Drawing conclusions: Representing and evaluating competing explanations

Alice Liefgreen[1 *] and David A. Lagnado[1]

[1]Department of Experimental Psychology, University College London, 26 Bedford Way, WC1H 0AP, London, UK.

[*]Corresponding Author

Author Note:

The authors declare no competing interests. Data and materials have been made publicly available via the Open Science Framework at https://osf.io/4evxn/?view_only=3155b287f51243b2a11649c86dc24eb5. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

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, their explanatory preferences are affected by whether they are required to draw causal models of the evidence. In addition, we identify 'mechanism' as an explanatory feature that people value when evaluating explanations. Although previous research has shown that people can reason correctly about causality, ours is one of the first studies to show that generating and drawing causal models directly affects people's evaluations of explanations. Our findings have implications for the development of normative models of legal arguments, which have so far adopted a singularly 'unified' approach, as well as the development of modelling tools to support people's reasoning and decision-making in applied domains. Finally, they add to the literature on the cognitive basis of evaluating competing explanations in new domains.

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

1. Introduction

In December 1996, the son of Sally and Stephen Clark died at the age of 11 weeks old – a death that was initially ruled as being due to natural causes, until the couple's second son died at eight weeks of age in 1998. At this point, a re-examination of the evidence of the first son's death was issued and Mrs Clark was tried and convicted for the murder of both children in 1999. This became one of the most famous miscarriages of justice in UK history, only resolved by a quashed conviction on second appeal (see Clark, R v 2003 précis for details). Sally's case demonstrates the deleterious consequences that fallacies in probabilistic and causal reasoning can generate within the criminal justice system[1]. In addition, on a broader scale, it showcases the complexities that juries must face when reasoning during a criminal trial.

Imagine being a member of the jury in Sally's trial. You are told that the medical examination on the first son revealed three distinct injuries – blood in the lungs, a torn frenulum (tissue between lip and jaw) and bruises on the arms and legs. In order to reconstruct *what happened*, what led to those injuries, you will draw on your own understanding and prior knowledge of how those injuries could occur to an infant – as well as your general intuitions about the case. Imagine that the prosecution lawyer then presents an explanation as to how those injuries occurred: Sally Clark caused all of them by smothering the child. You listen to expert witnesses and medical reports that support this explanation. Imagine that the defence lawyer then presents you with a *different* explanation for how those injuries occurred: they were caused by three independent incidences – all natural or accidental. You listen to different expert witnesses and reports that support the defence's explanation. Both sides offer *plausible* explanations, but only one of the two can be true. How do you evaluate them and determine whether the prosecution's explanation is compelling enough to lead to a verdict of *guilty*? What factors would you consider when comparing the two explanations? Evaluating and comparing competing explanations of the same evidence is not a trivial task – and the consequences of sub-optimal evaluation and

[1] See Nobles & Schiff (2005) and Saini (2009) for details on statistical/probabilistic errors that led to the miscarriage of justice. See also Lagnado (2021) for discussion of the causal and probabilistic reasoning in the case.

reasoning in these instances can be extremely damaging, as demonstrated by Sally Clark's case. This makes the above questions pressing ones to answer.

1.1. Simplicity and complexity in causal explanation

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, such as simplicity, coherence and breadth (for overview see Lipton, 2004; Mackonis, 2013). In brief, *coherence* relates to the consistency between an explanatory hypothesis and the relevant background knowledge. The more consistency there is between the two, the better the explanation. *Breadth* refers to how unifying an explanation is, i.e., how well it can explain a variety of different items. The more different kinds of items a hypothesis can explain, the better it is at predicting new items, and the more unifying it is. 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).

Ironically, simplicity is one of the most complicated explanatory virtues to study and comprehend given that there are multiple ways in which an explanation can be defined as *simple* (e.g., ontologically, syntactically, structurally; see Niiniluoto, 1999). 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 less flexible[2] (Blanchard, Lombrozo, Nichols, 2018; Lombrozo, 2007; 2016). Within psychology, only a few studies have directly investigated how people evaluate competing explanations that differ on some measure of simplicity. One of these studies tested a measure of simplicity – supported by philosopher Paul Thagard (1989) – according to which simpler explanations are ones that make fewer 'assumptions' (Read & Marcus-Newhall, 1993). In this study, participants were given a scenario in which a patient who suffers from three symptoms (recent nausea, weight gain and fatigue) could be either a) pregnant – single assumption – or b) suffer from a conjunction of three

[2] The notion behind a complex explanation being more flexible (and therefore a simple explanation being less flexible), is that a complex explanation in this sense has greater number of possible parameters or "free parameters" and can accommodate a wider array of possible data by adjusting or fine-tuning the parameters of the explanation. This means that for many parameter settings, the explanation hypothesis fits the actual data poorly.

distinct issues (stomach virus, lack of exercise and mononucleosis) – multiple independent assumptions– and asked to evaluate these explanations. Participants favoured a single assumption in the conjunctive explanation when the patient had only a single symptom (e.g., they preferred to explain *nausea* by appeal to a *stomach virus* when nausea was the only symptom) but they also favoured strongly the simpler common-cause explanation when all three symptoms were present. Although seemingly revealing a preference for simpler explanations – when simplicity is measured as number of independent assumptions – this effect could also be accounted for by probabilistic inference. As such, if all causes are equally rare, all else being equal, probabilistically one *should* favour the simpler explanation given that the probability of a conjunction of three rare occurrences is less than the probability of a single rare cause, i.e. P (Cause 1) > P (Cause 2, Cause 3, Cause 4). Research has confirmed that when probabilities of explanations are explicitly stated, participants do favour the most *likely* explanation irrespective of the number of causes it involves (Lagnado, 1994). However, when there remains uncertainty on the probabilistic parameters underlying the explanations, preferences are for simpler explanations, or a function of both simplicity and probability (Blanchard et al., 2018; Lombrozo, 2016). Blanchard et al. (2018) showed that people's intuitive probabilistic and explanatory judgments are sensitive to a different form of simplicity, namely inflexibility (i.e., hypotheses do not accommodate a wider range of possible data), in a way that is also consistent with a Bayesian account of inference.

Overall, despite disagreement on a formal definition of simplicity, there has been widespread empirical agreement that simpler explanations are preferred and are found more satisfying than complex explanations in a wide array of settings (Blanchard et al., 2018; Bonawitz & Lombrozo, 2012; Chater & Vitanyi, 2003; Lombrozo, 2007; 2016; Pacer & Lombrozo; 2017; Walker, Bonawitz, Lombrozo; 2017). More recent studies, however, have painted a more nuanced picture of people's explanatory preferences for simplicity. Zemla, Sloman, Bechlivanidis & Lagnado (2017) have shown that explanatory virtues such as simplicity are not good predictors of an explanation's quality in naturalistic settings e.g., when testing real-world explanations found on *Reddit*. Rather, they found that when evaluating these types of explanations, people prefer complexity – likely arising from a preference for explanations that invoke more causal mechanisms to explain an event/effect. In similar vein, Zemla, Sloman, Bechlivanidis & Lagnado (2020) showed that a preference for simplicity is influenced by the presence of mechanisms in an explanation. Whereas without mechanism the preference for simplicity over complexity holds, this preference is reversed or mitigated when explanations contain details of mechanisms underlying cause-effect relations. 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. Thus, people will prefer simple explanations for simple events and more complex explanations for more complex events. Although simple explanations might typically have a higher prior probability, complex explanations might have a higher Bayesian likelihood – thus fitting complex events better. Managing the trade-off between probability and likelihood is arguably what leads people to use and prefer more complex explanations as the complexity of the event increases (Johnson, Valenti and Keil, 2019). Generally, in certain situations, it seems reasonable that more complex explanations – or those that contain multiple mechanisms – are preferred if they substantially increase one's sense of understanding and provide a fuller account of the situation. Mechanisms represent qualitative understanding as they account for how and why a given cause produces the effect (Thagard, 2000). In many cases, we think of explanations as providing some sense of mechanism (Glennan, 2002). When comparing explanations, we therefore also ask questions about the plausibility of the mechanisms underlying the cause-effect relations – which can be driven by our familiarity of similar relations and prior knowledge (Keil, 2006). In a series of studies, Ahn and Khalish (2000) showed that people interpret patterns of association i.e., between factors and effects, in light of their beliefs about mechanisms. Thus, though particularly crucial when considering complex explanations, the mechanism appears to be an explanatory feature of substantial significance – though less studied than features such as simplicity and complexity. In our studies, we investigated through mixed methods, whether people naturally consider the mechanisms underlying cause-effect relationships when evaluating explanations.

While there is a substantial amount of psychological research assessing people's preferences for simple and complex explanations – and more recently, on the role of mechanism in mediating this preference – 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. This is something we will explore in the present research. Think back to Sally Clark's case presented at the beginning of this paper. As a juror on the case, you would have been required to evaluate competing explanations offered to you by the prosecution and the defence attorneys, which varied in degrees of simplicity (as well as other features). What explanatory features would you have valued when comparing the legal narratives of "what happened"? Given that how legal explanations are evaluated and compared, and how their merits are established, can ultimately determine a person's fate, this is a critical question to answer. The answer to this question might differ from answers derived using more abstract and constrained laboratory tasks given that, unlike most of the explanations tested in these tasks, legal explanations often concern motivated human behaviour. When participating in laboratory tasks using more abstract explanations, people might be less motivated to construct causal models of the information and to understand the causal relations involved and the mechanisms underlying these – compared to when reasoning in a legal case that describes human goals and actions unfolding in everyday physical environments. Exploring these questions within a legal sphere will therefore allow us to appraise the specificity of people's explanatory preferences – in a unique context in which people routinely make judgments of blame and culpability.

1.2. Evaluating legal explanations: descriptive and formal approaches

In their notable 'story model', Pennington and Hastie (1988) argue that evidence evaluation in complex diagnostic decisions such as juror decision-making is explanation-based, and that these explanations take the form of *stories*. Decision-makers construct a causal narrative to explain the available evidence at the outset, and subsequently base decisions on the causal interpretation they impose on the evidence (Pennington & Hastie, 1988; 2000). These causal narratives are constructed by combining relevant world knowledge as well as one's expectations of what would be an adequate explanation in the context of the decision domain. This includes evaluating an explanation on features including coverage, coherence and uniqueness (Pennington & Hastie, 2000). Coverage refers to the amount of evidence accounted for by a particular story. Coherence is determined by a story's consistency, plausibility and completeness. Finally, uniqueness refers to the presence (or absence) of other acceptable stories that could account for the evidence. Some of these factors overlap with the 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 research has established that juror's explanations of legal cases take the form of narratives which feature causal relations among events (Pennington & Hastie, 1986; 1988). These representations are comparable to the situation models proposed by van Dijk and Kintsch (1983) and the mental models proposed by Johnson-Laird (1983). These causal stories are spontaneously constructed by jurors and seem to mediate verdict decisions (though for discussion of this point see Vorms & Lagnado, 2019). 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 –which also influenced perceptions of evidence strength as well as confidence in verdicts (Pennington & Hastie, 1986; 1991). This suggests that order of evidence presentation is an important factor to consider when investigating how people construct and evaluate causal representations of legal evidence. Although in their studies, Pennington and Hastie proposed informal casual networks of the most prevalent stories given 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 (e.g., forensic analyses, medical diagnosis – see Smit, Lagnado, Morgan & Fenton, 2016; Costantinou, Fenton, Marsh & Radlinksi, 2016; Boneh et al., 2015; Zhang, Liu & Pang, 2018). Evaluating a single explanation and comparing multiple explanations are challenging tasks, however, and even formal approaches to evaluation do not provide a clear-cut metric for an explanation's quality (Neil, Fenton & Lagnado, 2019). Formal tools that have been used by researchers in a legal domain primarily include logic-based systems, (Zarri, 1998; Nissan, 2008), argumentation systems (see Prakken, 2004), and causal Bayesian networks (CBNs; Fenton & Neil, 2018; Fenton et al., 2013; Thagard, 2004).

Causal Bayesian networks (Pearl, 2000/2009) are comprised of two components: i) a causal graph that captures the qualitative causal relations between variables, and ii) a quantitative component that captures the probabilistic strength of these relations, and how multiple causes interact. In the causal graph, nodes represent variables, and arrows (directed links) represent the causal relations between these variables. For example, the graph of the prosecution's explanation for three of the injuries mentioned in the Sally Cark case is seen in Figure 1 (see Lagnado, 2021, for causal modelling of the wider case).

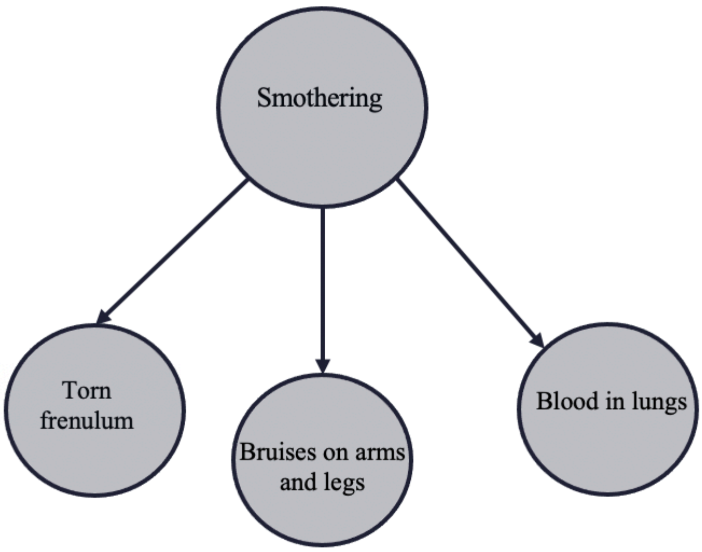


Figure 1. *Causal graph of the prosecution's account of how the three injuries found on the baby were caused.*

In this model, the node 'Smothering' is the cause of the three injuries, which are each represented as separate nodes. The arrow directed *from* 'Smothering' *to* the injury nodes indicates that the former is a potential *cause* of the latter nodes. Behind the arrows or directed links there are probabilities – the computational constituent of the network. As such, CBNs use conditional probability tables to quantify the relations between the variables (Fenton & Neil, 2018). To parameterise a network, one must specify the probability of each state of each node (i.e., 'blood in lungs' = True/False) given each state of its *parent* node (the node it is causally linked to i.e., 'smothering' = True/False). After specifying these probabilities, one can compute the probability of any node being true (given knowledge of any other variables) via Bayesian updating (Pearl, 1988). Once set up these networks can thus be used to make predictive, diagnostic, and counterfactual inferences. Though the underlying mathematical engine is crucial, the *structure* of the causal graph is also critical in guiding these inferences. Different graphical representations of the same information can yield different normative answers to the same probabilistic queries. The graphical modelling characteristic of these networks – though less researched – is thus arguably just as fundamental as the probabilistic machinery that underlies it.

Most Bayesian models of legal explanations developed so far have adopted a ‘unified’ approach, aiming to represent in a single model all of the arguments under consideration, such as those presented by the defence and prosecution in a trial (Aitken & Taroni, 2004; Fenton et al. 2013; Taroni et al. 2014). Fenton et al. (2016) however, have shown that this unified 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. In addition, representing competing arguments in a combined model assumes that the reasoner – in our running example this would be a jury member – is able to rationally combine all relevant information, even though this is often an iterative process that involves learning about information consecutively and updating one’s model(s) multiple times. A recent attempt to formally model and evaluate competing legal arguments when these are represented using separate CBNs is given by Neil, Fenton, Gill & Lagnado (2019), but this remains untested. This disjunctive approach allows one to account for the differences in variables and causal dependencies that the two arguments may contain. It also attempts to model the arguments from the perspective of a juror who observes the different arguments and facts presented separately by each adversarial side.

Though this work advances 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 they represent them using a unified approach or a disjunctive approach? Would the order of evidence presentation affect the causal structures drawn and the inferences made based on these? Most of the research on formal models of legal arguments has focused on the comparative aspect and less on the representation and integration aspects (Bex et al., 2007; 2010; Fenton et al., 2016) despite these being crucial to accurate reasoning. In this paper, we present studies that directly elicit the causal models that people construct and compare in legal domains. The findings of our studies can inform the development of formal tools that are able to support people's evidential reasoning and decision-making in these high-stake contexts (Mackor, Dahlman & Lagnado 2022).

The efficacy of causal models in helping people reason in real-world-like situations has not been studied in any depth. Previous research has shown that the use of visual displays such as influence diagrams to teach people about causal relationships of a process can improve performance when tested on that process (Hung & Jonassen, 2006). Zheng, Marsh, Nickerson, and Kleinberg (2020) studied how causal models may be used to support people's decision-making in health management. They showed that when given causal models, individuals without experience in the domain of interest, made more accurate decisions regarding that domain, whereas more knowledgeable people made poorer judgments (possibly because the given models were too simplistic and didn't correspond well with their prior models; Mackor et al., 2022). Relatedly, in a study utilising clinical scenarios, patients and laypeople showed better comprehension of causal information about treatments for Generalized Anxiety Disorder when using a causal model to accompany an auditory presentation than when given the auditory presentation alone (N. Kim et al., 2013). In this study, causal models were built using a flexible drawing tool ConceptBuilder, that has been used to successfully elicit the causal models of decision-makers in various contexts (N. Kim & Park, 2009; N. Kim, Luhmann, Pierce, & Ryan, 2009; Chen & Urminsky, 2019; Morais, Schooler, Olsson, & Meder, 2014). Research has also shown that learning causal structures improves probabilistic reasoning in learning, problem-solving and categorisation tasks (Krynski & Tenenbaum, 2007; Waldmann, Hagmayer, & Blaisdell, 2006). An effective way to convey information about causal structure is via visual causal models, which illustrate the relationship between items of information using arrows (Heiser & Tversky, 2006). Heiser and Tversky argued that visual models improved judgments because the presence of arrows primed people to pay particular attention to detecting functional or causal relationships when approaching new information. Similarly, causal models are argued to be effective as they support long-term memory, facilitate information processing, organize thoughts, and promote inference and discovery (B. Tversky, 2011). Hayes et al. (2018)'s study results suggest that while providing causal explanations does not result in correct normative judgments, it can still help improve people's causal models by drawing attention to the statistical information which gets incorporated into causal structures.

Overall, however, work in psychology has focused primarily on understanding how people learn causal structures, rather than how people use causal models in real-world decision-making tasks. The studies that have been carried out on the role of causal models in supporting decision-making and probabilistic inference, show promising findings that merit further study as building causal models could prove to be an effective means to support people's reasoning in real-world diagnostic tasks. Hence in the present work, we explore the influence of drawing causal models of competing legal explanations – on the explanatory preferences exhibited by participants when the competing explanations differ in terms of simplicity'. We expect that, although a more complex explanation might be preferred for a 'complex' event (Lim & Oppenheimer, 2020), drawing causal models might shift people's preferences to a simpler explanation by rendering the statistical relations implied by the causal models more explicit. This will occur because, all else being equal, a simpler explanation is more probable (Lombrozo, 2007; 2016). As such, we believe that using a visual causal model to externalize their reasoning helps people visualise and evaluate the model's assumptions – emphasising certain aspects, such as the model's plausibility and probability.

1.3. Present Work

Legal cases often require the factfinder to evaluate two competing accounts of *what happened* in a given instance. Was the baby smothered? Was his death accidental? This is not a trivial task and would undoubtedly benefit from the development of support tools with an underlying formal (Bayesian) causal reasoning framework. However, to help develop an effective support tool, we need to understand *how* people structurally represent this information – without assuming they adopt a unified approach. Research has shown that how people structure the causal relations between items of information in a clinical domain predicts how they search for new information (Morais, Olsson & Schooler, 2011). Relatedly, Pennington and Hastie (1988) showed that different verdicts in mock juror decision-making tasks were underpinned by different causal narratives. Thus, different structural representations of legal explanations can lead to different evaluative processes and decisions. Throughout the introduction, we presented various factors known to influence how people build causal representations and reason with causal structures – such as simplicity of the represented information (e.g., Johnson et al., 2019; Lombrozo, 2016; Walker et al., 2017), the presence of mechanisms (e.g., Ahn & Khalish, 2000, Zemla et al., 2020) or the order of presentation of information (e.g., Hogarth & Einhorn, 1992; Pennington and Hastie, 1988). In the present research, we investigated whether some of these factors, recounted in the psychological and philosophical literature, also influence how explanations are represented and evaluated within a legal domain.

More specifically, in this paper we present three studies in which we probe: 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. In Study 1 and Study 2, we compared participants who were asked to draw causal models of the explanations to participants who were required to describe the explanations. In Study 3 we replicate our findings in a different legal scenario, and counterbalancing which adversarial side (prosecution or defence) puts forth the simple explanation (vs. complex explanation). Prior to running these three studies, we ran two "proof-of-concept" preliminary studies which followed a similar methodology to that used in Study 1 and 2, but with a control condition that was less matched to our experimental condition (participants had to either draw the causal models of the explanations, or simply read the explanations). Findings from these preliminary studies, which closely match those of Studies 1 and 2 reported in this paper, can be read in Liefgreen and Lagnado (2021). Materials and data relating to all of the studies presented in this paper can be found on the project's OSF page: https://osf.io/4evxn/?view_only=3155b287f51243b2a11649c86dc24eb5.

2. Study 1 (Sequential presentation of explanations)

In Study 1, we explored how people graphically represent two competing explanations of the same evidence when the complete explanations are presented sequentially (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.

2.1. Methods

2.1.1. Participants

139 participants (Mean age = 35.2, SD = 11.4; *n* females = 94) 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). No participant dropped out of the study. All studies presented in this paper were processed by the Research Ethics Committee at University College London (project number: EP/2017/005).

2.1.2. Materials and Procedure

A between-subjects design was employed. As in our previous studies, 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, hereafter referred to as: 'describe' (*n*=70) and 'draw' (*n* = 69).

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. This introduction comprised of explaining what causal models were, and how they could be used to graphically represent information (see Appendix 1 for full instructions). For example, participants were told that in order to represent the information: "rain and a sprinkler can both make the grass wet" in the form of a causal model, you would firstly identify the elements of interests (nodes: RAIN, SPRINKER, WET GRASS) and use arrows to represent the relations between the variables i.e., RAIN and SPRINKLER as causes of the effect, WET GRASS. This would result in a causal model with two nodes (RAIN and SPRINKLER), and arrows from each of these nodes leading to a third node (WET GRASS).

Next, they were introduced to the online tool that they would be required to use during the task to draw their own causal models: Loopy[3] (https://ncase.me/loopy/). In order to learn how to use this tool they were shown examples of three causal structures drawn in Loopy: a two-node cause-effect model, a three-node common-cause model and a three-node common-effect model and asked to replicate them in a new page in Loopy and paste the links to their models in the designated text box on the survey page in Qualtrics. The final stage of the learning/practice block included drawing a causal model of the following information: "*Tom has a cough. The doctor thinks that it could be a symptom of either asthma or the flu*", saving the model, and pasting the link to their model in the designated text box on the survey page. For this practice task, participants were instructed to label their nodes meaningfully.

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:

> "*Eva and Theo Glaser are a married couple living together in a small town outside London. They have been married for 7 years before their first son, David, was born. Tragedy struck one evening when David was only 11 weeks old, and Eva was alone with the baby. Her statement read that she gave the baby a feed and placed him in his basket as she did each night. Soon afterward he became unwell and appeared to stop breathing. Eva called an ambulance, but baby David could not be saved. Eva is now the prime suspect in the death of David"*

After reading the case briefing, they were informed that the following distinct medical findings had been recorded after examining David's body:

1) *Bruises on the arms and legs*
2) *Torn lingual frenulum (fold of tissue attaching tongue to floor of mouth)*
3) *Fresh blood in the lungs*

[3] 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/).

These injuries were consisted with those present in the real Sally Clark case[4], though in the real case they comprised only a subset of the total evidence. After having received this information, participants in the 'draw' condition were instructed that there were two competing accounts to explain the evidence – one put forth by the prosecution and one by the defense. The two accounts were presented to participants sequentially, in counterbalanced order. The prosecution account posited that: *Smothering* (purposeful suffocation) caused the bruises, *Smothering* (purposeful suffocation) caused the torn frenulum and *Smothering* (purposeful suffocation) caused the blood in the lungs. The defence account 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. Thus, they were presented with a 'simple' common-cause explanation of the evidence, and a competing 'complex' explanation of the evidence comprised of multiple independent causes for each piece of evidence.

After reading about the first account, participants in the 'draw' condition were required to represent the information as a causal model in Loopy and paste the link in the designated box. Subsequently, they viewed the second account 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) and pasting the link to the causal model in the designated box. 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. To re-iterate, in the 'draw' condition, half of the participants participants saw the prosecution's account first and drew a causal model of it, the other half participants saw the defence's account first and drew a causal model of it, and all 139 participants drew another causal model/updated their initial model to include *all* the information i.e., comprising both accounts.

[4] We used a subset of the evidence and explanations relating to the first son's death in the Sally Clark case, though all names were changed to ensure participants did not have prior knowledge about the explanations or the case more generally.

After completing 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 reminded of the two accounts before answering this question. They were also asked to provide reasoning for their choice (think-aloud responses) 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 pre-determined 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 of the evidence (on two separate pages). In this condition, participants were required to describe and/or summarise information after it was presented. As such, after having seen the first explanation for the evidence (randomised order of prosecution and defence explanation) participants in the 'describe' condition were asked to: "*Please write, in your own words, a description or summary of the explanation of the evidence viewed so far*" in a text box with a minimum requirement of 100 characters. This was additionally asked subsequently to viewing the alternative explanation. Next, they were required to choose the best explanation for the evidence and provide reasoning for their choice. For a graphical representation of the procedure participants in each condition followed see Figure 2. All participants were de-briefed at the end of the task.

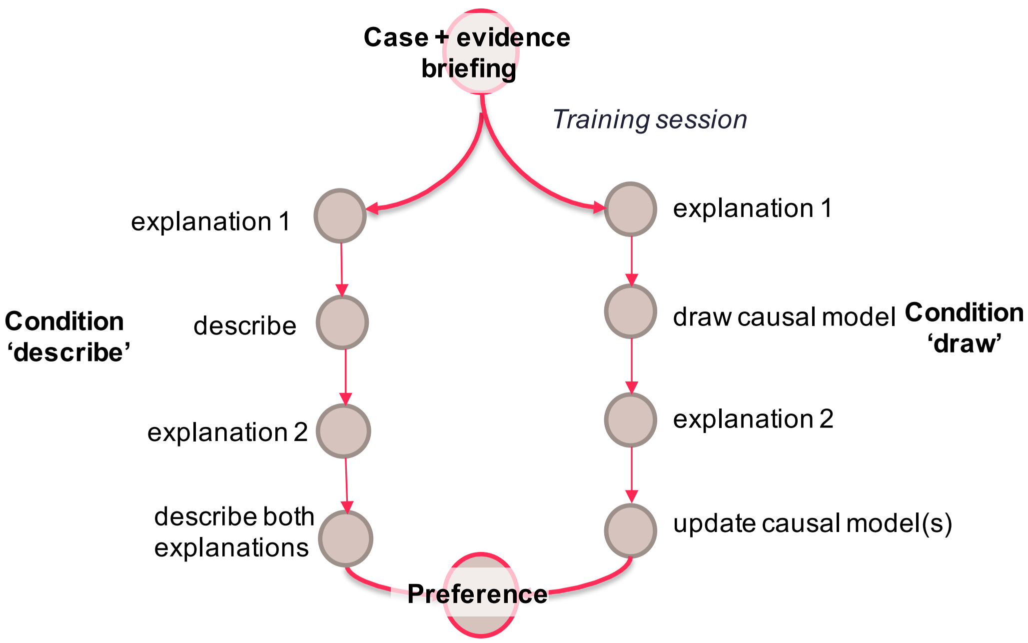


Figure 2. *Study 1: graphical depiction of experimental procedure in 'draw' and 'describe' conditions.*

2.2. Results

2.2.1. Explanation choice

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

Table 1

Study 1: Percentage of participants who chose each explanation in each condition.

Condition	Choice, *N (%)*		**Total**
	Complex (defence)	Simple (prosecution)	
Describe	39 (55.7%)	31 (44.3%)	**70**
Draw	21 (30.4%)	48 (69.6%)	**69**

A Chi Square test of independence revealed a significant difference in the distribution of participants' choices between the two conditions, χ^2 (1) = 8.05, p = 0.005, φ = 0.25. As can be seen from Table 1, the majority of participants in the 'draw' condition chose the prosecution's explanation as the best explanation for the evidence[5]. Conversely, participants in the 'describe' condition did not display a clear preference for either the simple explanation put forth by the prosecution or the complex explanation put forth by the defence[6]. These findings echo those described in Liefgreen and Lagnado (2021) and suggest that drawing causal models influences evaluative processes when comparing competing explanations. Given the use of a better-matched control condition, compared to that used in Liefgreen and Lagnado (2021), findings of Study 1 suggest that the effect of drawing causal models cannot be entirely explained by increased engagement and time efforts in the 'draw' condition.

2.2.2. Reasoning (think-aloud responses)

[5] A binomial test within the 'draw' condition revealed a significant preference for the prosecution's explanation, p = 0.0007.

[6] A binomial test within the 'describe' condition did not reveal a significant preference for the prosecution's explanation, or the defence's explanation, p = 0.2.

To probe the reasoning underlying participants' explanation preferences and what explanatory features they valued, we analysed their think-aloud responses. Participants' think-aloud responses were qualitatively analysed and coded with a single code that simultaneously categorized, summarised and accounted for the response (Charmaz, 2006) by a rater blind to both condition and the choice associated with each response. Each think-aloud response was therefore attributed a code, from a pool of codes that were drawn directly from the response and not a pre-existing set (e.g., was conjured to describe and summarise a response rather than chosen from a pre-existing set of codes), which acted as a descriptive label of participants' reasoning. The codes drawn from our participant sample, with a description of each, and its percentage in each condition can be seen in Table 2.

Table 2

Study 1: Reasoning codes with descriptions and percentage of codes across the two conditions.

Reasoning Code	Description of Code	Example participant response	Draw condition (%)	Control condition (%)
Complexity/ Specificity	Reasoning appealed to greater number of causes in explanation, greater specificity to evidence.	*"They gave a more detailed explanation."*	4.3	12.9
Mechanism	Reasoning questioned mechanism underlying proposed cause-effect relations in explanation.	*"I do not think that bruising can occur post-mortem and would be surprised if resuscitation attempts could cause a torn frenulum."*	23.2	22.9
No Intent/Motive	Reasoning referred to the lack of intention or motive for killing baby.	*"I believe the mum was innocent without motive."*	10	8.6
Probability	Reasoning appealed to likelihood of explanation.	*"Seems to be more likely."*	8.7	15.7
Simplicity / Probability	Reasoning appealed to smaller number of causes of explanation and greater likelihood of explanation given this.	*"Seems more likely if one thing was the cause of all observed injuries."*	40.6	12.9
Fits with Evidence	Reasoning relating to the fact that a given explanation seemed like	*"Fits better I think with what evidence was given."*	8.7	14.3

	a good fit for the evidence, without giving any details as to why.			
Other	Reasoning did not fall under any of above categories or was not elaborate enough to code.	*"It's the least horrible of the options."*	4.3	12.9

Fischer's exact test illustrated a significant difference in the distribution of reasoning codes between conditions, $p < 0.0001$. Bonferroni corrected post-hoc comparisons revealed the only significant difference to be that participants in the 'draw' condition employed reasoning that fell under the 'simplicity/probability' code significantly more than participants in the 'describe' condition, $p = 0.003$. This finding suggests that the act of drawing causal models – rather than merely describing the explanations more – increases the prevalence of reasoning that relates to the probabilistic connotations of causal structures underlying the competing explanations. For the percentage of codes underlying each explanation choice within each condition see Table 3.

Table 3

Study 1: Percentage of participants within each reasoning category who chose each explanation, per condition.

	Draw condition		Describe condition	
Reasoning Code	Complex (defence) explanation (%)	Simple (prosecution) explanation (%)	Complex (defence) explanation (%)	Simple (prosecution) explanation (%)
Participant Choice	***30.4***	***69.6***	***55.7***	***44.3***
Complexity/Specificity	14.3	0	23.1	0
Mechanism	23.8	22.9	15.4	32.2
No Motive/Intent	33.3	0	15.4	0
Probability	4.8	10.4	20.5	9.7
Simplicity/Probability	4.8	56.3	0	29
Fits with evidence	19	4.2	17.9	9.7
Other	4.8	4.2	7.7	19

2.2.3. Causal models (in Drawing Condition)

Next, we take a closer look at the causal models elicited from participants in the 'draw' condition, to evaluate their variability and ascertain whether structural differences in the models influenced explanatory preferences. Given that we are interested in whether representing competing explanations as intergrated or disjunctive influences how these explanations are evaluated, we only present analyses relating to participants' final models below. All final models were coded according to the features listed in Table 4. We focus on what *structure* participants used to represent the two competing explanations. Three models were coded as "n/a" given they were not complete. For the number of participants who drew each type of structure see Figure 3.

Table 4

Coded features of causal diagrams with description.

Coded Feature	Description
Number of nodes	Number count of nodes in causal model(s).
Structure	Whether explanations were represented as: (A) two models – separate for each account, (B) one unified model, (C) three separate models – one for each evidence piece, or (D) whether their representation didn't fit in any of these categories.
Causal Direction	Whether the causal direction of the links (flowing from cause → effect) was correct (1) or incorrect (0).
Nodes Missing (number)	Whether some variables in the model(s) were not represented (number count)[7].
Nodes Missing (names)	Names of variables that were not represented.
Nodes Added (number)	Whether some extra variables were added to model(s) – measured as number count.
Nodes Added (names)	Names of variables that were added to model(s).
Mechanism	Whether mechanisms were added to the model(s). For example, whether extra nodes were added between cause-effect nodes to provide details of mechanism.

[7] Considering the 3 items of evidence and the 4 causal models (3 in the defence's account and 1 in the prosecution's account), participants' models should have contained at least 7 nodes.

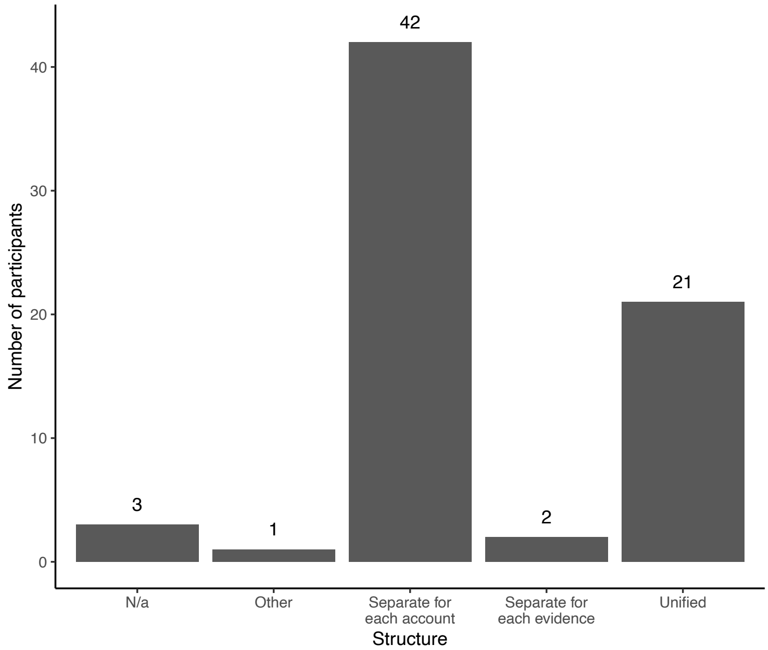


Figure 3. *Study 1: Frequency (count) of model structures in 'draw' condition.*

Fisher's Exact Test showed a significant preference for certain structures used by participants, $p < 0.001$. As such, we found that when explanations are presented sequentially, the majority of participants (60.8 %) represents these as two separate models – one for each account. In addition, 30.4 % represented them in a unified model, 2.9% drew three separate models (one for each piece of evidence) and 1.4 % of models were classified as 'other' given they did not fall into any of the above categories.

See Figures 4-6 for examples of the causal models that participants drew, coded as each of the main categories described in Table 4.

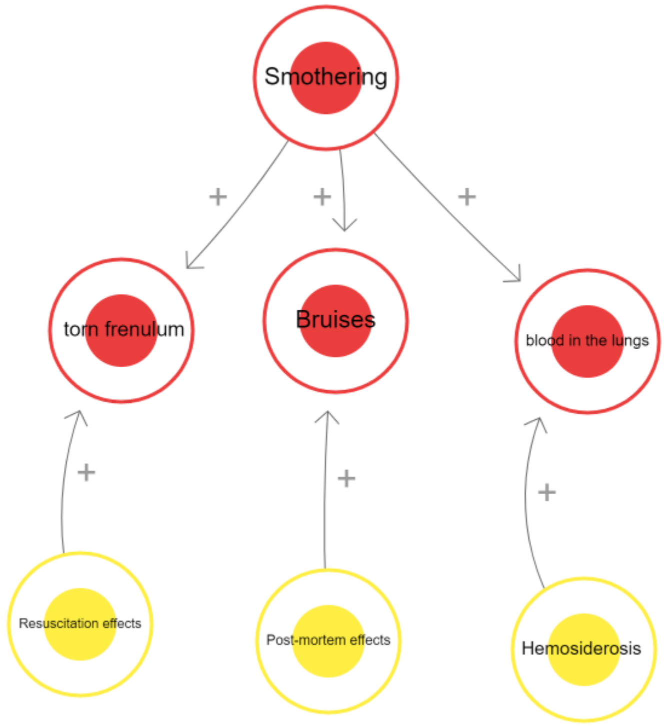


Figure 4. *Study 1: Participant drawn 'unified' causal model of defence and prosecution accounts of the evidence.*

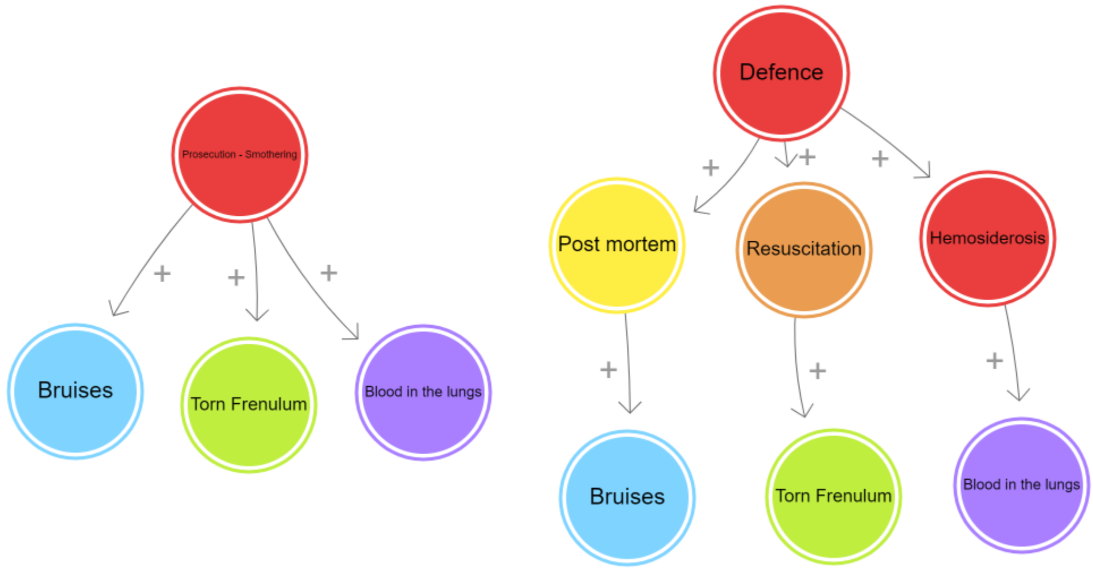


Figure 5. *Study 1: Participant drawn (separate) causal models of defence and prosecution accounts of the evidence.*

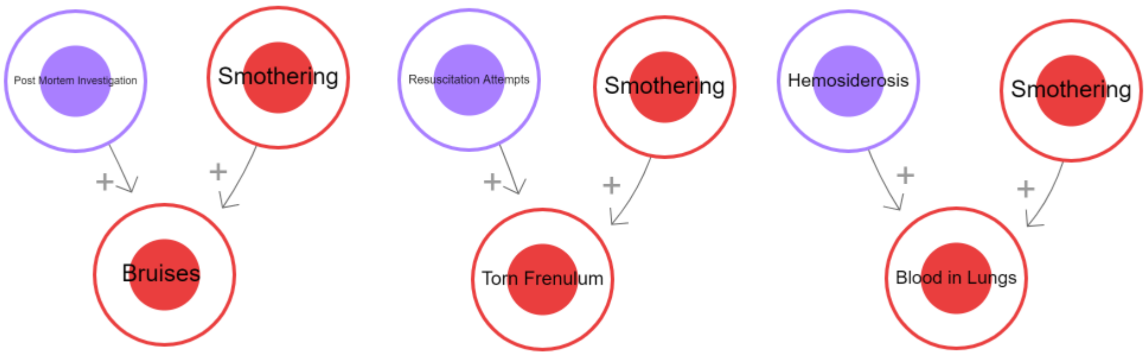


Figure 6. *Study 1: Participant drawn causal models of defence and prosecution accounts, representing these in separate models for each item of evidence.*

2.2.4. Causal structure and explanation choice

Finally, using Fisher's Exact Test, we investigated whether there is an association between causal structure drawn and explanation chosen. Our results showed there was no significant association between these two factors, $p = 0.08$. As can be seen in Table 5, the vast majority of participants drew the explanations either in separate models or each account or a unified model - and in both of these groups, the majority of participants preferred the simple explanation over the complex one. The extremely low number of participants who drew the explanations as three separate models, however, did not allow us to establish the significance of this group preferring the 'complex' explanation more than participants who represented the explanations in other structures.

Table 5

Study 1: Number of participants who chose each explanation and drew a certain causal model structure.

Causal Model structure	Explanation Preference: Complex (defence)	Explanation Preference: Simple (prosecution)	Total
n/a	1	2	3
Separate for each account	10	32	42
Separate for each evidence	2	0	2

Unified	7	14	21
Other	1	0	1
Total	21	48	69

3. Study 2

Here, we build on the findings of Study 1 by exploring whether another factor, namely the *order* that information about the competing accounts is 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* (see section 3.1.2. for more detail). This contrasts with how information was learned by participants in Study 1, in which the complete competing explanations were presented *for all the evidence* sequentially. Presenting information on the two competing explanations simultaneously might additionally allow participants to represent the information in more varied ways, e.g., as "separate for each piece of evidence". This will allow us to ascertain whether order of presentation influences the causal structures drawn, and whether a preference for representing explanations in unified models, or as two models, holds regardless of how information is presented to participants.

3.1. Methods

3.1.1. Participants

138 participants (Mean age = 37.2, SD = 12.2; *n* females = 99) completed Study 4 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). No participant dropped out of the study.

3.1.2. Materials and Procedure

A between-subjects design was employed. 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, hereafter referred to as: 'describe' (*n*=70) and 'draw' (*n* = 68). In both conditions, participants reasoned with the same criminal case, evidence, and explanations thereof, used in Study 1 (see Section 2.1.2).

Participants in the 'draw' condition received the same training on causal models and Loopy as that detailed in Section 2.1.2 of Study 1. Subsequently, they read the case briefing and report of the evidence found through the medical examination (bruises on the arms and legs, fresh blood in the lungs and a torn frenulum). Rather than presenting the two competing explanations of all of the evidence sequentially as was done in Study 1, participants 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 using Loopy. Subsequently, participants 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 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. Participants were told at each new information round that they could represent the information as one causal model, two models or in any way that seemed intuitive to them.

Participants in the 'control' condition received information in the same manner as that of participants in the 'draw' condition, however they were not required to draw models representing the information given to them at any stage. Rather, participants were required to describe/summarise the information after it was presented. As such, after having seen the two competing explanations (defence and prosecution) for the first piece of evidence participants in the 'describe' condition were asked to: "*Please write, in your own words, a description or summary of the explanations of the evidence viewed so far*" in a text box. This was asked again after participants viewed the two competing explanations for the second piece of evidence, as well as the third.

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 the evidence. They were additionally required to provide a think-aloud response justifying their choice for us to obtain an insight into the explanatory features that were valued. For a graphical representation of the procedure participants in each condition followed see Figure 7.

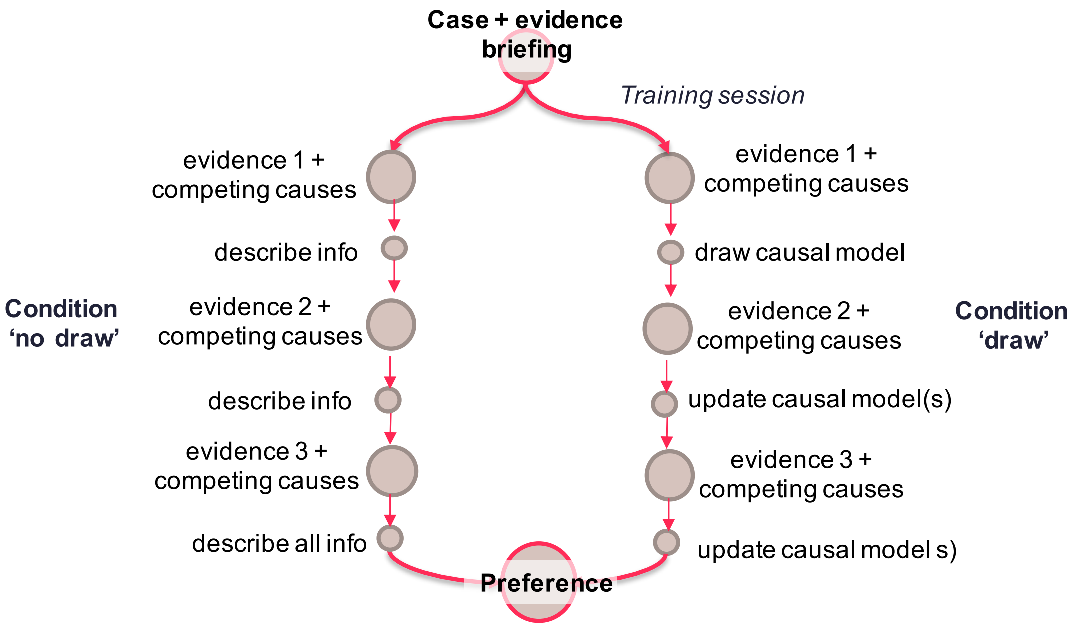


Figure 7. *Study 2: Graphical depiction of experimental procedure in 'draw' and 'no draw' conditions.*

3.2.Results

3.2.1. Explanation Choice

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 6.

Table 6

Study 2: Number of participants who chose each explanation in each condition.

Condition	Choice, *N* (%) Complex (defence)	Choice, *N* (%) Simple (prosecution)	Total
Describe	34 (48.6%)	36 (51.4%)	70
Draw	20 (29.4%)	48 (70.6%)	68

A Chi Square test of independence revealed a significant difference in the distribution of participants' choices between the two conditions, χ^2 (1) = 4.5, p = 0.03, φ = 0.19. As can be seen from Table 6, most participants in the 'draw' condition chose the prosecution's explanation as the best explanation for the evidence[8]. Conversely, participants in the 'describe' condition did display a clear preference for the simple explanation put forth by the prosecution or the complex explanation put forth by the defence[9]. These findings replicate those of Study 1 regarding a simplicity preference when drawing causal models.

3.2.2. Reasoning (think-aloud responses)

The reasoning codes drawn from our participant sample, with their frequency in each condition, can be seen in Table 7. For a description of the codes see Table 2 (in Study 1).

Table 7

Study 2: Percentage of reasoning codes found in each condition.

Reasoning Code	Draw Condition (%)	Describe Condition (%)
Complexity/ Specificity	7.4	8.6
Mechanism	16.2	27.1
No Motive/Intent	2.9	12.9
Probability	10.3	4.3
Simplicity / Probability	45.6	15.7
Fits with evidence	7.4	15.7
Other	10.3	15.7

A Chi-Square test of independence illustrated a significant difference in the distribution of reasoning codes between conditions, χ^2 (6) = 20.9, p = 0.002, V = 0.39. Bonferroni corrected post-hoc comparisons revealed the only significant difference to be that participants in the 'draw' condition employed reasoning that fell under the 'simplicity/probability' code significantly more than participants in the 'describe' condition, p = 0.001. This was in line with the findings of Liefgreen and Lagnado (2021) and once again suggests that it the act of drawing the causal models leads people to increasingly consider

[8] A binomial test within the 'draw' condition revealed a significant preference for the prosecution's explanation, p < 0.001.
[9] A binomial test within the 'describe' condition did not reveal a significant preference for the prosecution's explanation, or the defence's explanation, p = 0.09.

the probabilistic connotations of the causal structures underlying the two competing explanations. For the percentage of codes underlying each explanation choice within each condition see Table 8.

Table 8.

Study 2: Percentage of participants within each reasoning category who chose each explanation, per condition.

	Draw Condition		Describe Condition	
Reasoning Code	Complex (defence) explanation (%)	Simple (prosecution) explanation (%)	Complex (defence) explanation (%)	Simple (prosecution) explanation (%)
Participant Choice	***30.4***	***69.6***	***48.6***	***51.4***
Complexity/Specificity	25	0	11.8	0
Mechanism	25	12.5	23.5	22.2
No Motive/Intent	10	0	26.5	0
Probability	10	10.4	5.9	2.8
Simplicity/Probability	5	62.5	0	38.9
Fits with evidence	15	4.2	5.9	23
Other	10	10.4	26.5	11.1

3.2.3. Causal models (in 'draw' condition)

Next, we take a closer look at the causal models that participants produced in the 'draw' condition, in order to evaluate their variability and ascertain whether structural differences in the causal models drawn influenced participants' explanatory preferences. As mentioned in Section 3.1.2., in the drawing condition participants viewed the two competing explanations for each piece of evidence sequentially. As such, they drew *three* models. One including the two competing explanations for the *bruises*, one updated model including the competing explanations for the *bruises* and *the torn frenulum* and the final model with the competing explanations for the *bruises, the torn frenulum* and *the blood in the lungs.* Given that once again our question of interest pertained to whether representing competing explanations in disjunctive or unified models influences how these explanations are evaluated, we only present the analyses carried out on participants' final models below. All final models were coded according to the features listed in Table 4 (Section 2.2.3). Our main feature of interest was what *structure* participants utilised to embody the information found in the two competing explanations. In terms of structure, models were again coded according to whether the competing accounts (defence vs. prosecution) were represented as: (i) two separate models, (ii) one unified model, (iii) three separate models, one for each evidence piece, or (iv) 'other' - a structure that didn't fit in any of these above categories. For the number of participants who drew each type of structure see Figure 8.

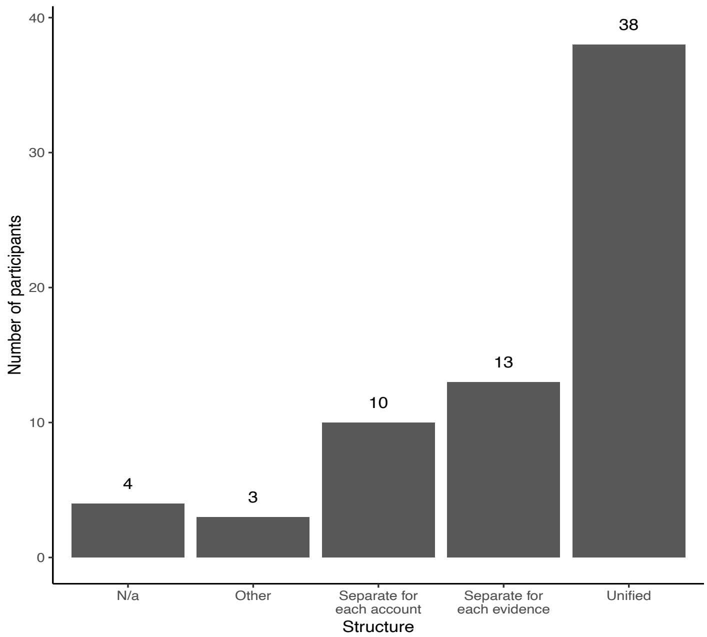


Figure 8. *Study 2: Frequency of model structures in 'draw' condition.*

We found that when the explanations were presented simultaneously for each piece of evidence, the majority of participants (55.8 %) represented the explanations in a unified model. In addition, 14.7 % represented them in two separate models – one for each account - and19.1 % as three separate models (one for each piece of evidence). 4.4 % of models were classified as 'other' given they did not fall into any of the above categories. A Chi-Square goodness of fit test showed a significant difference in the χ^2 (4) = 59.79, p < 0.001 in the distribution of structures used by participants.

3.2.4. Causal structure and explanation choice

Finally, using Fisher's Exact Test, we investigated whether there is an association between causal structure drawn and chosen explanation. Our results showed there was no significant association between these two factors, p = 0.09. As can be seen in Table 9, most participants drew the explanations either in separate models or each account or a unified model - and in both groups, the majority of participants preferred the simple explanation over the complex one. Once again, likely due to the lower number of participants who drew three separate models, we were unable to determine the significance of the finding that the "separate for each evidence" subgroup prefers the complex explanation over the simple one.

Table 9.

Study 2: Number of participants who chose each explanation and drew a certain causal model structure.

	Explanation Preference		Total
	Complex (defence)	Simple (prosecution)	
n/a	2	2	4
Separate for each account	6	10	16
Separate for each evidence	5	8	13
Unified	5	27	32
Other	2	1	3
Total	20	48	68

3.2.5. Model structure in Study 1 vs Study 2

In order to ascertain whether the order in which information relevant to the two competing explanations was presented to participants affected how they graphically represented the information we used Fisher's Exact Test, comparing the structure of final models in Study 1 (n = 69) and Study 2 (n = 68) - see Table 10.

Table 10

Percentage of causal model structures drawn in Study 1 and Study 2.

Causal model structure	Study 1 (%)	Study 2 (%)
Unified	30.4	55.9
Separate for each account	60.9	14.7
Separate for each evidence	2.9	19.1
Other	1.4	4.4

n/a	4.3	10.3
Total N	**69**	**68**

Our analysis yielded a significant difference in the frequency with which each model structure was adopted by participants between the two studies, < 0.001. Bonferroni corrected post-hoc pairwise comparisons showed that there was a difference in the proportion of 'unified' 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 evidence' category, p = 0.004. As can be seen from Table 10, 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 (19 %) compared to Study 1 (2.9%).

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

4. Study 3

In the studies we presented so far, the simple explanation always corresponded to the prosecution, and the complex explanation to the defence. Although an analysis of participants' think-a-loud responses did not suggest that this confounded our results, we were nonetheless not able to generalize our results in their current form, given that the explanations did not differ only in terms of complexity, but also burden of proof and asymmetric cost of falsely accepting one explanation over the other[10]. For this reason, we ran a final study, in which we counterbalanced whether the simple explanation is advanced by the defence or the prosecution. To anticipate our findings, even in this case we replicated the main

[10] We thank one of our reviewers for pointing this out and encouraging us to include an additional experiment to address this issue.

fining that drawing causal models promoted a preference for a simpler model (whether for prosecution or defence).

4.1. Methods

This study replicates Study 1 – in which explanations are presented sequentially – but counterbalances whether the simple explanation was advanced by the defence or the prosecution. We did not include a condition in which explanations were presented simultaneously, as our previous studies did not find a main effect of this manipulation. Rather, we created a new legal scenario and focused on replicating our Drawing effect, crossing simple and complex explanations with adversarial side (prosecution or defence), to increase the generalisability of our previous findings.

4.1.2. Participants

202 participants (Mean age = 37.5, SD = 13.7; *n* females = 153) completed Study 5 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). No participant dropped out of the study.

4.1.3. Materials and Procedure

A between-subjects design was employed. As in previous studies, 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, hereafter referred to as: ‘describe’ (*n*=101) and ‘draw’ (*n* = 101). The procedure of Study 3 was identical to that of Study 1 (see Section 2.1.2), bar for the fact that half of the participants within each condition saw a scenario in which the simple explanation was put forth by the prosecution, and the other half saw a version of the scenario in which the simple explanation was put forth by the defence[11].

Given the new counterbalancing conditions, we designed a new legal scenario for this study (this also enabled us to generalise our findings beyond a single legal scenario). This scenario was

[11] This meant that 50 participants were in the ‘Describe’ condition and saw a version of the scenario in which the complex explanation was the prosecution’s, 51 participants were in the ‘Describe’ condition, and saw a version in which the complex explanation was the defence’s. Similarly, 50 participants in the ‘Draw’ condition saw a version in which the complex explanation was the prosecution’s, and the remaining 51 saw a version in which the complex explanation was the defence’s.

inspired by the real case of Michael Peterson (NC v Peterson, 2003; depicted in the Netflix series 'The Staircase'), in which a man was accused of the murder of his wife who was found lifeless at the bottom of their home staircase. Participants in any condition were given the initial briefing of:

"Eric and Laura have been married for more than a decade and live alone in their family home. On the night of Saturday 5th July, after spending the evening together, Eric found Laura unconscious at the bottom of the staircase in their home. He called the ambulance which arrived on the scene shortly after, but they were unable to save Laura. Eric is now the prime suspect in Laura's death."

For a graphical representation of the two versions of the scenarios we developed (changing whether the simple or complex explanation was the prosecution's or the defence's), see Figure 9. As can be seen from Figure 13, the evidence was kept the same in each version of the scenario, as all participants were told that the following distinct findings were reported by professionals at the scene of the crime:

"(i) damage to the banister of the stairs in one location

(ii) Lacerations on Laura's head

(iii) Bruises on Laura's arm"

We only varied whether the simple explanation was put forth by the prosecution (She was pushed) or the defence (she fell).

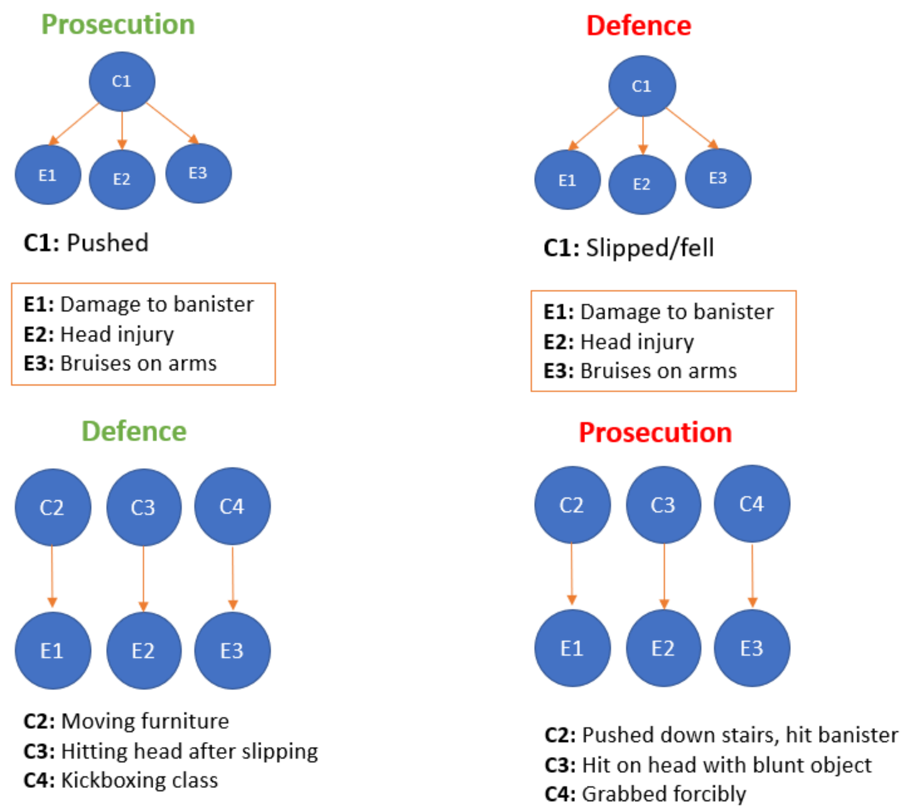


Figure 9. *Study 3: Graphical representation of the evidence and causes presented to participants, varying which adversarial side put forth the simple vs. complex explanation.*

Once again, in both the describe and the draw condition, 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.

4.2. Results

4.2.1. Explanation Choice

A Chi Square test of independence revealed a significant difference in the distribution of participants' choices between the two conditions, χ^2 (1) = 30.48, $p < 0.001$, $\varphi = 0.39$. As can be seen from Table 11, the majority of participants in the 'draw' condition chose the prosecution's explanation as the best explanation for the evidence[12]. Conversely, participants in the 'describe' condition displayed a preference for the complex explanation[13].

Table 11.

Study 3: Explanatory preference within each condition (Draw vs Describe) and broken down by which adversarial presented the complex explanation.

	Explanation Preference					
Method Condition	Complex		Complex Preference Total N (%)	Simple		Simple Preference Total
	Defence Complex	Prosecution Complex		Defence Complex	Prosecution Complex	
Describe	28	37	65 (64.3%)	22	14	36 (35.7%)
Draw	11	14	25 (24.7%)	40	36	76 (75.3%)

These findings replicate those of our previous studies, regarding a simplicity preference when drawing causal models. A Chi Square test of independence revealed no difference in the distribution of participant's explanatory choices when the simple explanation was put forth by the prosecution vs. the defence, χ^2 (1) = 2.42, p = 0.12, φ = 0.12. This crucially indicates that participants' explanatory preferences were not influenced by whether a specific adversarial side put forth the simple vs. complex explanation, allowing us to consolidate our findings regarding the influence of drawing causal models on people's explanatory preferences (and replicating these in a new legal scenario).

4.2.2. Reasoning (think-a-loud)

The reasoning codes drawn from our participant sample, with their frequency in each condition, can be seen in Table 12. For a description of the codes see Table 2 (in Study 1).

[12] A binomial test within the 'draw' condition revealed a significant preference for the simple explanation, $p < 0.001$.

[13] A binomial test within the 'describe' condition did revealed a significant preference for the complex explanation, $p = 0.005$.

Table 12.

Study 3: Percentage of reasoning codes in each method condition.

Reasoning Code	Method Condition		Total N
	Describe (%)	Draw (%)	
Complexity/Specificity	16.8	4.9	22
Fits with Evidence	16.8	17.8	35
No Intent / Motive	12.9	6.9	20
Other	8.9	10.9	20
Probability	18.8	18.8	38
Questioning Mechanism	15	3.9	19
Simplicity/Probability	14.9	35.6	45
n/a	1.9	1	3
Total N	101	101	202

A Chi-Square test of independence illustrated a significant difference in the distribution of reasoning codes between conditions, $p < 0.001$. Bonferroni corrected post-hoc comparisons revealed the only significant difference to be that participants in the 'draw' condition employed reasoning that fell under the 'simplicity/probability' code significantly more than participants in the 'describe' condition, $p < 0.001$. This was in line with the findings of our previous studies and once again suggests that it is the act of drawing the causal models, rather than thinking more about the explanations or engaging more with the task, that leads people to increasingly consider the probabilistic connotations of the causal structures underlying the two competing explanations.

Through Fisher's Exact Test, we did not find a significant difference in the distribution of reasoning codes between groups who saw a scenario in which the complex explanation was presented by the prosecution vs. defence, $p = 0.42$. This once again allowed us to conclude that changes to whether the simple or complex explanation of the evidence is put forth by a specific adversarial side, does not influence people's explanatory preferences or reasoning underlying these preferences, which instead are influenced by drawing causal models, and reasoning increasingly in term of ‘simplicity’ and affiliated probabilistic connotations.

4.2.3. Causal models (in `draw' condition)

As per our previous studies, all final models drawn by participants in the ‘draw’ condition after having viewed both explanations for the evidence were coded according to the features listed in Table 4 (section 2.2.3). Our main feature of interest was once again what *structure* participants utilised to embody the information found in the two competing explanations. For the number of participants who drew each type of structure see Figure 10.

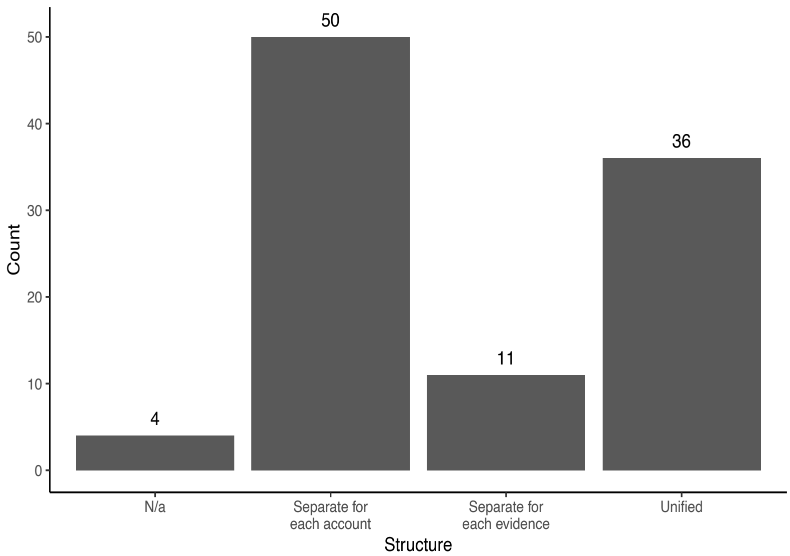


Figure 10. *Frequency of model structures in ‘draw’ condition Study 3.*

Replicating our previous findings involving the sequential presentation of the explanations, we found that most participants (49.5%) represented the explanations as two separate models. In addition, 35.6 % represented them in as a unified model, and 10.9 % as three separate models (one for each piece of evidence). 3.9 % of models were classified as ‘other’ given they did not fall into any of the above categories. A Chi-Square test of Independence showed a significant difference in the χ^2 (4) = 59.79, p < 0.001 in the distribution of structures used by participants.

4.2.4. Causal structure and explanation choice

Finally, using a Chi-Square test of Independence, we investigated whether there is an association between the causal structure drawn and chosen explanation. Our results showed there was no significant association between these two factors, χ^2 (3) = 0.66, p = 0.08, V = 0.08. As can be seen in Table 13, the vast majority of participants drew the explanations either in separate models for each account or a unified model - and in both of these groups, the majority of participants preferred the simple explanation over the complex one.

Table 13

Study 3: Model structure codes by explanation preference.

Model Structure	Explanation Preference		**Total**
	Complex	Simple	
n/a	1	3	4
Separate per evidence	2	9	11
Separate per explanation	14	36	50
Unified	8	28	36
Total	25	76	101

We additionally did not find a difference in the distribution of model structure codes depending on whether the simple explanation shown to participants was put forth by the prosecution or the defence, χ^2 (3) = 6.65, p = 0.08, V = 0.25.

5. General Discussion

In three studies, we investigated: i) how people represent competing explanations of the same legal evidence, 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 simply reading the explanations and describing them.

In terms of explanatory preferences, in all of our studies we observed a preference for the 'simple' explanation when participants were asked to draw causal models of the explanations. This preference was less clear in participants required to 'describe' the explanations. The large proportion of participants who preferred the complex explanation in the 'describe' conditions of our three studies (and in the control condition of the studies presented in Liefgreen and Lagnado (2021)) suggests that more parsimonious explanations may not be favoured over complex ones in certain domains (i.e., 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. In the present studies, 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 et al. (2019), positing that people use features of complexity in an explanation as a cue for goodness-of-fit and Bayesian likelihood.

In addition to 'simplicity' and 'complexity/specificity' a large cluster of participants across all of studies, cited mechanism-related factors as reasons for their explanatory choices. Our findings reinforce the notion that mechanism is a factor that people consider important when evaluating explanations (Zemla et al., 2020), by showing that when evaluating explanations that contain no details of mechanisms, people spontaneously deliberated and questioned the possible mechanisms involved in bringing about the given effects (injuries in our scenario). As such, when no details of these mechanisms are provided, people seemingly use their own prior knowledge and intuitions about the cause-effect relations to evaluate the explanations. This fits with the idea that people's beliefs about causal relations naturally includes beliefs in causal processes that take place between the cause and effects as well as beliefs about the nature of the mechanisms underlying these relations (Ahn & Kalish, 2000). Across our studies, when the simple explanation was not favoured this was partly due to the fact that some of the subjectively inferred mechanisms underlying the cause-effect relations in this explanation were being questioned, such as how *smothering* could lead to a *torn frenulum* (materials used in Studies 1 and 2). Similarly, a significant number of people stated that they did not favour the complex explanation because they did not believe that *bruises* could occur *post-mortem*, as the defence's explanation implied. Given that details of mechanisms in explanations have been found to increase one's subjective sense of understanding (Varsilyeva & Lombrozo, 2015; Zemla et al., 2017), it is not surprising that questioning the plausibility of conjectured mechanisms underlying cause-effect relations in one explanation leads to a preference for the alternative explanation. Relevant to this finding is also the notion that if something cannot be accommodated within one's causal model, due to its conflict with prior knowledge or because it lacks a suitable explanation relative to the current causes in the model, it is largely ignored or underweighted (Krynski & Tenenbaum, 2007). People's engagement with questioning the mechanisms underlying the causal explanations decreased when participants were drawing or describing the explanations – at which point, especially when drawing the explanations, they prioritised evaluating explanations based on other features such as simplicity/probability. This could be because drawing links between causes and effects facilitates the acceptance that a link between these exists, regardless of the mechanism, making questioning underlying mechanisms less appropriate[14]. To rule out the possibility that drawing naturally favours structural simplicity vs. representing complexity within the modelling framework, future studies should manipulate the presence/absence of mechanism in the explanations as well as people's possibility of representing mechanisms within their causal representations. This would also help us ascertain whether drawing links between causes and effects facilitates the acceptance that a link between these exists, regardless of the mechanism. Overall, however, our findings suggest that details of mechanism are an important component to consider when providing people with explanations, especially in high-stake (e.g., legal) domains. This would help to maximise people's sense of understanding of the arguments under consideration and avoid diverting their attention away from probabilistic inference when reasoning under uncertainty.

[14] We thank an anonymous reviewer for this suggestion.

Across all studies, compared to participants in the ‘describe’ conditions, we found an increase in the frequency of reasoning relating to the simplicity and probability of the two explanations in the ‘draw’ condition. As such, drawing causal models of the explanations led to a shift in explanatory preference – in favour of the simple explanation – and in (probabilistic) reasoning. This was true regardless of the order in which evidence was presented. 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., Jefferys & Berger, 1992; Lagnado, 1994; 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. Future work should directly test this hypothesis by providing participants with the necessary probabilistic information to enable a comparison of their reasoning against a normative (Bayesian) benchmark. It is possible that, having a probable defence narrative is less critical than having a probable prosecution story because the burden of proof is on the prosecution. As such, drawing causal models facilitated participants’ engagement with the probabilistic connotations of the causal structures representing the explanations – which led them to favour the simpler (and thus likely more probable) explanation.

Although our control condition arguably satisfyingly matched – though not perfectly so – the experimental condition in terms of deliberation time or engagement, to rule out the possibility that shift in explanatory preference and reasoning was due to increased deliberation time or engagement with the task in the draw condition compared to the control condition, running further studies with other types of comparative conditions which vary in terms of engagement and deliberation time, and time spent drawing causal models is explicitly measured. Importantly, we replicated the effect of drawing causal models on people’s explanatory preferences in Study 3, in which we a) utilised a different legal case and b) counterbalanced which adversarial side (prosecution or defence) put forth the simple vs. complex explanation. This allowed us to generalise our findings beyond a single scenario and ascertain that the effect observed in Study 1 and Study 2 was not driven by factors associated with the fact that the simple explanation was always put forth by the prosecution rather than the defence.

As previously mentioned, in all our studies, drawing causal models facilitated participants' engagement with the probabilistic connotations of the causal structures representing the explanations – which led them to favour the simpler (and thus likely more probable) explanation. Although this could be for several reasons, including allowing participants to process the information on a deeper level and perhaps easing up the demands on participants' working memory by being able to compare the two explanations using the resulting diagram, we favour the hypothesis that this was due to drawing causal models allowing participants to scaffold probabilistic computations over the diagram. As such, we propose that graphically representing information using nodes and directed links boosts one's understanding of the relation between the items of information represented by the nodes (e.g., independence of causes in the disjunctive complex explanation) and the probabilistic and statistical connotations implied by these relations (e.g., the disjunctive structure most probably implies the explanation to be less probable than the explanation with an underlying common-cause causal structure). In the case of our studies, drawing causal graphs seemed to boost participants' understanding that an explanation invoking only one root cause is likely to be more probable than an explanation that invokes three independent root causes. This was corroborated by an increase in reasoning relating to the probability of the explanations in the 'draw' conditions. This hypothesis would be in line with studies showing that learning of causal relations improves performance in probabilistic reasoning tasks (Krynski & Tenenbaum, 2003). 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 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. 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 to other types of causal structures and more broadly. In addition, our findings do not at present allow us to establish whether diagram structure affects explanatory preferences, or vice-versa, or whether there is a common factor that influences both. To begin to address this query, future studies could elicit measures of people's explanatory preferences before reasoning within a scenario such as that presented in this study – and verify whether these preferences predict the structures that are drawn and mediate the influence of drawing on a simplicity preference. Moreover, to determine whether it is *drawing* the causal models that leads to a shift in people's explanatory preferences, a 'drawing' condition should be compared to a condition in which participants are shown the causal structures (perhaps the three variations that they primarily drew in our experiments) and are not required to build these themselves. We predict that in this instance participants' preferences would be comparable to those of participants in the 'draw' conditions of our studies. This, however, would likely hold only if participants were asked to represent only the information provided to them in the task – not including their own beliefs/existing knowledge or additional variables. A recent study has shown that simply presenting information to participants as a causal model might not boost their decision-making, if the information/structure presented disagrees with their existing knowledge and beliefs (Zheng et al., 2020).

Graphical models have featured in accounts relating to human categorization (Rehder, 2001) and causal structure learning (Lagnado & Sloman, 2004, 2006; Sloman & Lagnado, 2015; Tenenbaum & Griffiths, 2001) but have not yet been largely explored in the study of reasoning under real-world uncertainty (though see Zheng et al., 2020). In addition, to our knowledge, few studies have allowed people to actually *draw* causal models of the information they were required to reason with. Morais et al. (2011) investigated whether the structure of people's knowledge of causal relations between the features of categories predicts how they search for information in a categorization task. Participants were asked to draw a causal model that described how the symptoms of depression are causally related to one another, and to estimate the strengths of those relationships. Additionally, they were asked to categorize a series of patients as suffering from depression or not, after searching their symptoms. The results showed that the structurally more important a symptom was in a causal model, the more frequently and the earlier in search it was inspected, ultimately concluding that causal model structure predicted information search behaviour. Relatedly, Cruz et al. (2020) conducted a large-scale laboratory experiment illustrating that a Bayesian network modelling tool adapted to provide basic training and guidance on the modelling process helped lay people reach normative Bayesian solutions to complex reasoning problems. Given the tool had numerous features this effect cannot be specifically attributed to the graphical drawing component, though it undoubtedly played a role given that an incorrectly drawn causal structure would have prohibited participants from reaching optimal Bayesian solutions. Our results notably contribute to this growing pool of studies demonstrating that graphical causal models are helpful tools when reasoning under uncertainty, even without the underlying computational components. Future work should replicate our findings including more complex explanations comprising larger amounts of evidence to increase the ecological validity of the materials – perhaps adopting Pennington and Hastie's (1988) approach of manipulating order of evidence presentation according to whether it was temporal or random. Broadly, however, our findings can hold initial implications for the legal domain, and related domains in which one is required to generate and test explanations for items of evidence/information under uncertainty. As such, they emphasize the importance of considering the complexity of a proposed explanation and ensuring that an explanation fits the facts of the case. Especially in cases when the proposed explanation is complex, our findings also suggest that accompanying the explanation with details of the mechanisms underlying cause-effect relations may help people evaluate the explanation and its probability/plausibility.

In terms of the causal structures drawn to represent the explanations, our findings indicate that individuals do not uniformly represent these using a unified framework. As such, although Bayesian models of legal explanations have so far mostly adopted a 'unified' approach, representing in a single model all of the arguments under consideration (Aitken & Taroni, 2004; Fenton et al. 2013; Taroni et al. 2014; but see Neil et al., 2019), we have shown that people represent competing explanations in a variety of ways when asked to draw their own causal models. Participants primarily represent these competing explanations as one unified model, or as two separate models (one for each explanation), with a minority group representing them as three separate models (one for each piece of evidence and its competing causes). We should note that in our instructions (see Procedure sections of our studies), participants in the 'Draw' condition were told they could represent the information in Loopy as one model, two models, or in any way that seemed intuitive to them. This was to allow participants creative freedom when representing the information, rather than guiding them to draw a specific type of structure. Nonetheless, we should note the instructions might have influenced participants who predominantly drew unified models or two separate models. Similarly, the instructions participants received in Study 2 might have confounded the influence of the order of presentation on the causal structures drawn by participants. Nonetheless, our findings might 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 engage in different inferential and evaluative processes. Finally, we showed that the way 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, posing obvious implications for the delivery of adversarial arguments in court trials. Overall, we advocate that future work modelling legal explanations should elicit the models of the reasoners involved in order to optimise the development of normative solutions to the problem at hand, and 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.

The findings presented in this paper, especially if replicated utilising different scenarios, could prove insightful to other disciplines in which people are required to reason with, and evaluate competing explanations of, large amounts of interrelated information under uncertainty – such as Judicial decision-making, medical diagnosis and intelligence analysis. As such, our research contributes to ongoing work illustrating how causal explanatory theories and Bayesian probability theory can complement each other in supporting decision-making practices of Judges (Mackor et al., 2022). Finding that people can easily represent competing legal narratives as causal models (receiving minimal training), is an encouraging first step towards developing intuitive methods that can help judges and professionals to structure their legal arguments and inferences, and – once transforming these graphical models into Bayesian Networks, guard against probabilistic fallacies.

6. References

Ahn, W. K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. *Explanation and cognition*, 199-225.

Aitken, C., & Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*. John Wiley & Sons.

Bex, F., Van den Braak, S., Van Oostendorp, H., Prakken, H., Verheij, B., & Vreeswijk, G. (2007). Sense-making software for crime investigation: how to combine stories and arguments? *Law, Probability & Risk*, *6*(1-4), 145-168.

Blanchard, T., Lombrozo, T., & Nichols, S. (2018). Bayesian Occam's Razor Is a Razor of the People. *Cognitive Science*, *42*(4), 1345–1359.

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*(4), 1156–1164.

Boneh, T., Weymouth, G. T., Newham, P., Potts, R., Bally, J., Nicholson, A. E., & Korb, K. B. (2015). Fog forecasting for Melbourne Airport using a Bayesian decision network. *Weather and Forecasting*, *30*(5), 1218-1233.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

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.

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.

Dijk, T. A. V., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic.

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.

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.

Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, *69*(S3), S342–S353.

Hayes, B. K., Ngo, J., Hawkins, G. E., & Newell, B. R. (2018). Causal explanation improves judgment under uncertainty, but rarely in a Bayesian way. *Memory & cognition*, *46*(1), 112-131.

Hung, W., & Jonassen, D. H. (2006). Conceptual Understanding of Causal Reasoning in Physics. *International Journal of Science Education*, *28*(13), 1601–1621.

Jefferys, W. H., & Berger, J. O. (1992). Ockham's razor and Bayesian analysis. *American Scientist*, *80*(1), 64-72.

Jensen, F. V. (1996). *An introduction to Bayesian networks* (Vol. 210, pp. 1-178). London: UCL press.

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

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press.

Krynski, T. R., & Tenenbaum, J. B. (2003). The role of causal models in statistical reasoning. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society,* 693-698.

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

Lagnado, D. (1994). *The psychology of explanation: A Bayesian approach* (Master's Thesis). Schools of Psychology and Computer Science, University of Birmingham.

Lagnado, D. (2021). *Explaining the evidence: how the mind investigates the world*. Cambridge University Press.

Liefgreen,A., and Lagnado,D. (2021). The role of causal models in evaluating simple and complex legal explanations. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 43(43).

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

Lipton, P. (2004). *Inference to the best explanation*. Taylor & Francis.

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.

Mackor, R.A.,Dahlman, C. & Lagnado, D. (2022). *The Evaluation of Evidence in Criminal Law* [Workshop Presentation]. Bochum Bayesian Learning & Reasoning Workshop 2022, Online. https://www.youtube.com/watch?v=wT1gjqQw1cs.

Morais, A. S., Olsson, H., & Schooler, L. J. (2011). Does the structure of causal models predict information search? *European perspectives on cognitive science*, 1-6.

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.

Niiniluoto, I. 1999. Critical Scientific Realism. New York: Oxford University Press.

Nissan, E. (2008). Select topics in legal evidence and assistance by artificial intelligence techniques. *Cybernetics and Systems: An International Journal*, *39*(4), 333-394.

Nobles, R., & Schiff, D. (2005). Misleading statistics within criminal trials: The Sally Clark case. *Significance*, *2*(1), 17-19.

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. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, *51*(2), 242–258.

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), 521–533.

Pennington, N., & Hastie, R. (1991). A Cognitive Theory of Juror Decision Making: The Story Model Decision and Interference Litigation. *Cardozo Law Review*, *13*(Issues 2-3), 519–558.

Prakken, H. (2004). Analysing reasoning about evidence with formal models of argumentation. *Law, Probability and Risk*, *3*(1), 33-50.

Clark, R v (2003) *England and Wales Court of Appeal Crim 1020*. http://www.bailii.org/ew/cases/EWCA/Crim/2003/1020.html

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.

Rehder, B., & Hastie, R. (2001). Causal knowledge and categories: The effects of causal beliefs on categorization, induction, and similarity. *Journal of Experimental Psychology: General*, *130*(3), 323.

Saini, A. (2009). The mathematical mistakes that could be undermining justice. *New Scientist*, *204*(2731), 42-45.

Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual review of psychology*, *66*, 223-247.

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.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., & Aitken, C. (2014). *Bayesian networks for probabilistic inference and decision analysis in forensic science*. John Wiley & Sons.

Thagard, P. (1989). Explanatory coherence. *Behavioral and brain sciences*, *12*(3), 435-502.

Thagard, P. (2000). *How Scientists Explain Disease*. Princeton University Press.

Vasilyeva, N., & Lombrozo, T. (2015). Explanations and Causal Judgments are Differentially Sensitive to Covariation and Mechanism Information. In *CogSci*.

Vorms, M., & Lagnado, D. (2019). Coherence and Credibility in the Story-Model of Jurors' Decision-Making: Does Mental Simulation Really Drive the Evaluation of the Evidence? In Á. Nepomuceno-Fernández, L. Magnani, F. J. Salguero-Lamillar, C. Barés-Gómez, & M. Fontaine (Eds.), *Model-Based Reasoning in Science and Technology* (pp. 103–119). Springer International Publishing.

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.

Zarri, G. P. (1998). Representation of temporal knowledge in events: The formalism, and its potential for legal narratives. *Information and Communications Technology Law*, *7*(3), 213-241.

Zhang, H., Liu, J., & Pang, A. C. (2018). A Bayesian network model for data losses and faults in medical body sensor networks. *Computer Networks*, *143*, 166-175.

Zheng, M., Marsh, J. K., Nickerson, J. V., & Kleinberg, S. (2020). How causal information affects decisions. *Cognitive Research: Principles and Implications*, *5*(1), 6.

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

Zemla, J., Sloman, S. A., & Lagnado, D. (2020). Not so simple! Mechanisms increase preference for complex explanations. Pre-print: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4130789.

Appendices

Appendix 1: Instructions given to participants on causal models during the training block

Page 1.

Forms of causal reasoning (reasoning about causes and effects) are extremely common in everyday situations. For example, you might use causal reasoning when deciding what to eat (e.g. avoiding certain **foods** you know might **cause** you **stomach pains**).

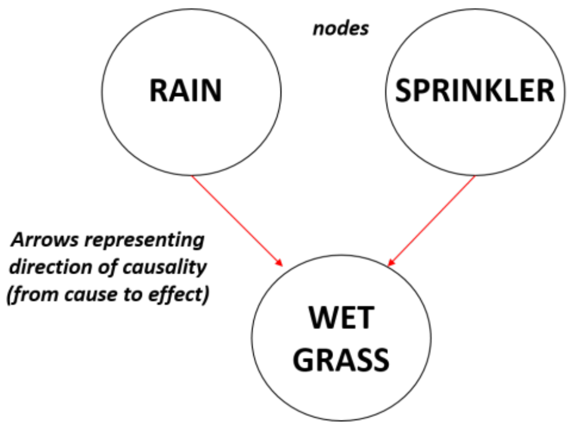
Any situation involving reasoning about causes and effects can be represented in what we call a **"causal model".** Causal models are essentially diagramming that allow you to represent objects, events, items of information etc. as" NODES**" (circles)** and draw arrows between these nodes to represent the relationship between them (e.g. cause and effect).

For example, if you wanted to represent the following information in a causal model:

"rain and a sprinkler can both make the grass wet"

you would first identify your elements of interest (nodes; RAIN, SPRINKLER, WET GRASS) and using arrows represent **RAIN and SPRINKLER** as **causes** of the **effect**: **WET GRASS**.

The causal model would therefore look like the one below:

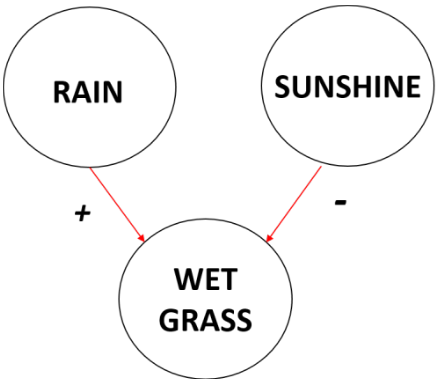


Page 2.

Sometimes, you might also want to say whether the cause makes the effect **MORE** or **LESS** likely e.g. rain makes wet grass MORE likely but if you had another node representing "sunshine" this would make the effect (wet grass) LESS likely.

To show this, you can either have a plus (+) or (-) sign next to the arrow to say if the cause makes the effect more likely (+) or less likely (-)

This can be shown in the causal model diagram below:



Page 3.

In today's task, you will be presented with a scenario and asked to draw your own causal model for it using a tool called Loopy.

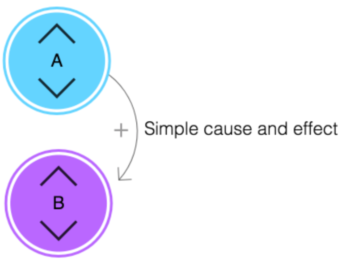
Loopy is an online tool that allows you to draw causal models by simply drawing nodes as circles and arrows to show the relation between these.

Please open Loopy now in a new tab following this link: https://ncase.me/loopy/v1.1/ so that you can switch back and forth between Loopy and this survey.

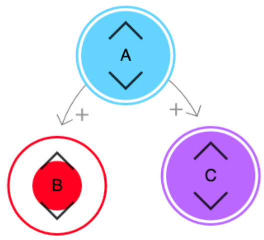
Page 4

Here are how some example structures would look like in Loopy:

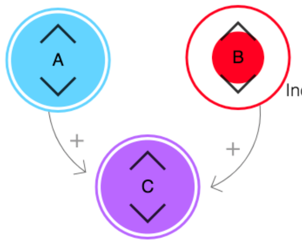
1) Simple cause and effect where the node A is the cause and B is the effect



2) Simple common cause diagram where node A is the cause of both B and C



<u>3) Simple common effect diagram where node A and Node B both cause the effect C</u>



Now open loopy the loopy tab (following this link: https://ncase.me/loopy/v1.1/) and keep it open in a separate tab for the duration of this experiment.

Delete anything on the initial screen using the ERASER tool and **replicate each of the above diagrams as best as you can on the same page**. **To draw a node simply use the PENCIL tool to draw a circle**. **You can rename your node and change its colour using the options on the right hand of the screen**.

To draw an arrow between two nodes simply use the PENCIL tool to draw a line between them. The default arrow will have a **"+"** sign. This is all you will need for the present study so there is no need to ever change it to a **"-”.**

When you are done drawing each diagram (on the same page), go to the right-hand menu on the Loopy webpage and click "save as link".

Page 5.

One last practice round before you get started!
Clear your loopy page and start a new one.

Please represent the information below in a causal diagram on a fresh loopy page.

" Tom has a cough. The doctor thinks that it could be a symptom of either asthma or the flu".

<u>Remember to rename your nodes with informative names (e.g. "cough").</u>

Once you have finished, click "save as link" on the right-hand side menu and paste the link in the text box below.

Keep the diagram as simple as you can.