AP, United Kingdom
[3]University of Oxford, School of Geography and the Environment, South Parks Road, Oxford, OX1 3QY, United Kingdom

Abstract

Independent pieces of corroborating evidence should provide stronger support to a hypothesis than dependent pieces of evidence. Overlooking the inferiority of dependent relative to independent items of evidence can lead to a chain reaction of double-counting evidence, over-estimating the probability that the fact under consideration is true, and making wrongful decisions. Within one medical and one criminal scenario, we investigate people's sensitivity to the independency advantage. We assess their ability to integrate multiple items of evidence that come from (in)dependent sources who differ in reliability. We find that participants properly perceive dependencies when explicitly asked but fail to distinguish the probative value of dependent versus independent evidence in their belief updating. Still, individuals who perceive a strong dependence between sources treat the evidence as being more redundant. We find no dependency-related effects on participants' individual Bayesian network model predictions derived from their own conditional probability assumptions. Potential reasons why participants perceive (in)dependencies and yet (mostly) fail to discount for them are discussed.

Keywords: dependence, evidential reasoning, probative value, uncertainty, belief updating, Bayesian network models

Medical errors linked to false positives (commission errors) and false negatives (omission errors) constitute the third leading cause of death in the U.S. (Makary & Daniel, 2016; Singh et al., 2017). The high error prevalence permeating medicine can be ascribed mainly to cognitive errors (Graber et al., 2005). Combining independent diagnostic evidence of multiple physicians (even as few as two) substantially enhances diagnostic performance (Barnett et al., 2019; Wolf et al., 2015). There is little doubt that merging multiple independent assessments (i.e., items of evidence) from reasonably reliable individuals (i.e., sources) are useful for fact-finding and to finalise decisions (Soll, 1999; Van De & Delbecq, 1971). Less clear is the benefit of integrating two assessments if one expert knows about her colleague's evidential report before examining the case by herself. In this case the report of the second expert (partially) depends on her colleague's report compared to a counterfactual example in which both experts reach a corroborative conclusion entirely on their own and independently from each other. People who neglect directional dependency between evidence variables are at risk to over-value or double-count evidence (Schum & Martin, 1982). In this study we assess whether people's evidential reasoning is sensitive to the directional dependence between evidence and whether this is moderated when reliabilities differ between sources.

(In)Dependence Between Reports

Consider a patient concerned about her state of health who sees two physicians working in different hospitals. Doctor A examines the patient and diagnoses a malignant tumour. To obtain a second opinion, the patient sees Doctor B. Imagine two distinct cases:

(1) Doctor B reads Doctor A's report (i.e., knows about Doctor A's diagnosis) before examining the patient and diagnosing a malignant tumour, or

(2) Doctor B does not read Doctor A's report (i.e., does not know about Doctor A's diagnosis) before examining the patient and diagnosing a malignant tumour.

In case (1), the evidential report of Doctor B is *dependent* on and potentially influenced by Doctor A's report. In case (2) there is no direct information flow between the physicians and thus their medical reports are *independent* (Bovens & Hartman, 2003). Normatively, from a third party's perspective the dependence should be incorporated by adjusting the evidential strength of B's report, which affects the probability estimate that the patient suffers from a tumour (see Schum & Martin, 1982). The probabilistic belief of the tumour being present should increase to a greater extent if B's evidence is independent from A's evidence compared to the dependent analogue.

Now, imagine a highly reliable senior receives a medical report from a less accurate novice. The dependence may be negligible given the assumption that a less reliable source may provide neither benefit nor harm to the more reliable source (see Pilditch et al., 2020). Conversely, when the novice receives a medical report from the senior, the novice may disregard a greater proportion of his own private evidence and instead rely more heavily on the findings of his senior colleague (larger dependence) given he is aware of his own and his colleague's level of expertise (see Beauchamp et al., 2024).

Statistical Dependence Between Sources

Even the most conscientious experts are not spared from being biased by contextual information that is case relevant

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

but irrelevant for the objective investigation itself (Dror, 2016). For instance, forensic expert B who is aware about his colleague A's report (e.g., DNA matches forensic evidence) is inclined to erroneously arrive at a result that is consistent with A's report (e.g., fingerprint matches forensic evidence; see Dror et al., 2006; Dror & Cole, 2010) and thus retains the direction in which the evidence works (e.g., suspect being guilty instead of being innocent; see Dror & Murrie, 2018). This form of 'biasability' between examiners is not restricted to criminal justice but generalises to various tasks and contexts that are rooted in human perception and judgment (Dror, 2016; Schum, 2001). The impact of biasing information is especially high when the conclusion is based on vague observations that are close to the decision threshold (i.e., lacking from sufficient certainty; Dror, 2016).

People's theories about the dependence of evidence and its impact are usually opaque. How much and in which form B is influenced by A is uncertain and must be intuitively judged by reasoners. People may inter-individually differ when considering the point on the continuum between B relying entirely on his own investigations and B simply copying A's report (see Dror & Murrie, 2018).

Bayesian Networks (BNs)

Bayesian networks (BNs) are graphical models of uncertainty that capture the essentials of normative evidential reasoning: They enable us to represent the structural and probabilistic interrelations between variables in the proposed underlying model of a system (Lagnado et al., 2013; Pearl, 2000).

As depicted in Figure 1, the arrows indicate probabilistic dependencies between nodes (variables). The direction of arrows represents the stream of influence a 'parent' variable exerts onto its 'child'. A node without any parents, namely the root node, is unconditional. In our simple cases the hypothesis of interest (H) is a root node and carries the prior probability of H being true (in our cases either tumour present or suspect guilty). Each child node is a piece of evidence (E) which is associated with a conditional probability table (CPT) incorporating all possible states of its parent(s). E_A and E_B are items of evidence provided by investigator A and B, respectively.

The dashed arrow represents the possible dependency between E_B and E_A. In the dependent scenario B has access to A's report but not the reverse. Thus, E_B has two parent nodes that need to be considered for E_B's underlying CPT. In other words, E_B's state depends not only on H but also on the state of E_A. The two expertise conditions (i.e., Fig. 1a and 2b) differ in the sense that the expertise asymmetry is converse. Note that in case of an independence, the dashed arrow of direct dependence is removed, and the two models become structurally identical.

BNs provide an optimal way to analyse the subtleties of the relationship between variables, such as combining dependent pieces of evidence (Lagnado, 2011) described in the example above. In keeping with that theme, BNs allow us to establish the evidence's inferential direction (i.e., favours one hypothesis over another) and inferential force (the degree to which the evidence supports a hypothesis; Juchli et al. 2012). In the Bayesian analysis this is usually given as likelihood ratio (LR), which conceptualises the weight of evidence for or against a hypothesis (i.e., diagnosticity) and thus carries discriminative power for the states of the hypothesis (see Lagnado et al., 2013). The LR is the ratio of the two conditional probabilities: the probability of the evidence given the hypothesis is true $P(E|H)$ compared to if the hypothesis is false $P(E|\neg H)$. LR > 1 provides positive evidence for H, whereas LR < 1 provides negative evidence for H. Multiplying the prior odds (i.e., the naïve state) with the likelihood ratio yields the posterior.

In the current study, the impact of evidence is scaled by the reliability of its source, the dependence between sources, and their interaction. Going back to the medical and forensic examples, let's assume the LR of the evidence provided by the novice (4 in 5 times accurate; 1 in 5 times wrong) is LR = .8/.2 = 4, whereas the report of the senior investigator (19 in 20 times accurate; 1 in 20 times wrong) giving a LR = .95/.05 = 19. Both items of evidence, independently from each other, provide positive evidence for H. Nonetheless, people should, as prescribed by Bayes' rule, update their prior more heavily when they receive evidence from the senior compared to the novice given the sources are independent.

However, when B depends on A (dashed line), E_B loses some evidential weight compared to the independent case (no dashed line), ceteris paribus. The reduction in E_B's weight due to the dependence should be more pronounced when the novice receives the report from the senior (bold dashed line signifies strong dependence) than when the senior receives the report from the novice (thin dashed line signifies little dependence).

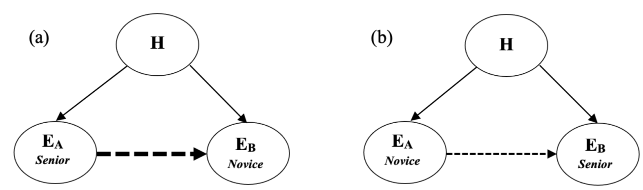

Figure 1: Bayesian networks (BNs): H corresponds to the hypothesis, E_A to the first item of evidence and E_B to the second item of evidence. The dashed arrow signifies the (potential) dependence between E_A and E_B given H. (a) information flows from the senior to the novice (b) information flows from the novice to the senior. If E_A and E_B are independent, (a) = (b).

Previous Research

What has been explicated so far, is the dependence as one-way information transference between sources, such that source B can see the report of source A but not the reverse. There are also other forms of dependencies that are not focal in the present study such as correlated or shared information (i.e., mutually observed evidence; Enke & Zimmermann, 2019; Whalen et al., 2018), sharing a common background or a common motivational ground, and dependence as informational consistency between otherwise independent sources (Bovens & Hartman, 2003; Madsen et al., 2018; Tindale & Kameda, 2000).

Research on people's sensitivity to dependency cues has yielded mixed results. Some data suggest that people fail to value independent social information more than non-independent information (Sulik et al., 2020) and that they seem to be insensitive to the inferior quality of a consensus if sources reach their conclusion dependently as compared to independently (Yousif et al., 2018). Even if individuals are clearly aware that low quality of consensus (i.e., due to dependence of sources) *should* be discounted, they still fail to do so by displaying an overreliance on dependent consensus (Yousif et al., 2019). In contrast, other work shows that that people can be responsive to violations of informational independencies (Mercier & Miton, 2019). Individuals value the opinion of informants more if they had no mutual access to each other's opinion compared to those formed with mutual informational access (Bloomfield & Hales, 2009). Given the unclear picture, Bayesian comparisons have proven useful to elucidate people's (in)appropriate belief updating when facing dependencies (e.g., Fränken et al. 2020).

Importantly, Pilditch et al. (2020) recently demonstrated how people reason about information of two investigators who assess the case of a plane crash to figure out its cause: either a sabotage or an accident (50% probability each). In the dependent scenario, one investigator has access to her colleague's report, whereas in the independent scenario her report is based on her assessment alone. For both cases participants estimated the probability for the sabotage and whether the independent or dependent case provided stronger support for the sabotage hypothesis. Additionally, their underlying conditional probability judgments were measured to compute normative BNs and to identify the nuances of potential (beneficial/harmful) influence between reports. Given the two investigators' reports were corroborative, participants were sensitive to the advantage of independence. The BNs reflected a similar pattern. In a subsequent experiment they introduced differing reliabilities among the two investigators. Although participants incorporated reliability into their probability estimates, the dependency cue was heavily under-utilised. This implies a vulnerability of people's sensitivity to dependencies when reliability is manipulated.

Overview of Present Research

We examined people's sequential belief formation within the (in)dependent legal and medical scenarios described above. We focus on Pilditch et al.'s (2020) finding that a clear preference for independence in corroborating evidence disappears (in qualitative and probability judgments) when the two sources were of unequal reliability. Currently, we aimed to test those findings using different scenarios and adopting some novel method: To examine people's appreciation of structural dependencies, we probed not only their probability estimates of the hypothesis being true and their conditional probability assumptions (see Pilditch et al., 2020), but also their perceived dependence (and the confidence thereof) by asking explicitly how much they think that B's report depends on the content of A's report given that B has or has not read A's report. By not putting the dependent and independent case into direct comparison, we assessed how robust people's intuitions might be when being exposed to a more implicit paradigm. In addition to that, we measured participants' final treatment/culpability decision.

Predictions of Current Study

We predict that people incorporate, at least partially, the subtlety of statistical (in)dependence in their probability judgments as would be expected by the Bayesian model. Namely, when corroborative evidence is statistically dependent, people will raise the probability less in their posterior compared to when it is independent. In addition, participants will indicate a negligibly low perceived dependence with high confidence ratings in the independent case, whereas in the dependent case their judgment of perceived dependence will be higher, but with lower confidence ratings.

The pattern of sensitivity to statistical dependency between sources is anticipated to be moderated by the sources' reliability.

(a) If the first source is a senior and the second source a novice (see Figure 1a), we predict an independency advantage when contrasting dependent and independent cases. When the novice receives correct information from the senior, he is more likely to be correct, whereas when receiving incorrect information, he is more likely to be misled – the novice is (partially) mirroring the senior's report and contributes less uniquely.

(b) If the first source is a novice and the second source is a senior (see Figure 1b), the dependent and independent cases are much alike. If information flows from the novice to the senior, the senior may be less reliant on the report that originates from a source that is in terms of statistical accuracy inferior to his own. The senior is neither assisted by correct information from the novice, nor misled by incorrect information by the novice. Thus, the independency advantage will diminish or even vanish when comparing dependent and independent cases.

Nonetheless, we do not rule out that participants generally fail to incorporate the subtlety of dependence given the additional complexity with varying levels of expertise of the evidence providers when mentally modelling the totality of the evidence (see Exp. 3, Pilditch et al., 2020).

Method

Participants We recruited 80 participants (65% female; M_{age} = 31.5; age range 18–80) via Prolific (>95% approval rate). Participants identified as native English speakers based in the UK, US, Canada, Australia, and New Zealand. Participants were compensated for their time \$2.23 (*Mean* = 12.51 minutes, *SD* = 7.07).

Design We used a mixed design, with Dependency (independent vs. dependent) and Time Stage (t_0, t_1, t_2) as within-subject factors and Expertise (novice sending source vs. senior sending source) as between-subject factor. We controlled for order effects. The key dependent variable (DV) was participants' *probability judgment* of H (i.e., tumour present/suspect guilty) at t_0 (baseline), t_1 (first evidence) and t_2 (second evidence). *Perceived dependency* of reports and *confidence* thereof were measured at t_2. Moreover, the treatment/culpability *decision* was measured at t_2. Finally, participants' *conditional probability judgments* were taken to construct Behaviorally Informed Bayesian Networks (*BIBNs*; see Pilditch et al., 2020).

Materials and Procedure

Overview Materials and data are available at https://osf.io/7xc9w/. Participants were instructed to reason with information about two different scenarios: A fictitious medical and criminal case. Dependent measures were taken using a step-by-step method (see Hogarth & Einhorn, 1992). The order of scenario type (medical/forensic) and dependency (independent/dependent) was counterbalanced.

Dependence was manipulated within-subject. Participants received one dependent and one independent scenario, of which one was medical, and one was forensic.

Participants were randomly allocated to one of two expertise between-subjects conditions whereby the second investigator was either lower (novice: error rate of 1 in 5 [20%]; accuracy rate of 4 in 5 [80%]) or higher (senior: error rate of 1 in 20 [5%]; accuracy rate of 19 in 20 [95%]) in expertise than the first investigator when providing reports independently (see full scenarios at OSF). The accuracy was symmetrical and hence true positive rates equal true negative rates. The same applied for error rates (i.e., false positive rates equal false negative rates).

Scenarios At t_0 (baseline) participants were presented with the prior probability of H being true that is 1 in 10 or 10% (tumour present/suspect is guilty). They were informed that two investigators will examine the case in question. At t_1 participants were exposed to the first piece of evidence supporting the hypothesis provided by the first investigator A (report of Doctor/DNA investigator who was a novice/senior). At t_2, investigator B (senior/novice) provided the second piece of corroborating evidence. In the dependent condition B knew about A's report and whether A was a novice or a senior. In the independent condition B was completely blind about A's report and did not know that an investigation prior to his own had taken place.

Measures Participants provided probability judgments across the three time stages (reminding them of their initial judgments), t_0: P(H), t_1: P(H|E_A), and t_2: P(H|E_A,E_B): "Based on what you know at this point, how likely is it that the patient suffers from a malignant tumour/the suspect is guilty?" (0 [*not likely at all*] – 100 [*certain*]). After the complete scenario has been presented, participants were asked to indicate the extent to which they perceive a dependency: "How much do you think that Doctor B's diagnosis depends on the content of Doctor A's medical report given that he has (not) read Doctor A's medical report?" (0 [*not dependent at all*] – 100 [*entirely dependent*]) and to provide the confidence thereof: "How confident are you that your previous response is correct?" (0 [*not confident at all*] – 100 [*extremely confident*]). Participants were asked to finally decide whether the patient should undergo treatment (0 [*definitely no treatment*] – 100 [*definitely treatment*]) or whether the suspect should be found guilty (0 [*definitely not guilty*] – 100 [*definitely guilty*]). Finally, participants provided conditional probability judgments: (I) "Suppose the tumour is present and Dr A reports correctly that the tumour is indeed present, how likely is it that Dr B reports that the tumour is present?"; (II) "Suppose the tumour is present but Dr A reports erroneously that the tumour is NOT present, how likely is it that Dr B reports that the tumour is present?"; (III) "Suppose the tumour is NOT present and Dr A reports correctly that the tumour is indeed NOT present, how likely is it that Dr B reports that the tumour is present?"; (IV) "Suppose the tumour is NOT present but Dr A reports erroneously that the tumour is present, how likely is it that Dr B reports that the tumour is present?" (0 [*not likely at all*] – 100 [*certain*]).

Results

We conducted separate repeated measures ANOVAs with the following factors: 3 (time stage: t_0, t_1, t_2) × 2 (dependency: dependent vs. independent) × 2 (expertise: A < B vs. A > B) × 2 (response type: participant vs. BIBNs) × 2 (dependence order: dependent first vs. independent first) × 2 (scenario order: medical first vs. forensic first) with expertise, dependence order, and scenario order as between-subjects factor.[1] Additional tables and figures of dependent variables can be found at OSF.

Probability Estimates There was no main effect of dependence nor a significant dependence × expertise interaction ($p > .05$) on probability estimates (black lines in Figure 2). The analysis revealed a main effect of time stage, $F(1.41, 101.26) = 429.91$, $p < .001$, $\eta_p^2 = .857$, indicating an increase of probability estimates across time, $p < .001$. There was a main effect of expertise $F(1, 72) = 9.03$, $p = .004$, $\eta_p^2 = .111$. Participants' belief updates increased more when receiving evidence from the senior compared to the novice. These main effects were qualified by an interaction of time stage and expertise, $F(2, 144) = 13.1$, $p < .001$, $\eta_p^2 = .154$. Simple effect analyses showed that probability estimates between the expert reports differed solely at t_1 (lower for the novice's report than for the senior's report, $p < .001$). At t_2,

[1] Unexpectedly, we found a series of interactions with the order of dependence and scenario on some measures.

the probability estimates between levels of expertise converged and were not significantly different.

BIBNs Conditional probabilities were used to compute BIBNs (via gRain package in R). Each participant has a separate BIBN for the independent and the dependent scenario, respectively (see grey lines Figure 2). Corroborating participants' belief adjustments, the two main effects of expertise (senior > novice) and time stage ($t_0 < t_1 < t_2$) of the model predictions were qualified by an interaction, $F(2, 156) = 44.37$, $p < .001$, $\eta_p^2 = .363$, reflecting higher posteriors when source A was a senior (M = 70. 52, SD = 16.04) compared to a novice (M = 43.17. 52, SD = 16.51) solely at t_1, $p < .001$. We found no significant dependency related effects on the model predictions.

Model Comparisons As illustrated in Figure 2, participants' probability estimates deviated significantly from their BIBNs, (participants > BIBNs), $F(1, 78) = 20.36$, $p < .001$, $\eta_p^2 = .207$. Participants overestimated the probative force of evidence relative to the model predictions (except from t_0), more evident at t_1 compared to t_2 as the Response Type × Time interaction revealed, $F(2, 78) = 17.12$, $p < .001$, $\eta_p^2 = .18$.

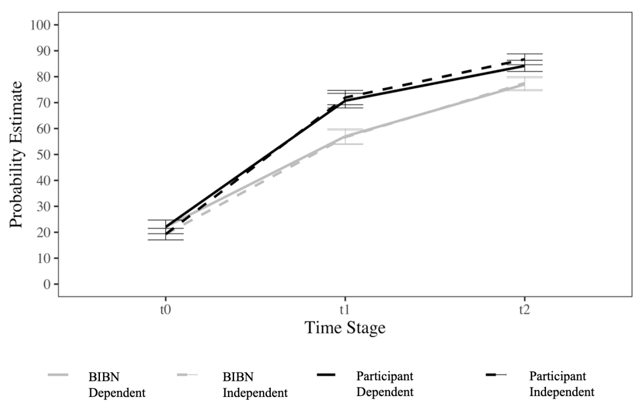

Figure 2: Participants' mean estimates of the probability (black line) and individually fitted BIBNs (grey lines) across Time Stage. Dashed lines correspond to independent cases and solid lines to dependent cases. Expertise data are aggregated. Error bars represent ±1 standard error of mean.

Conditional Probabilities Several t-tests were conducted to assess whether receiving correct or incorrect information from the sending source were beneficial or detrimental to the recipient's accuracy and error rate, respectively. Following Pilditch et al.'s (2020) approach, all values were log transformed prior to the analysis (log (x +1)) to combat values of 0. In sum, participants were sensitive to the influence A had on B's accuracy/error rate. To our surprise, the influence was also captured when sources were independent.

Perceived Dependency Perceived dependency was judged higher in the dependent case (M = 46.34, SD = 39.1) relative to the independent case (M = 11.23, SD = 2.88; $p < .001$), fitting the manipulation check, $F(1, 72) = 109.97$, $p <.001$, $\eta_p^2 = .604$. The main effect of expertise revealed that perceived dependency was generally judged higher when the sending source was a senior (M = 39.34, SD = 20.5) compared to a novice (M = 18.75, SD = 27.36; $F(1, 72) = 23.1$, $p < .001$, $\eta_p^2 = .243$). As expected, the dependence × expertise interaction indicated that the dependence sensitivity was even stronger if the sending source was a senior, $F(1, 72) = 57.67$, p < .001, $\eta_p^2 = .445$. As hypothesised, confidence in perceived dependency was significantly lower in the dependent (M = 68.27, SD = 3.09) compared to the independent scenario (M = 85.05, SD = 2.46; $p < .001$; $F(1, 72) = 27.7$, $p < .001$, $\eta_p^2 = .262$).

In an exploratory analysis, we found that perceived dependency predicted participants' magnitude of belief updating ΔPE (probability estimate t_2 - probability estimate t_1), $r(160) = -.205$, $p = .009$, $R^2 = .042$. The higher participants judged perceived dependence, the smaller was their probabilistic belief revision ΔPE.

Decision (treatment/culpability) Decisions did not differ significantly irrespective of dependence and expertise treatments. Nonetheless, participants' decisions in the dependent case (M = 84.20, SD = 16.82) correspond well to their t_2 posterior probability judgments (M = 84.18, SD = 19.29). Similarly, in the independent case participants' decisions (M = 86.73, SD = 17.67) and t_2 posteriors (M = 86.71, SD = 18.83) are highly in agreement with each other.

Discussion

Are our beliefs about the truth differently shaped when integrating corroborating evidence from dependent relative to independent sources? How is this affected when sources vary in reliability? We anticipated that participants exploit the information of directional dependence when making probabilistic inferences about a singular event, but more so when a less reliable source depends on a more reliable source than the reverse.

Probability Estimates, BIBNs, and Decision Participants failed to sufficiently incorporate the nuance of (in)dependency in either their belief updating or their BIBNs (see also Pilditch et al., 2020). Our findings imply not only that individuals chronically overestimate dependent and/or underestimate independent pieces of evidence, but they also base their final decision of the intervention fully on their (arguably) suboptimal posterior belief that results.

Perceived Dependency Participants' perceived dependency judgments correctly tracked the actual dependency. They also perceived the dependence to be higher when the novice received information from the senior than the reverse. This fits with the assumption that the novice has more capacity for accuracy improvement and relies more heavily on the report of the senior than the reverse. When the senior was the recipient, the perceived dependence faded but was not muted entirely. Thus, it seems that individuals were sensitive to the dependence and its interaction with the source's reliabilities under certain conditions.

As evidenced in the exploratory analysis, the subset of participants who strongly perceived a dependence in the dependent scenario (in absolute terms but not relative to the independent scenario) believed it to imply more redundancy by updating their beliefs proportionally less compared to those who perceived little dependence. These tentative findings suggest that the novel measure of perceived dependency can detect inter-individual differences in people's intuitions about the degree of dependency when B has access to A's report, and that this is coherently linked to their belief revisions.

Conditional Probabilities For the counterfactual questions, B's accuracy and error rates were, according to participants, differentially influenced by A's report depending on the sources' reliability configuration. Since these effects were similar when the scenario was *independent* and in fact no influence could take place, the BIBNs could not capture dependency related differences. It seems as if participants wrongfully inferred a dependency in the independent scenario. Potentially, the nature of counterfactual questions could have blunted dependency related effects in the model predictions: Since it may appear odd to ask about B's conclusion given A's conclusion in the independent scenario, participants may have been misled into conjecturing a dependence.

Capturing Dependency Effects Schum and Martin (1982) demonstrated that people double count corroboratively redundant evidence in their holistic assessment, which is the overall likelihood judgment summarised as a single number, whereas in more fine-grained estimates participants' sensitivity of redundancy became apparent. Currently, although participants seem to be aware of the dependence in the low-dimensional measure of perceived dependency, it is inadequately reflected in their global probability estimate, so that the subtlety of (in)dependence may have been lost when they reasoned based on the totality of evidence.

When participants are explicitly asked about perceived dependence, they have a firm sense of the distinction between dependent and independent items of evidence and its interaction with reliability, but it seems as if they generally struggle to appreciate the consequence thereof. The direct influence dependence *should* exert on their belief revision may be blunted by the additional complexity faced with unequal source reliabilities. This claim cannot be made from the present data alone without baseline condition of two equally reliable sources, which calls for adding such a condition in the future. Introducing a direct measure of 'perceived redundancy', 'perceived influence', and 'perceived evidential weight' given the dependency, would be a way to explore the extent to which individuals are aware about the impact of dependence relative to independence. Further, the link between the dependence and its consequences could be examined through qualitative inquiry.

Professionals may be already associated with a high standard of trustworthiness, credibility, and competence in their position. As a result, a direct dependence between them may be perceived not to compromise the truthfulness of their work - which relates to argumentum ad verecundiam (i.e., appeal to authority; Woods & Walton, 1974). Given that people feel themselves immune against the influence of contextual information (Dror, 2016), they may underestimate potential biasing effects of dependence in experts even more. We acknowledge that the complexities involved in relation to people's Theory of Mind and the way they model dependencies between other minds are not fully disentangled and need to be addressed in the future.

Implications Giving individuals instructions of the relevant variables can improve their reasoning (Nisbett, 1993). It seems advisable to raise the awareness of neglected evidential details of dependency which may appear to be small in nature but can generate substantial effects. One way to overcome at least some obstacles in evidential reasoning tasks is to make use of BNs. The usage of BNs in healthcare and legal systems has been well recognized (e.g., Kyrimi et al., 2020; Richens et al., 2020). It can assist experts to maximally exploit the often incomplete and uncertain pieces of interrelated evidence at hand (Constantinou et al., 2016) without the requirement to understand the precise mathematical mechanisms (Cruz et al., 2020). Nonetheless, as seen in the present data, human reasoners must justify the input regarding the quantitative and qualitative aspects of evidence, so that BNs cannot fully compensate for human errors at earlier stages.

Conclusion

Participants perceived the difference between dependent and independent sources and its interaction with their perceived reliabilities. Moreover, the variance of people's perceived dependence between sources successfully predicted the magnitude of their belief updating. Yet individuals failed to acknowledge a clear-cut distinction between the probative value of dependent compared to independent evidence in their probability estimates and their individual model fits. Further work is needed to delineate possible boundaries of computational difficulty that may underpin this inconsistency. The ignorance of dependency cues can have grave repercussions across various disciplines. Future investigations are highly encouraged to extend this work to hopefully elucidate the manifold facets of people's (in)sensitivity to direct dependence.

References

Barnett, M. L., Boddupalli, D., Nundy, S., & Bates, D. W. (2019). Comparative accuracy of diagnosis by collective intelligence of multiple physicians vs. individual physicians. *JAMA Network Open*, *2*(3), 133- 174.

Beauchamp, N., Shugars, S., Swire-Thompson, B., & Lazer, D. (2024). How expertise mediates the effects of numerical and textual communication on individual and collective accuracy. *Decision, 11*(1), 194–211.

Bloomfield, R., & Hales, J. (2009). An experimental investigation of the positive and negative effects of mutual observation. *The Accounting Review*, *84*(2), 331-354.

Bovens, L., & Hartmann, S. (2003). Solving the riddle of coherence. *Mind*, *112*(448), 601-633.

Constantinou, A. C., Fenton, N., Marsh, W., & Radlinski, L. (2016). From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. *Artificial Intelligence in Medicine*, *67*, 75-93.

Cruz, N., Desai, S. C., Dewitt, S., Hahn, U., Lagnado, D., Liefgreen, A., ... & Tešić, M. (2020). Widening access to Bayesian problem solving. *Frontiers in Psychology*, *11*, 660.

Dror, I. E. (2016). A hierarchy of expert performance. *Journal of Applied Research in Memory and Cognition*, *5*(2), 121-127.

Dror, I. E., & Cole, S. A. (2010). The vision in "blind" justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review*, *17*(2), 161-167.

Dror, I. E., & Murrie, D. C. (2018). A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law*, *24*(1), 11.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*(1), 74-78.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies, 86*, 313-332.

Fränken, J. P., Theodoropoulos, N. C., Moore, A. B., & Bramley, N. R. (2020). Belief revision in a micro-social network: Modeling sensitivity to statistical dependencies in social learning. *Proceedings of the Annual Conference of the Cognitive Science Society*.

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of Internal Medicine, 165*(13), 1493-1499.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1-55.

Juchli, P., Biedermann, A., & Taroni, F. (2012). Graphical probabilistic analysis of the combination of items of evidence. *Law, Probability and Risk*, *11*(1), 51-84.

Kyrimi, E., Neves, M. R., McLachlan, S., Neil, M., Marsh, W., & Fenton, N. (2020). Medical idioms for clinical Bayesian network development. *Journal of Biomedical Informatics*, 103495.

Lagnado, D. A. (2011). Thinking about evidence. In *Proceedings of the British Academy* (Vol. 171, No. 183, p. e223). Oxford, UK: Oxford University Press.

Lagnado, D. A., Fenton, N., & Neil, M. (2013). Legal idioms: a framework for evidential reasoning. *Argument & Computation*, *4*(1), 46-63.

Madsen, J. K., Hahn, U. & Pilditch, T. (2018). Partial source dependence and reliability revision: the impact of shared backgrounds*, Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ*, *353*.

Mercier, H., & Miton, H. (2019). Utilizing simple cues to informational dependency. *Evolution and Human Behavior*, *40*(3), 301-314.

Nisbett, R. E. (Ed.). (1993). *Rules for Reasoning*. Psychology Press.

Pearl, J. (2000). *Causality*. New York: Cambridge University Press (2nd edition).

Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (2020). Dependencies in evidential reports: The case for informational advantages. *Cognition*, *204*, 104343.

Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, *11*(1), 1-9.

Schum, D. A. (2001). *The evidential foundations of probabilistic reasoning*. Northwestern University Press.

Schum, D. A., & Martin, A. W. (1982). Formal and empirical research on cascaded inference in jurisprudence. *Law and Society Review, 17,* 105-151.

Singh, H., Schiff, G. D., Graber, M. L., Onakpoya, I., & Thompson, M. J. (2017). The global burden of diagnostic errors in primary care. *BMJ Quality & Safety*, *26*(6), 484-494.

Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology*, *38*(2), 317-346.

Sulik, J., Bahrami, B., & Deroy, O. (2020). Social influence and informational independence. *Proceedings of the Cognitive Science Society Conference*.

Tindale, R. S., & Kameda, T. (2000). 'Social sharedness' as a unifying theme for information processing in groups. *Group Processes & Intergroup Relations*, *3*(2), 123-140.

Van De, A., & Delbecq, A. L. (1971). Nominal versus interacting group processes for committee decision-making effectiveness. *Academy of Management Journal*, *14*(2), 203-212.

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive Science*, *42*(1), 168-187.

Wolf, M., Krause, J., Carney, P. A., Bogart, A., & Kurvers, R. H. (2015). Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. *PloS one*, *10*(8), e0134269.

Woods, J., & Walton, D. (1974). Argumentum ad verecundiam. *Philosophy & Rhetoric*, 135-153.

Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science*, *30*(8), 1195-1204.

Yousif, S. R., Aboody, R., & Keil, F.C. (2018). Any consensus will do: the failure to distinguish between 'true' and 'false' consensus. *Proceedings of the Cognitive Science Society Conference*.