The role of causal models in evaluating simple and complex legal explanations

Alice Liefgreen (alice.liefgreen@ucl.ac.uk)
Department of Experimental Psychology, 26 Bedford Way, WC1H 0AP, London, UK

David Lagnado (d.lagnado@ucl.ac.uk)
Department of Experimental Psychology, 26 Bedford Way, WC1H 0AP, London, UK

Abstract

Despite the increase in studies investigating people's explanatory preferences in the domains of psychology and philosophy, little is known about their preferences in more applied domains, such as the criminal justice system. We show that when people evaluate competing legal accounts of the same evidence that vary in complexity, their explanatory preferences are affected by: i) whether they are required to draw causal models of the evidence, and ii) the actual structure that is drawn. Although previous research has shown that people can reason correctly about causality, ours is one of the first studies that shows that generating and drawing causal models directly affects people's evaluations of explanations.

Keywords: explanation; causal models; evidential reasoning; simplicity; mechanism

Introduction

Imagine being a juror in a criminal trial in which a mother is accused of killing her infant son. You are told that the medical examination on the son revealed three distinct injuries – blood in the lungs, a torn frenulum (tissue between lip and jaw) and bruises on the arms and legs. The prosecutor offers an explanation as to how those injuries occurred: the mother caused all of them by smothering the child. The defence lawyer then offers a different explanation for how those injuries occurred: three independent incidences – all natural or accidental – brought about the injuries. Both sides put forth plausible explanations, but only one of the two can be true. How would you evaluate these explanations and what factors would you take into account when comparing them?

Research in philosophy and cognitive science has, over the past decades, suggested that we judge and evaluate explanations partly based on how well they satisfy a set of explanatory virtues, or features, including simplicity, coherence and breadth (for overview see Mackonis, 2013). The present work concerns itself primarily with the simplicity virtue. According to this virtue, a hypothesis is a better explanation, the simpler or more parsimonious it is – reflecting the principle known as "Ockham's Razor" (Pacer & Lombrozo, 2017). Simplicity, however, can be defined in more than one way. An explanation can be simple in the sense that it appeals to few entities (or few different types of entities), in the sense that it involves fewer number of causes, or in the sense that it is highly inflexible (Lombrozo, 2007, 2016). In the present paper we adopt a definition of simplicity used by (Read & Marcus-Newhall, 1993) - according to which simpler explanations are ones that make fewer 'assumptions'. Despite disagreement on a formal definition of simplicity, there has been overall widespread empirical agreement that simpler explanations are preferred and are found more satisfying than complex explanations in a wide array of settings (Chater & Vitányi, 2003; Lombrozo, 2007; Walker, Bonawitz, & Lombrozo, 2017). More recent studies, however, have painted a slightly more nuanced picture of people's explanatory preferences for simplicity showing that this virtue is not a good predictor of an explanation's quality in naturalistic settings e.g., when testing real-world explanations found on Reddit (Zemla, Sloman, Bechlivanidis, & Lagnado, 2017). Lim and Oppenheimer (2020) recently put forth a unifying account dubbed the 'complexity-matching hypothesis' suggesting that people believe a "good" or satisfying explanation should be as complex as the event being explained.

Notwithstanding this increase in studies assessing people's preferences for simple versus complex explanations within the fields of philosophy and cognitive science, little is known about people's explanatory preferences in applied domains, such as the criminal justice system – despite explanations being an integral part of how this system functions. Think back to the case presented to you at the beginning of this paper. As a juror on the case, you would have been required to evaluate the competing explanations offered to you by the prosecution and the defence, which varied in degrees of simplicity (as well as other features). How would you have evaluated these two explanations of "what happened"? Given that how legal explanations are evaluated and compared, how their merits are established, can ultimately determine a person's fate, this is a critical question to answer. In addition, answering this question within a legal sphere will enable us to appraise the domain specificity of people's explanatory preferences (e.g., for virtues such as simplicity).

In their notable 'story model', Pennington and Hastie (1988) argue that jurors construct a causal model – that resembles a story – to explain the available evidence at the outset, and subsequently base decisions on the causal interpretation they impose on the evidence. This includes evaluating an explanation on features including coverage, coherence and uniqueness (Hastie & Pennington, 2000). Some of these factors overlap with the aforementioned explanatory virtues considered by researchers in the domains of psychology and philosophy of science, whereas others, like simplicity, remain unexplored within a legal context. Empirical work testing the story model has shown that these causal stories are spontaneously constructed by jurors and seem to mediate verdict decisions (Pennington & Hastie, 1988). Among jurors who choose a particular verdict, substantial overlap in their story structures was found – suggesting that representational aspects influence evaluative processes. Research also

found that story structures are influenced by the order of evidence presentation via affecting people's perceptions of evidence strength, as well as confidence in decisions (Hastie & Pennington, 2000). Although in their studies, Pennington and Hastie developed informal casual networks of the most prevalent story detailed by participants – no research has so far elicited causal graphical models directly from participants when engaging in a legal reasoning task, a gap which we aim to fill.

As well as being a natural way for people to represent evidence in legal domains – and the tool then used to guide inference in these contexts – causal models have also been adopted by researchers as formal systems to evaluate evidence (Smit, Lagnado, Morgan, & Fenton, 2016; Constantinou, Fenton, Marsh, & Radlinski, 2016). Evaluating a single explanation, and comparing multiple explanations, are challenging tasks. Even formal approaches to evaluation such as Causal Bayesian networks (CBNs, Pearl, 2009) do not provide a clear-cut metric for an explanation's quality (Neil, Fenton, Lagnado, & Gill, 2019). So far, the majority of Bayesian models of legal explanations that have been developed have adopted an 'integrated' approach, aiming to represent in a single unified model all of the arguments under consideration, such as those presented by the defence and prosecution in a trial. Fenton et al. (2016) however, have shown that this integrated approach can pose certain modelling difficulties regarding e.g., the mutual exclusivity of certain variables and ensuring that causal dependencies between variables remain consistent despite the competitive nature of the arguments. Only recently, has an approach been developed to model and evaluate competing legal arguments when these are represented using separate CBNs (Neil et al., 2019). This disjunctive approach allows one to account for the differences in variables and causal dependencies that the two arguments may contain. Though this is a notable advancement in the formalisation of legal arguments using causal models, it is still unclear how lay people, i.e., jurors, would structurally represent competing arguments in the first place. Would their causal models take on an integrated or a disjunctive form?

The vast majority of the research on formal models of legal arguments has focused on the comparative aspect and less on the representation and integration aspects despite these being crucial to accurate reasoning. There is thus a need for empirical research to directly elicit the causal models that people construct, and subsequently compare, in legal domains. Importantly, findings could inform the development of formal tools that are able to support people's evidential reasoning and decision-making in these high-stake contexts. The efficacy of causal models in helping people reason in real-world-like situations has thus far been scarcely studied. So far, learning causal structures has been shown to improve probabilistic reasoning in learning, problem-solving and categorisation tasks (Krynski & Tenenbaum, 2007; Waldmann, Hagmayer, & Blaisdell, 2006). In addition, the use of visual displays such as influence diagrams to teach people about causal relationships of a process has also been shown to improve performance when tested on that process (Hung & Jonassen, 2006). In none of the above-mentioned studies, however, were people required to draw their own causal models of the information – something which could prove to be an effective means to support people's reasoning in real-world diagnostic tasks.

In two experiments[1], we investigate: (i) how people represent competing explanations of the same legal evidence by asking them to draw causal models, (ii) whether this information is represented (structurally) differently depending on the order in which it is presented, (iii) people's preferences for simple vs. complex legal explanations, (iv) whether people's explanatory preferences differ depending on what causal structure is drawn and finally (v) whether drawing causal models of explanations engages different explanatory preferences and reasoning patterns than not drawing causal models.

Study 1

In our first study, we explored how people graphically represent two competing explanations of the same evidence when these are presented sequentially for all evidence at once (i.e., the prosecution's full explanation of the evidence is presented, followed by the defence's full explanation of the evidence). In addition, we investigated whether the process of drawing these explanations, in the form of causal models, influences how they are evaluated.

Methods

Participants and Design 214 participants (Mean age = 31.8, SD = 10.8; *n* females = 147) completed Study 1 through Prolific Academic. All participants provided informed consent and were compensated at a rate of £7/h for their time. The study was completed in Qualtrics (`https://www.qualtrics.com`).

Materials and Procedure A between-subjects design was employed. All participants were told they would be presented with information about a criminal case and would be required to answer some questions about the case. Participants were randomly allocated to one of two conditions, hereafter referred to as: 'control' (n=108) and 'draw' (n = 106).

Participants in the 'draw' condition were given a short introduction to causal models and completed a learning/practice block at the outset of the task explaining what causal models were, and how they could be used to graphically represent information. Next, they were introduced to the online tool that they would be required to use during the experiment to draw their own causal models:Loopy [2] . In order to learn how to use this tool they were shown examples of various causal structures and asked to replicate them. As part of the

[1]All data and materials are publicly available via OSF at `osf.io/25quj/?view_only=cea7a787a6ed45d2b574fbda452811f0`

[2]Loopy is an open-source online learning software that allows one to draw causal models and build interactive simulations of how systems work (see `https://ncase.me/loopy/`)

learning/practice block, they were asked to reproduce verbal scenarios (e.g., "Tom has a cough. The doctor thinks that it could be a symptom of either asthma or the flu") in the form of causal models in Loopy.

After having completed the learning/practice block, participants in the 'draw' condition were introduced to the legal scenario they would be required to reason with through a case briefing that described a mother being accused of the death of her infant son. Participants were informed that three distinct medical findings had been recorded after examining the infant's body: i) bruises on the arms and legs, ii) torn lingual frenulum (tissue attaching tongue to floor of mouth) and iii) fresh blood in lungs. After receiving this information, participants in the 'draw' condition were presented (sequentially, in counterbalanced order) with two competing accounts that explain the evidence. As such they learned that the prosecution posited that: *Smothering* (purposeful suffocation) caused all three injuries and that the defence posited that: *Post-mortem effects* (injuries caused during the autopsy) caused the bruises, *Resuscitation effects* (injuries caused during resuscitation attempts) caused the torn frenulum and *Hemosiderosis* (natural condition that leads to blood clustering in organs) caused the blood in the lungs. In this manner they were presented with a 'simple' common-cause explanation of the evidence (prosecution), and a competing 'complex' explanation of the evidence (defence) comprised of multiple independent causes for each piece of evidence. After learning about the first explanation, participants in the 'draw' condition were required to represent it as a causal model in Loopy.

Subsequently, they viewed the next explanation and were asked to draw all the information obtained so far as a causal model. This entailed drawing a model including the information presented by both accounts (prosecution and defence). Participants were instructed that they could draw one model with all the information in it, two different models, or simply represent the information in a way they found most intuitive. After having completed their final causal model drawing, participants were asked which account (defence vs. prosecution) "is the best explanation for the evidence" and had to indicate their answer in a dichotomous forced-choice question. They were also asked to provide reasoning for their choice in a free-form text box. This allowed us to obtain an insight into what explanatory virtues people valued, without constraining them to a set of predetermined selections.

Participants in the 'control' condition started off the task by reading the case briefing and the summary of the evidence and subsequently saw in counterbalanced order the prosecution's and the defence's account for the evidence (on two separate pages). After viewing these, participants were required to choose the best explanation for the evidence and provide reasoning for their choice. All participants were de-briefed at the end of the task.

Results

Explanation Preference The proportion of participants in each condition who chose each explanation as the best explanation for the evidence can be seen in Table 1.

Table 1: Number of participant choices in each condition.

	Complex (defence)	Simple (prosecution)
Control	65	43
Draw	42	64

A Chi Square test of independence with continuity correction revealed a significant difference in the distribution of participants' choices between the two conditions, $\chi^2(1) = 8.24, p = 0.004, V = 0.2$. As can be seen from Table 1, the majority of participants in the 'control' condition chose the defence's explanation as the best explanation for the evidence. Conversely, the majority of participants in the 'draw' condition chose the prosecution's explanation as the best explanation for the evidence. These preferences did not vary depending on the order in which the two explanations were viewed (i.e., defence first vs. prosecution first), $\chi^2(1) = 0.4, p = 0.8, V = 0.02$. Overall, our findings suggest that drawing causal models of the competing explanations affects how these were evaluated.

Reasoning underlying explanation preferences To probe the reasoning underlying participants' explanation preferences and what explanatory features they valued, we analysed their think-aloud responses and extracted six codes. See Table 2 for frequency description of codes and frequency across conditions. A Chi-Square test of independence illustrated a significant difference in the distribution of the six reasoning codes between conditions, $\chi^2(5) = 27.3, p < 0.001, V = 0.36$. Bonferroni corrected post-hoc comparisons revealed the only significant difference was between the percentage of people whose reasoning fell under the 'simplicity/probability' code, with this reasoning code being employed significantly more in the 'draw' condition than in the 'control' condition, $p < 0.001$.

Causal models in 'draw' condition Next, we evaluated the structure of the final causal models drawn in the two conditions. A Chi-Square test of Independence showed a significant difference in the distribution of structures used by participants, $\chi^2(3) = 61.1, p < 0.001, V = 0.53$. The largest cluster of participants (n = 53) represented the competing explanations in two separate causal models, one for each legal account – see Figure (1). Out of these, 66% preferred the simple explanation. 39 participants drew them in an integrated 'combined' model (Fig. 2) – out of these, 71% preferred the simple explanation. 7 participants drew them in three separate models, one for each piece of evidence (Fig. 3) – all of these participants preferred the complex explanation.

Using a Chi Square test of Independence, we investigated whether there is an association between causal structure drawn and chosen explanation.Our results showed there was

Table 2: Reasoning codes with description and frequency across conditions.

Code	Description	Draw (n)	Control (n)
Complexity/Specificity	Greater number of causes of explanation, greater specificity to evidence.	12	20
Mechanism	Questioning mechanism underlying proposed cause-effect relations.	28	43
No Intent/Motive	Lack of intention or motive for killing baby.	6	9
Probability	Likelihood of explanation.	6	9
Simplicity/Probability	Smaller number of causes of explanation and greater likelihood of explanation.	39	9
Other	None of the above codes.	15	19

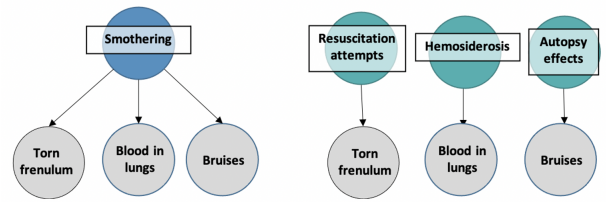

Figure 1: Example of disjunctive representation of two explanations as two models.

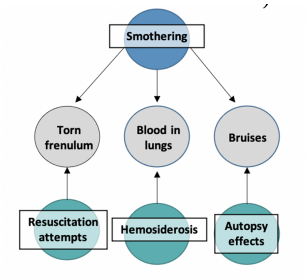

Figure 2: Example of integrated representation of two explanations as a single model.

a significant association between these two factors, $\chi^2(4) = 11.8, p = 0.018, V = 0.3$. The vast majority of participants represented the explanations either in separate models for each account or in a unified model – and, in both of these groups, the majority of participants preferred the simple explanation over the complex one. The only significant post-hoc comparison pertained to the 'separate for each evidence' sub-group ($p = 0.007$), which was associated with a preference for the complex explanation significantly more than the simple explanation. Overall, it appears that representing the explanations in two separate models or as a unified model are not associated with different explanatory preferences – but drawing them as three separate models is (though the low sample size in this sub-group does not allow us to draw strong conclusions from this result).

Study 2

We build on the findings of Study 1 by exploring whether the order that information about the competing accounts is

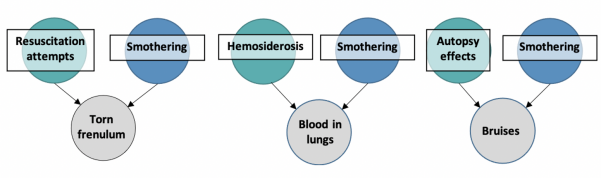

Figure 3: Example of disjunctive representation of two explanations as three separate models.

presented in, affects how the explanations are represented as causal models, and how they are evaluated. As such, in Study 2, participants learned of the two competing explanations simultaneously, for each piece of evidence. This is in contrast to how information was learned by participants in Study 1, in which the complete competing explanations were presented for all of the evidence sequentially.

Methods

Participants and Design 214 participants (Mean age = 35.3, SD = 11.8; n females = 129) completed Study 2 through Prolific Academic.

Materials and Procedure A between-subjects design was employed. As in Study 1, all participants were told they would be presented with information about a criminal case and required to answer some questions about the case. Participants were again randomly allocated to one of two conditions, referred to as: 'control' (n=110) and 'draw' (n = 104). In both conditions, participants reasoned with the same criminal case, evidence, and explanations thereof, used in Study 1. Participants in the 'draw' condition received the same training on causal models and Loopy as that presented to participants of Study 1. Subsequently, they read the same case briefing and report of the evidence found through the medical examination. However, rather than presenting the two competing explanations of all of the evidence sequentially as was done in Study 1, participants in Study 2 saw, for each piece of individual evidence, the two possible explanations as posited by the defence and prosecution simultaneously. As such, they were first told that the prosecution posited that the bruises were caused by smothering and the defence posited that the bruises were caused by autopsy effects. They were then asked

to draw this information in the form of a causal model using Loopy. Subsequently, they were told that the prosecution posited that the torn frenulum was caused by smothering and the defence posited it was caused by resuscitation attempts. They were then asked to draw all the information obtained thus far in a causal model in Loopy. Finally, participants were given the two competing explanations for the final piece of evidence (the blood in the lungs) and were asked to represent all the information obtained so far in Loopy in the manner that seemed most intuitive to them

Participants in the 'control' condition received information in the same manner, however they were not required to draw models representing the information given to them at any stage. In both conditions, at the end of the task, after having viewed both accounts of what caused each injury, participants were asked to choose (dichotomous forced-choice question) which account (defence or prosecution) best explained all of the evidence. They were additionally required to provide a think-aloud response justifying their choice in order for us to obtain an insight into the explanatory features that were valued.

Results

Explanatory Preferences The proportion of participants who chose each explanation (defence and prosecution) as the best explanation for the evidence in each condition can be seen in Table 3. A Chi-Square test of independence showed that the distribution of choices differed between the two conditions, $\chi^2(1) = 13.8, p = 0.0002, V = 0.31$. Analogously to our findings in Study 1, here, the majority of participants in the 'control' condition chose the defence's account as the best explanation of the evidence and conversely the majority of participants in the 'draw' condition chose the prosecution's account as the best explanation of the evidence.

Table 3: Number of participant choices in each condition.

	Complex (defence)	Simple (prosecution)
Control	67	43
Draw	36	68

Reasoning underlying explanation preferences To probe the reasoning underlying participants' explanation preferences, we once again analysed their think-aloud responses using the six codes extracted from Study 1 (see Table 2). Results will not be reported in full as they closely mirrored with those of Study 1, in which we found a significant increase in the frequency of the 'simplicity/probability' code in the 'draw' condition compared to the 'control' condition, $p < 0.001$.

Causal models in 'draw' condition Whereas in Study 1 we found that the majority of participants represented the competing accounts as two separate models, in the present study we found that the majority of participants ($n = 59$) represented them in an integrated model (Fig. 2) - and out of these, 88 % preferred the simple explanation. In addition, 17 participants represented them in two separate models Fig. (1) - out of these,72 % preferred the simple explanation. Finally, 25 participants drew three separate models - one for each piece of evidence - (Fig. 3) - out of these, 72 % preferred the complex explanation. A Chi-Square test of Independence showed a significant difference in the distribution of structures used by participants, $\chi^2(3) = 65.4, p < 0.001, V = 0.54$, with the majority of participants representing them in an integrated model. In addition, structure drawn was significantly associated with explanation choice, $\chi^2(3) = 21.6, p < 0.001, V = 0.45$. As in study 1, the majority of participants who drew separate models for each piece of evidence chose the defence's explanation ($p < 0.001$).

Structures of models in Study 1 vs Study 2 Perhaps more of significance, our analysis yielded a significant difference in the frequency with which participants structurally represented the explanations in the two studies, $\chi^2(4) = 35.8, p = 0.001, V = 0.41$. Bonferroni corrected post-hoc pairwise comparisons showed that there was a difference in the percentage of 'combined' models category, $p = 0.003$, the percentage of 'separate models for each account' category, $p < 0.001$ and the percentage of 'separate models for each piece of evidence' category, $p = 0.004$. When participants were presented with the two competing explanations sequentially for all evidence (Study 1), they primarily drew these as two separate causal models. Comparatively, when participants were presented with the two competing explanations simultaneously for each piece of evidence, they primarily drew these in one unified causal model. In addition, the percentage of participants who drew three separate models – one for each piece of evidence – significantly increased in Study 2 compared to Study 1.

Overall, these findings suggest that the manner in which information relating to competing explanations for the same evidence is presented affects how this information is represented in one's own mental causal model which in turn influences how the information is evaluated.

General Discussion

In two studies, we investigated people's preferences for simple vs complex legal explanations and how they represent these explanations in the form of causal models. In addition, we investigated whether representing these in the form of causal models influences how they are evaluated as well as whether structural differences in the causal models drawn influences evaluative practices. Finally, we explored whether the order that information is presented in influences the above-mentioned representational and evaluative processes. Our findings have shown that: (i) drawing causal models influences people's explanatory preferences in favour of the 'simpler' explanation, (ii) drawing causal models influences people's reasoning when evaluating the competing explanations in favour of 'probabilistic' reasoning, (iii) par-

ticipants who draw causal models represent the same information using different structures and (iv) this latter process is influenced by the order that information is presented in

When not required to draw causal diagrams of the information, in both of our studies we observed a preference for the disjunctive 'complex' explanation put forth by the defence. This finding suggests that more parsimonious explanations may not be favoured over complex ones in certain domains (i.e., medical/legal) involving more realistic situations than those typically explored within the psychological research on explanation. This is in line with the complexity-matching hypothesis proposed by Lim and Oppenheimer (2020), predicting that for more complex events, complex explanations are preferred. When analysing the reasoning underlying participants' explanatory preferences for the disjunctive explanation, we found that a meaningful cluster described 'complexity' as a favourable feature when accounting for the evidence and appealed to the fact that the complex explanation was more 'specific' to the evidence. This resonates with the 'opponent-heuristic account' advanced by Johnson, Valenti, and Keil (2019), positing that people use features of complexity in an explanation as a cue for goodness-of-fit and Bayesian likelihood i.e., $P(Evidence|Causes = True)$.

In terms of the simplicity virtue, across both studies, only a small percentage of participants in the 'control' condition referred to this feature when delineating the reasoning behind their explanatory preferences (for the prosecution's account). This was always in combination with probabilistic considerations relating to the fact that the prosecution's explanation is more probable due to only one cause (rather than three) having to be true to bring about the given pattern of evidence. In contrast, in the 'draw' condition of both studies, we found a significant increase in the frequency of reasoning relating to the simplicity and probability of the two explanations. As such, in both studies, drawing causal models of the explanations led to a shift in explanatory preference – in favour of the prosecution's explanation – and in (probabilistic) reasoning. Since we gave our participants no information relating to the prior probability of each of the causes or of the conditional probabilities of the evidence, we are not able to make claims on the normativity of participants' preferences. However, in the absence of explicit probabilistic information, and assuming all things being equal, one should arguably infer that – in line with probabilistic accounts (e.g., Lombrozo, 2007) – the explanation relying on one cause rather than three, is the 'best' explanation for the evidence given that it is likely to be the most probable one. Though the present work provides initial proof for the concept that the quality of people's reasoning is influenced by drawing causal models of the competing explanations under consideration, future work should directly test whether this is in the direction of 'better' reasoning by providing participants with the necessary probabilistic information to enable a comparison of their reasoning against a normative (Bayesian) benchmark.

Though future work is needed to unveil the mechanisms underlying this effect, we propose that graphically representing information using nodes and directed links boosts one's understanding of the relation between the items of information (e.g., independence of causes in the disjunctive complex explanation) and the probabilistic and statistical connotations implied by these relations. This would be in line with studies showing that learning of causal relations improves performance in probabilistic reasoning tasks (Krynski & Tenenbaum, 2007). Drawing causal models allows one to visualise the fact that one explanation needs only one cause to be present to bring about all the evidence, whereas the alternative explanation needs a conjunction of three independent causes– and even but for one of the causes being absent, the pattern of evidence would not be accounted for completely by the explanation. This would facilitate people's inferences relating to the probability of the explanations being true. The fact that participants who drew causal models separately for each account or in a unified way preferred the simple explanation over the complex one, and people who drew three separate models – one for each piece of evidence - preferred the complex explanation, suggests that certain structures particularly facilitate deliberations on simplicity and probability. As such, the latter two structural representations comprise one root node for the simple explanation and three root nodes for the complex explanation whereas the former method comprises three root nodes for each explanation (albeit the prosecution's explanation would have three of the same nodes to represent smothering). More research, however, is needed to establish whether the influence of drawing the diagrams is thus about the number of causes of multiple effects, or whether it extends for other types of causal structures and more broadly.

Future studies should use alternative comparison groups that are more closely matched on factors such as depth and time of processing to the 'draw' condition. This would allow us to rule out that the observed differences in explanatory preferences between participants who were drawing causal models and those who were simply reading the information and not engaging in any activity, are not due to this condition being more engaging and allowing for increased processing time. In addition, although in our experiments only a small number of participants (see Table 2) cited reasoning related to reluctance for a mother to kill a child ('no motive/intent' code), additional studies should also replicate these findings utilizing scenarios that vary in terms of emotional valence and 'moral load'. This would help increase the generalizability of our findings that so far is limited due to our hypotheses being tested on only one,particularly morally loaded, legal scenario.

In terms of the causal structures drawn to represent the explanations, our findings indicate that individuals do not uniformly represent these in a unified framework. As such, although Bayesian models of legal explanations have so far mostly adopted an 'integrated' approach, representing in a single unified model of all the arguments under consideration (Fenton, Neil, & Lagnado, 2013), we have shown that people

represent competing explanations in a variety of ways when asked to draw their own causal models. In addition, we found that the type of causal structure drawn is associated with participants' choice of which explanation best explains the evidence, reinforcing the view that causal structure plays a key role in inferential processes. These findings imply that, even when the individuals i.e., jurors learn the same arguments in the same exact manner, they can represent these in different mental models, and ultimately this might lead to different inferential and evaluative processes. Finally, we showed that the manner in which the competing explanations are presented to participants (simultaneously for all the evidence or sequentially for each piece of evidence) influences the causal structure that is drawn. Future work modelling legal explanations should therefore elicit the models of the reasoners involved in order to optimise the development of normative solutions to the problem at hand, and in order to understand the causal structures that underlie the inferences and judgments being made. This would also help to elucidate whether shortcomings in reasoning are the product of skewed mental models.

References

Chater, N., & Vitányi, P. (2003). Simplicity: A unifying principle in cognitive science? *Trends in cognitive sciences*, *7*(1), 19–22.

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.

Fenton, N., Neil, M., Lagnado, D., Marsh, W., Yet, B., & Constantinou, A. (2016). How to model mutually exclusive events based on independent causal pathways in bayesian network models. *Knowledge-Based Systems*, *113*, 39–50.

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using bayesian networks. *Cognitive science*, *37*(1), 61–102.

Hastie, R., & Pennington, N. (2000). Explanation-based decision making.

Hung, W., & Jonassen, D. H. (2006). Conceptual understanding of causal reasoning in physics. *International Journal of Science Education*, *28*(13), 1601–1621.

Johnson, S. G., Valenti, J., & Keil, F. C. (2019). Simplicity and complexity preferences in causal explanation: An opponent heuristic account. *Cognitive psychology*, *113*, 101222.

Krynski, T. R., & Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology: General*, *136*(3), 430.

Lim, J. B., & Oppenheimer, D. M. (2020). Explanatory preferences for complexity matching. *PloS one*, *15*(4), e0230929.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive psychology*, *55*(3), 232–257.

Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, *20*(10), 748–759.

Mackonis, A. (2013). Inference to the best explanation, coherence and other explanatory virtues. *Synthese*, *190*(6), 975–995.

Neil, M., Fenton, N., Lagnado, D., & Gill, R. D. (2019). Modelling competing legal arguments using bayesian model comparison and averaging. *Artificial intelligence and law*, *27*(4), 403–430.

Pacer, M., & Lombrozo, T. (2017). Ockham's razor cuts to the root: Simplicity in causal explanation. *Journal of Experimental Psychology: General*, *146*(12), 1761.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 521.

Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*(3), 429.

Smit, N. M., Lagnado, D. A., Morgan, R. M., & Fenton, N. E. (2016). Using bayesian networks to guide the assessment of new evidence in an appeal case. *Crime science*, *5*(1), 1–12.

Waldmann, M. R., Hagmayer, Y., & Blaisdell, A. P. (2006). Beyond the information given: Causal models in learning and reasoning. *Current Directions in Psychological Science*, *15*(6), 307–311.

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children's preference for simpler hypotheses. *Psychonomic bulletin & review*, *24*(5), 1538–1547.

Zemla, J. C., Sloman, S., Bechlivanidis, C., & Lagnado, D. A. (2017). Evaluating everyday explanations. *Psychonomic bulletin & review*, *24*(5), 1488–1500.