**Interdependent diffusion creates polarization in network experiments**

James Houghton

**Introduction**

Simulations of social contagion *(1-7)* predict that when beliefs interact with one another as they diffuse, new sociological processes emerge which may lead to polarization in the population. These processes lead sets of beliefs to be widely shared amongst subsets of the population, suggesting the emergence of “worldviews”. At the same time, the most similar individuals become yet more similar, and the least similar individuals increase their differences. Ideological “camps” become more self-consistent, and differentiated from one another. Additionally, individuals’ beliefs can be more easily described as positions along a “left-right” axis, such that a “political spectrum” emerges.

These simulations use highly simplified models of human behavior to demonstrate the sufficiency of belief interaction for generating new outcomes. However, their representations of human decision processes may be mistaken in ways that undermine the conclusions of the simulations. An empirical test is needed to build confidence in these simulations’ predictions.

Unfortunately, the mechanisms and outcomes described in simulation are difficult to observe and isolate in the field, because they interact strongly with factors such as demographics, social network structure, and existing belief systems. This paper reports on a laboratory experiment that placed 2,400 human actors anonymously in 120 artificial social networks, each with a set of carefully constructed beliefs to share with one another. This tightly controlled context makes it possible to study the effect of belief interaction with minimal spillover from polarization in the outside world.

It is impossible to conduct a perfect experiment in which all beliefs are allowed to interact in the treatment world, and the same identical beliefs are constrained to be independent within control world. However, by manipulating the structure of the belief set, this experiment encourages interaction between a subset of core beliefs in the treatment condition by explicitly introducing additional cross-linking beliefs to connect them to each other. In the control condition, filler beliefs are substituted in place of the cross-linking beliefs to maintain independence between the core subset of beliefs. The above predictions are then tested on this core subset of beliefs to place a lower-bound on the effect of belief interaction on diffusion outcomes.

**Theory**

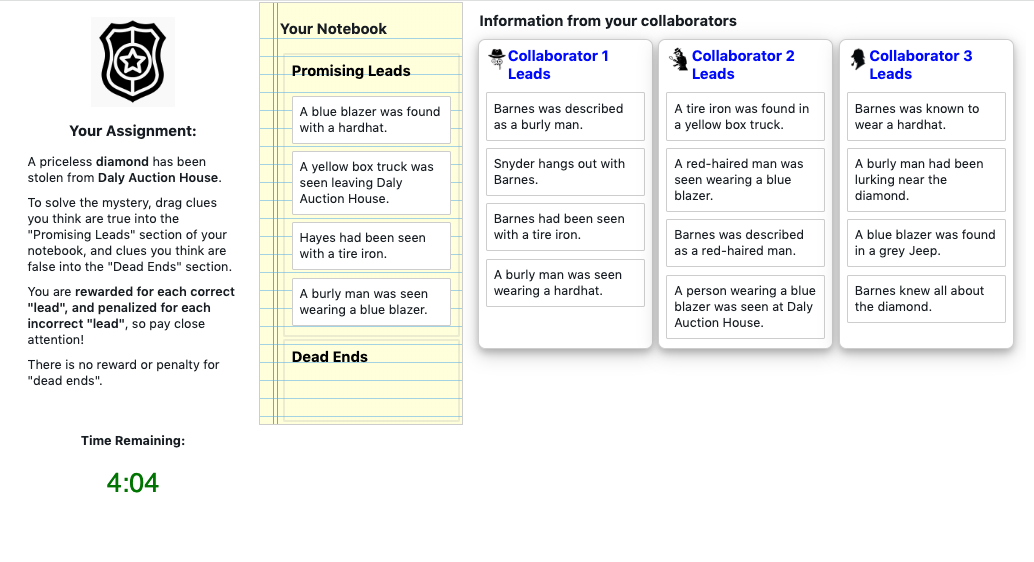
In the second paper of this dissertation *(7),* I use a simulation model to describe the mechanisms by which interdependence between beliefs leads to polarization in the population. The primary mechanism discussed is an “Agreement Cascade”, which can be described loosely as follows. When individuals adopt a belief from a contact, they become more similar to that contact than they would have otherwise been. If both parties then use their existing beliefs to decide whether to adopt a belief to which they are subsequently exposed, they will thus be more likely to respond in a similar way (to adopt or ignore the candidate belief) than they would have done in the counterfactual world.

This tendency creates a reinforcing dynamic that drives similar individuals to become yet more similar in the future. Crucially, the reinforcing dynamic occurs with no conscious desire of individuals to imitate similar neighbors, or conscious assessment of that similarity at all. The reinforcing mechanism leads some sets of beliefs to frequently cooccur, such that positions on one belief dimension become predictive of positions along other dimensions. Additionally, it leads individuals who are within the same camp to become more self-similar, and when the population begins with relatively sparse adoption of beliefs, prevents beliefs from spreading between camps to the extent they do in the counterfactual world.

**Experiment Design**

In this experiment, participants join a 20-player interactive game using their computer's browser, built on the Empirica experimental framework and shown in Fig. 1. They are tasked with finding a solution to a ‘mystery’ by identifying a burglar's name, description, clothing, burglary tool, and getaway vehicle.

Participants are given clues to the mystery, and asked to sort those clues into "Promising Leads" and "Dead Ends". They are told that their bonus depends upon the accuracy of their categorizations. When participants sort a clue into the "Promising Leads" category, the clue is immediately shared with three neighbors in the 20 person social network. The game is played in real-time over 8 minutes. Following the game, participants are asked to assess the likelihood of each suspect, vehicle, etc. being involved in the crime. Details of the participants' experience are outlined further in the `Gameplay` section of the supplement.

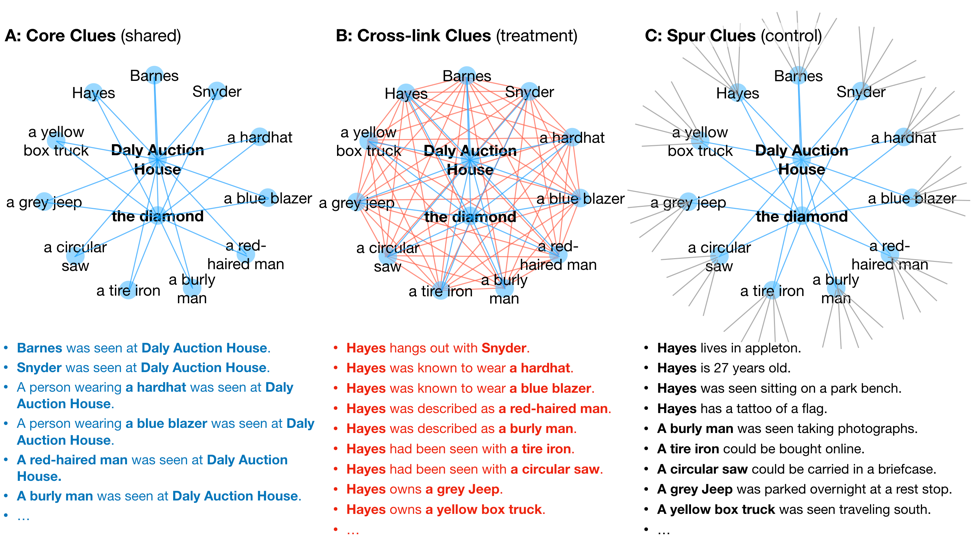


*Fig. 1. The primary user interface of the “Detective Game”. Clue cards can be dragged into categories in the player’s “Notebook”. Promising Leads are immediately shared with 3 neighbors, who can drag them into their own notebooks.*

Data are collected in three ways. First, a “behavioral” outcome is constructed from the clues that an individual has classified as Promising Leads at the end of the game, reflecting the cumulative choices that the individual has made. Secondly, a “self-report” outcome is constructed from the participant’s responses to the post-game evaluation, reflecting how the messages the individual classifies are internalized to create opinions. Lastly, a continuous-time “process” measure records the state of the player’s screen at each moment during the game, and computes the effect of factors such as the number of alters exposing the clue on an individual’s hazard of classifying it as a promising lead.

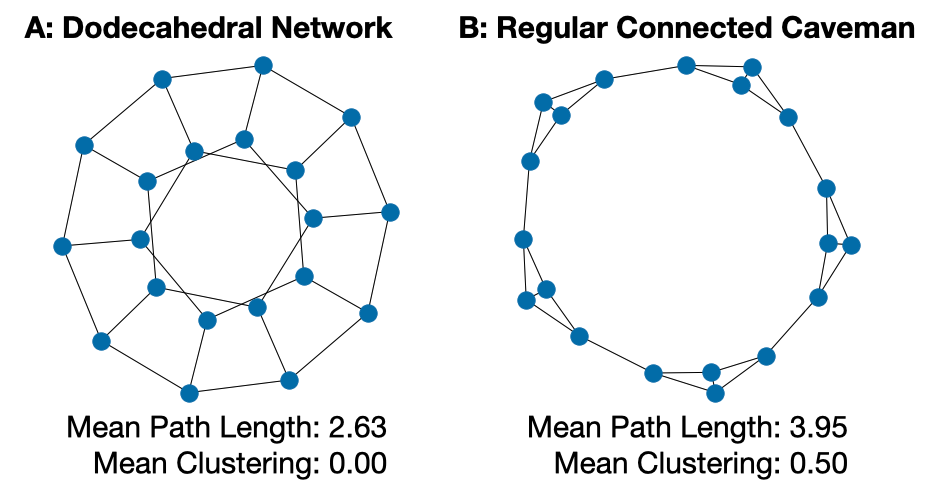
Two types of manipulation are made. First, a set of clues are constructed to interact strongly in a treatment world, and to maintain independence from one another in a control world. This is accomplished by constructing 22 clues that link the crime scene and stolen object to each of the suspects, descriptions, clothes, tools, and vehicles, as illustrated in Fig. 2a. These clues are common to both conditions, and form the ‘core’ set of clues upon which the outcomes will be assessed. To create interdependence between these clues in the treatment world, 55 additional ‘cross-linking’ clues are added that link each of the suspects, vehicles, etc. to one another (Fig. 2b). To create a control world with as much similarity to the treatment world as possible, the cross-linking clues are replaced with 55 ‘spur’ clues which comment on the suspect, vehicle, etc. without creating links (explicit or implicit) to other elements of the mystery (Fig. 2c). Further information can be found in the “Clue Generation Procedure” of the supplement.

Clues are extensively pre-tested to minimize bias from outside world, such that each element of each clue is perceived to be approximately equally likely to be involved in a burglary absent other information. At the start of the game, each clue is present in exactly one player’s notebook.



*Fig. 2: Clues are designed such that in a Treatment world, “core” clues are connected to one another by “cross-linking” clues, and in the Control world, these core clues have no connection to one another.*

Secondly (and orthogonally), the network structure is manipulated with a "low polarization" (dodecahedral) network designed to minimize the characteristic path length of the network and its average clustering, and a "high polarization" (regular connected caveman) network with long characteristic path lengths and high clustering. This manipulation serves as a point of comparison for the effect of interdependent diffusion on social contagion, and allows us to see how the effect of interdependence compares to another effect we are familiar with: the effect of social network structure. The two networks are selected intentionally to represent the extremes from a distribution of networks (that meet the constraints of the experiment), such that the difference between them might represent the maximum effect we could expect to see due to social network structure alone. Further discussion of the choice of network structures can be found in the supplement.



*Fig 3. Social relationships form 20-node regular degree-3 networks chosen to (A) minimize characteristic path length and clustering, and (B) maximize these attributes.*

Together these manipulations create four experimental conditions. The “low polarization” network and “independent” clues form the baseline condition, while the other three conditions reveal the marginal and joint effects of interdependence and network structure on the various outcome measures.

Blocks of four games were constructed with one game in each condition. Within each block, the content of the clues is designed to vary as little as possible between conditions, such that to the fullest extent possible, the clues seen in one position in the network of one game correspond to the clues seen by the same network position in the other games within the block. Individuals are randomly assigned to games and positions within each block. Games within each block are then treated as matched samples. To guard against any latent (systematic) external biases, each block contains a different set of clues generated randomly from a pool of pretested concepts (suspects, vehicles, etc.) representing over 18 million possible mysteries.

The experiment was fully preregistered at <https://osf.io/239ns>, including code for clue generation, the game itself, and for data processing and analysis. All code to replicate the experiment as presented is available at <https://github.com/JamesPHoughton/detective-game-interdependent-diffusion>. Further details of the experimental protocol can be found in the supplement.

**Measures**

While there are many measures in the literature that operationalize the process of polarization *(1,2,4, 9-15 for a sample)*, they generally attempt to represent three basic intuitions. First, that individuals within the same ideological camp come to be more similar to one another. Secondly, that individuals in different ideological camps become more *dis*similar to one another. Lastly that beliefs become associated with one another, such that knowing an individual’s position on one dimension of belief is informative about their position on other dimensions. As my purpose is not to argue that a population is (or is not) polarized, but to suggest that one set of conditions is more generative of polarization than another, many of the complex measures found in the literature do not add value. Instead I report heuristic measures characterizing the above three intuitions.

In order to assess the similarity of individuals who share an ideological camp, we need to know which relationships between individuals to classify as “within-camp”. Absent exogenous labels such as demographic or party, a simple and reproducible way to do this is to measure the similarity between all pairs of individuals (described below), and define a certain percentile as belonging to the same ideological camp. The more exclusive we are (i.e. the higher the percentile) the more conservative the claim that these represent “within-camp” relationships, at the expense of statistical noise in the margins of the distribution of subject-pair similarities. For the 20-person network in this experiment, the 95th percentile value of similarity balances these competing factors well.

Similarly, to assess the level of similarity “across-camps” we can define a percentile that we believe represents relationships between individuals in different ideological camps. The 5th percentile is a conservative choice.

The measure of similarity between individuals must also be selected from numerous options in the literature. The “self-reported” beliefs of participants fall on a continuous scale from 0 to 100, and so it is natural to use Pearson’s correlation on the vectors of individuals’ beliefs. This measure has the advantage of being easily interpretable and having a well-defined range that is independent of the number of features in the vector of attributes being compared, and the negative region of which can be interpreted as expressing dissimilarity. To assess the similarity of the binary “behavioral” data I use the Phi coefficient, an analogous measure to Pearson’s correlation with the same interpretable range.

To measure how strongly an individual’s position on one belief informs us of their position on another belief, we can use the percentage of the total variance in the belief space that can be explained by the first component in a principal component analysis. Intuitively, this measure describes how well the population can be mapped to a “left-right” axis. When the first principal component explains a large fraction of the variance, the population falls tightly along such a political spectrum, and individual’s beliefs along one dimension are strongly predicted by their beliefs on other dimensions.

The ”behavioral” measures are sensitive not only to interdependence and network structure, but also to the average level of diffusion of beliefs. To minimize noise due to differences in the level of activity between games, each of the behavioral measures is assessed compared to what would be expected due to chance, keeping the number of adopters of each clue and the number of clues adopted by each participant fixed.

Each of these measures was computed for each network, and pairwise one-sided T-tests conducted to assess the difference between the baseline condition (independent clues and low polarization network) and the other conditions. Further discussion of these choices can be found under “Choice of Measures” in the supplement.

**Results**

30 batches of 4 games each were conducted over the course of 8 days. 2768 workers were recruited from Amazon mechanical turk, of which 2400 players completed training and played the game. (*45% female; mean 37.1 years old; 27% high-school, 49% bachelors, 16% masters+*). 96.8% of players who began the game completed all steps of the experiment (<0.4% difference in dropout between conditions).

The experiment provides strong evidence that interdependence between beliefs contributes to the population’s alignment along a “left-right” axis, amongst both behavioral and self-report opinion measures (Table 1 and Fig. 4). In the context of this experiment, the effect size is approximately one-sixth the effect of moving from a strongly non-polarizing social network to a strongly polarizing network amongst the behavioral measures, and just over one-third amongst the self-report measures.

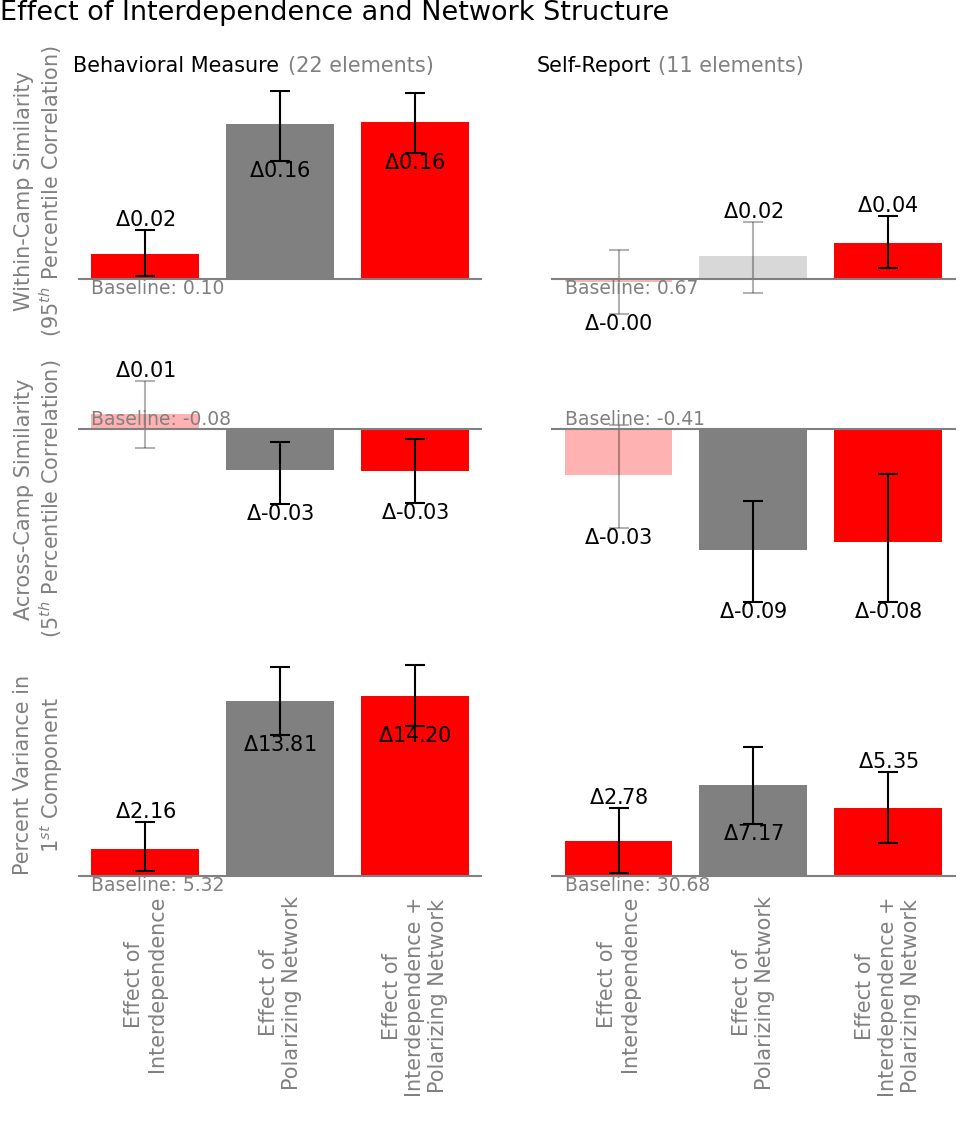
Evidence for increases in within-camp similarity is mixed, with behavioral measures supporting an increase due to polarization again approximately one-sixth of the effect of moving from low to high polarization networks. Self-report measures show no significant effect of either interdependence or network structure alone on the within-camp similarity, although the joint effect is significant and positive.

The experiment did not detect a significant (p < .05) decrease in across-camp similarity due to interdependence, although the effect of network structure was significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Effect over baseline (independent clues and non-polarizing network)** | | | | | | | | |
|  |  |  | Interdependent Clues Alone | | Polarizing Network Alone | | Interdependent Clues and Polarizing Network | |
| Within-Camp Similarity | (95th percentile correlation) | Behavioral1 | +0.0249\*\* | | +0.155\*\*\* | | +0.158\*\*\* | |
| Self-Report1 | | -0.00389 | | +.0231 | | +.0357\*\*\* |
| Across-Camp Similarity | (5th percentile correlation) | Behavioral1 | | +0.0112 | | -.0307\*\*\* | | -.0310\*\*\* |
| Self-Report1 | | -0.0345\* | | -.0904\*\*\* | | -.0844\*\*\* |
| Alignment with Political Axis | (% Variance in 1st Component) | Behavioral1 | | +2.16%\*\* | | +13.8%\*\*\* | | +14.2%\*\*\* |
| Self-Report1 | | +2.77%\*\* | | +7.16\*\*\* | | +5.34%\*\*\* |
| Confidence |  | Self-Report | | +2.17%\*\* | | +0.682% | | +3.74%\*\* |
| Consensus |  | Self-Report | | -0.624% | | +1.75%\* | | +1.03% |
| 1 Preregistered analysis; \*P value <.1, \*\*P value <.05, \*\*\*P value <.01; n=30 matched pairs | | | | | | | | |

Individuals’ overall confidence in their beliefs are slightly increased by clue interdependence, but no strong effect is detected on how individuals perceive consensus among in their social network.

One final result of this experiment may be easy to overlook. Despite the fact that there is no solution to the presented mystery, and each of the clues are symmetric with respect to one another, after only eight minutes of gameplay participants can come to strongly-held beliefs about which party is guilty and how they performed the crime. For example, over half of participants across conditions reported confidence in at least one aspect of the mystery (suspect, vehicle, etc.) of 95% or greater. This in itself is a remarkable demonstration of the way confirmation bias and social contagion can together lead to false confidence in in one’s beliefs.



**Discussion**

The fundamental challenge in designing this experiment was to create a control condition in which beliefs are effectively independent of one another. There are two ways in which beliefs interact in this experiment. The first is that beliefs may be supported by *logical* pathways created by other beliefs (i.e., if you believe that “A red-haired man was seen at the Daly Auction House”, and that “Barnes is a red-haired man”, there is some support for the belief that “Barnes was seen at the Daly Auction House”). The second is a *familiarity* interaction: the more clues an individual holds that reference “Barnes”, the more likely they are to adopt other clues referencing him. In the control condition, the filler clues eliminate logical interaction between the core set of clues. However, they cannot fully eliminate the effect of familiarity.

The difference between treatment and control conditions in this experiment thus is not a perfect test of the difference between interdependent diffusion and independent diffusion, but merely the closest approximation that could be achieved in a naturalistic setting. This experiment should thus be interpreted as setting a lower bound on the effect size that would have been observed in a perfect experiment.

The difficulty of creating a perfect control condition also reminds us of the practical relevance of belief interaction. Even in a laboratory setting in which the researcher has complete power over the social and informational environment, we struggle to create the independent social contagion assumed by much of the literature. Interaction between beliefs is less a product of the beliefs themselves and more a product of the human mind – always seeking connection and meaning in the beliefs it holds. It is reasonable to assume that when multiple beliefs spread in the real world, interaction is the norm rather than the exception.

**References**

1. D. Baldassarri, P. Bearman, Dynamics of political polarization. *Am. Sociol. Rev.* **72** 784-811 (2007).
2. D. DellaPosta *et al.*, Why do liberals drink lattes? *AM. J. Sociol.* **120** 1473-1511 (2015).
3. N. E. Friedkin *et al.*, Network science on belief system dynamics under logic constraints. *Science.* **345**, 321-326 (2016).
4. A. Goldberg, S. Stein, Beyond Social Contagion: Associative Diffusion and the Emergence of Cultural Variation. *Am. Sociol. Rev.* **83** 897-932 (2018).
5. C. T. Butts, Why I know but don’t believe. *Science.* **354** 286-287 (2016).
6. S. E. Parsegov *et al.*, Novel Multidimensional Models of Opinion Dynamics in Social Networks. IEEE Trans. Automat. Contr. **62** 2270-2285 (2017).
7. J. Houghton. Interdependent Diffusion: The Social Contagion of Interacting Beliefs. Doctoral Thesis, Massachusetts Institute of Technology, September 2020.
8. Almaatouq, A., Becker, J., Houghton, J. P., Paton, N., Watts, D. J., & Whiting, M. E. (2020). Empirica: a virtual lab for high-throughput macro-level experiments. *arXiv preprint arXiv:2006.11398*.
9. J. Becker, E. Porter, and D. Centola. The wisdom of partisan crowds. *Proc National Acad Sc*i. **116**, 10717–10722 (2019).
10. Permanyer, I. The conceptualization and measurement of social polarization. J Econ Inequal 10, 45–74 (2012).
11. DiMaggio, P., Evans, J. & Bryson, B. dimaggio.pdf. Am J Sociol 102, 690–755 (1996).
12. Sikder, Orowa, et al. "A minimalistic model of bias, polarization and misinformation in social networks." Scientific reports 10.1 (2020): 1-11.
13. 1.Baldassarri, D. & Gelman, A. Partisans Without Constraint: Political Polarization and Trends in American Public Opinion. *Ssrn Electron J* **114**, 408–446 (2008).
14. Poole, K. T. & Rosenthal, H. The Polarization of American Politics. *J Politics***46,** 1061–1079 (1984).
15. 1.Esteban, J.-M. & Ray, D. On the Measurement of Polarization. *Econometrica* **62**, 819 (1994).

**Supplement:**

**Secondary Analyses**

The preregistration for this experiment contained a secondary analysis to assess whether the effect of interdependence on polarization was mediated by an increased tendency to adopt beliefs from those with whom an individual already holds beliefs in common. The experiment as realized did not have sufficient power to assess the mediation claim.

The preregistration also included a secondary analysis to test whether the effect of interdependence was moderated by social network structure. A significant negative interaction effect was detected for the effect on alignment with a “left-right” axis amongst self-reported opinions, but interaction effects were not significant amongst the other measures. A negative interaction effect would suggest that interdependence and social network structure operated via similar mechanisms, or that the effect could ‘saturate’.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Interaction Effect** | | | | | | | |
|  |  |  | Interaction Effect | |
| Within-Camp Similarity | (95th percentile correlation) | Behavioral1 | -0.0223 | |
| Self-Report1 | | -0.0164 | |
| Across-Camp Similarity | (5th percentile correlation) | Behavioral1 | | -0.0114 | |
| Self-Report1 | | +0.0406 | |
| Alignment with “Left-Right” Axis | (% Variance in 1st Component) | Behavioral1 | | -1.76% | |
| Self-Report1 | | -4.60%\*\* | |
| Confidence |  | Self-Report | | +0.887% | |
| Consensus |  | Self-Report | | -0.0925% | |
| 1 Preregistered analysis; \*P value <.1, \*\*P value <.05, \*\*\*P value <.01 | | | | | | | |

**Gameplay**

1. Participants are shown a consent screen (Fig. S1) instructing them about how the HIT operates and what they will need to complete, along with a consent statement.

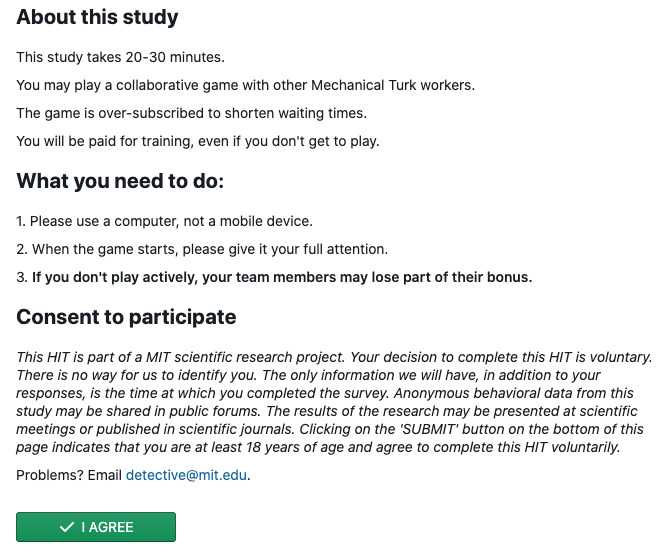
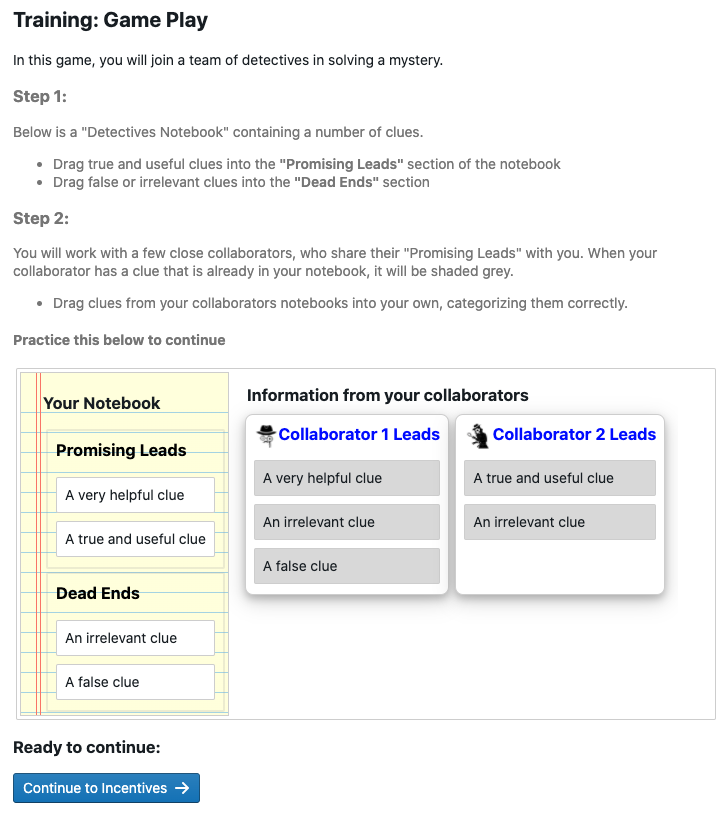


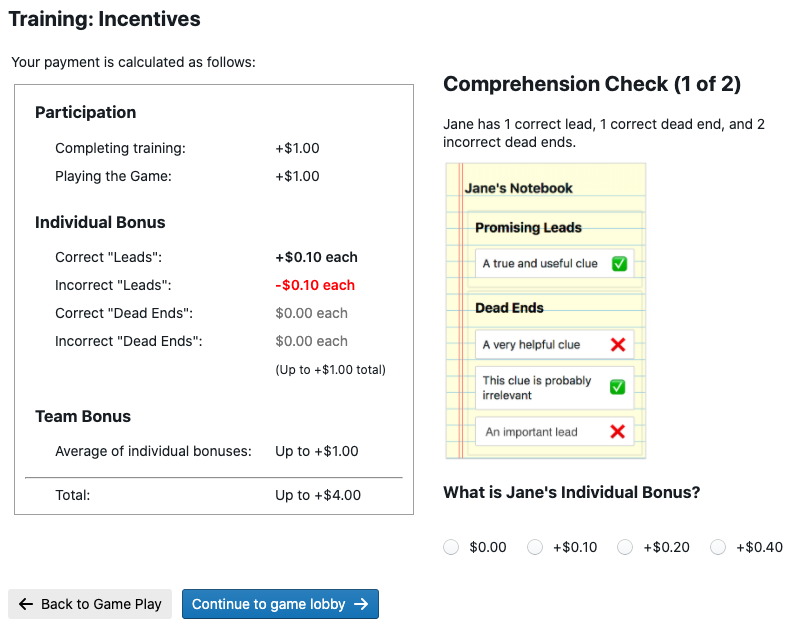
Fig S1. Consent to participate

1. The first training screen (Fig. S2) instructs participants in how to interact with the Detective Game interface. They are asked to sort clues into “Promising Leads” and “Dead Ends” by dragging and dropping them into labeled sections of their “Detective’s Notebook”. In addition to the clues each individual is seeded with, they also see which clues two of their collaborators have categorized as “Promising Leads”. Each participant must correctly sort the practice clues before they can continue to the next training screen.



*Fig. S2: Training screen 1 after participant completion*

1. The second training screen (Fig. S3) teaches participants how they will be rewarded for their performance. Individuals are told that they will receive $0.10 for each clue correctly categorized as a promising lead, and will be penalized $0.10 for each clue categorized as a promising lead that is actually false. They are also told that they will be rewarded for their team’s average performance, receiving the average of all players’ individual bonuses as a Team Bonus. These incentives encourage individuals to carefully sort clues according to their best estimate of their veracity, and to share clues with their neighbors that they believe will improve the team’s collective sensemaking ability. Setting the reward for success to be equal to the penalty for mistakes works to encourage participants to most accurately assess each statement, rather than ‘hedge’ by keeping too many or too few clues. Participants are compensated $1 for training.



*Fig 4: Training Screen S3 - Incentives (Individual)*

1. After completing training (taking between 2 and 4 minutes), participants enter a waiting room until there are 40 individuals who have completed training and are ready to play. The training is oversubscribed so that if some participants are unable to complete the training the game can still launch.
2. When the game launches, the 40 players are divided into two groups, and assigned to locations in two identical social networks. Each individual is given a “Detective’s Notebook” in which 4 clues start in the “Promising Leads” section. They are also show a “Police Bulletin” (Fig. S4) which gives them background information about the mystery and reminds them of their task. Showing the participant their own clues and the mystery premise before launching them into the game helps them orient to the task.

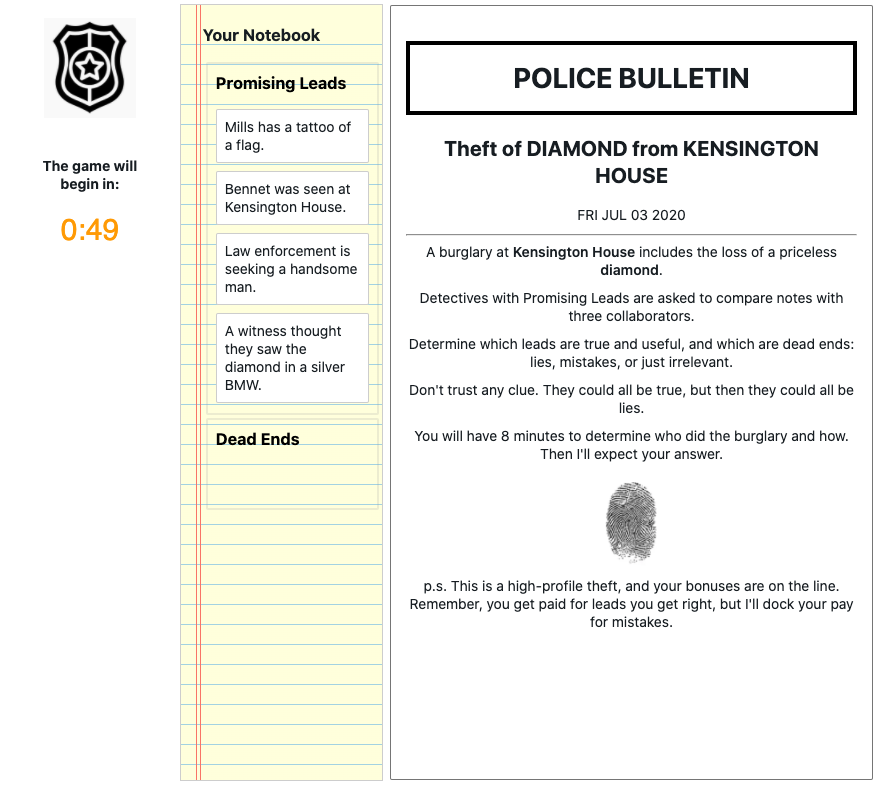


Figure S4: Exposition for the mystery

1. When the game launches, the “police bulletin” is replaced with the “Promising Leads” sections of their neighbors’ notebooks, showing the participants 16 unique clues at the start of the game. Individuals at corresponding positions in the two social networks are given clues that are as similar as possible while allowing for the intervention. These are shown for players in the treatment and control conditions in Figs. S5a and S5b respectively.

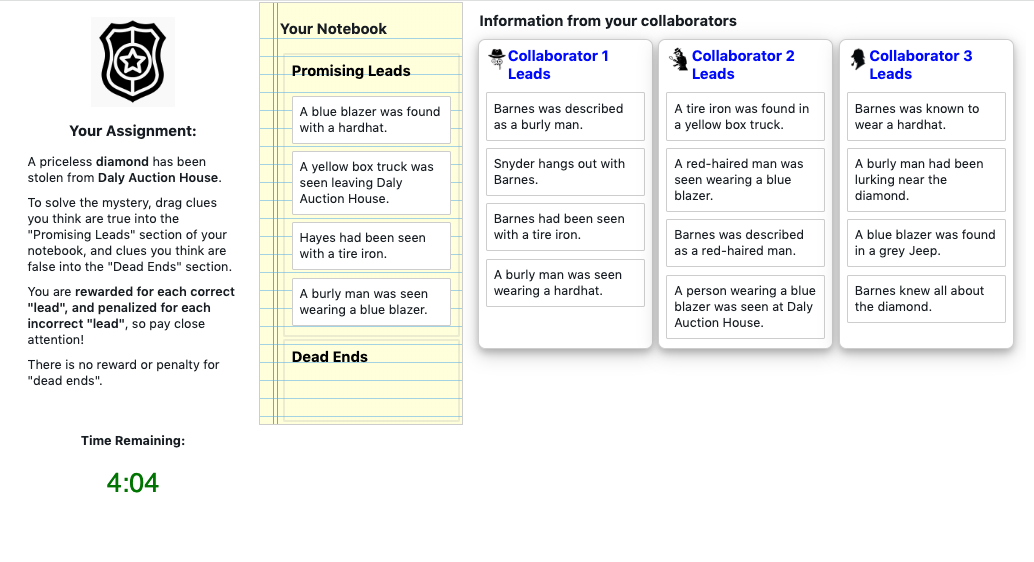


Fig S5a: Game screen – Treatment case

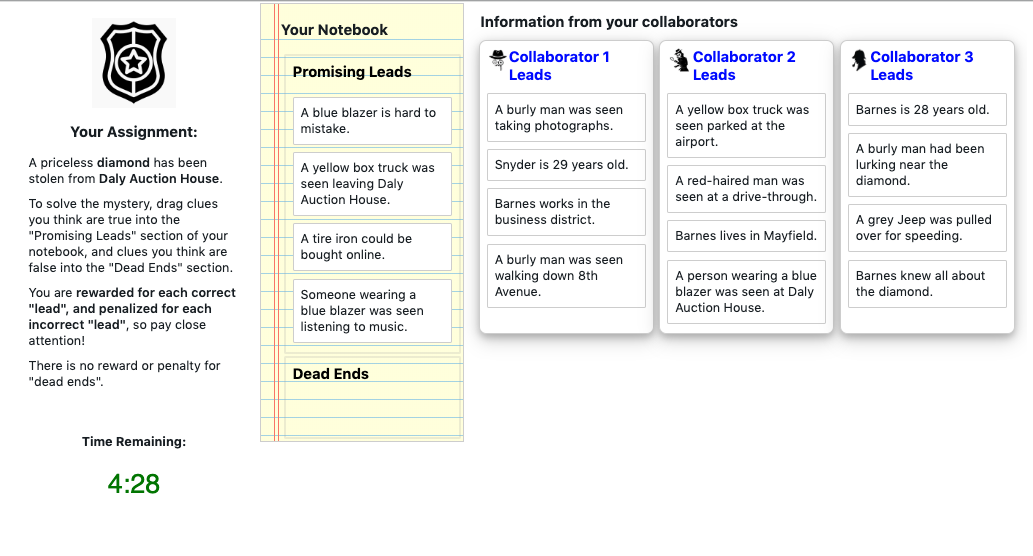


Fig S5b: Game screen – Matching control case

1. The game is played in real-time over 8 minutes. When a participant changes their “Promising Leads”, their neighbors immediately see the change on their own screen. The starting clues of every individual are recorded, and every change to every player’s “Detective Notebook” is logged, such that the state of every player’s notebook can be reconstructed at each moment in the game. Participants are compensated $1.00 for playing the game.
2. Following the game, participants are asked to assess using a slider how likely it is that certain individuals referenced in the game were the burglar, and how likely it is that they used various tools, vehicles, and disguises in the task. The first few of these questions are shown in Fig. S6a. Sliders are labeled from Extremely Unlikely to Extremely Likely, and their positions recorded on a scale from 0 to 100. Participants are also asked to assess their confidence in their solution, and their estimate of the level of consensus among their team, both using similar sliders, as shown in Fig. 7b.

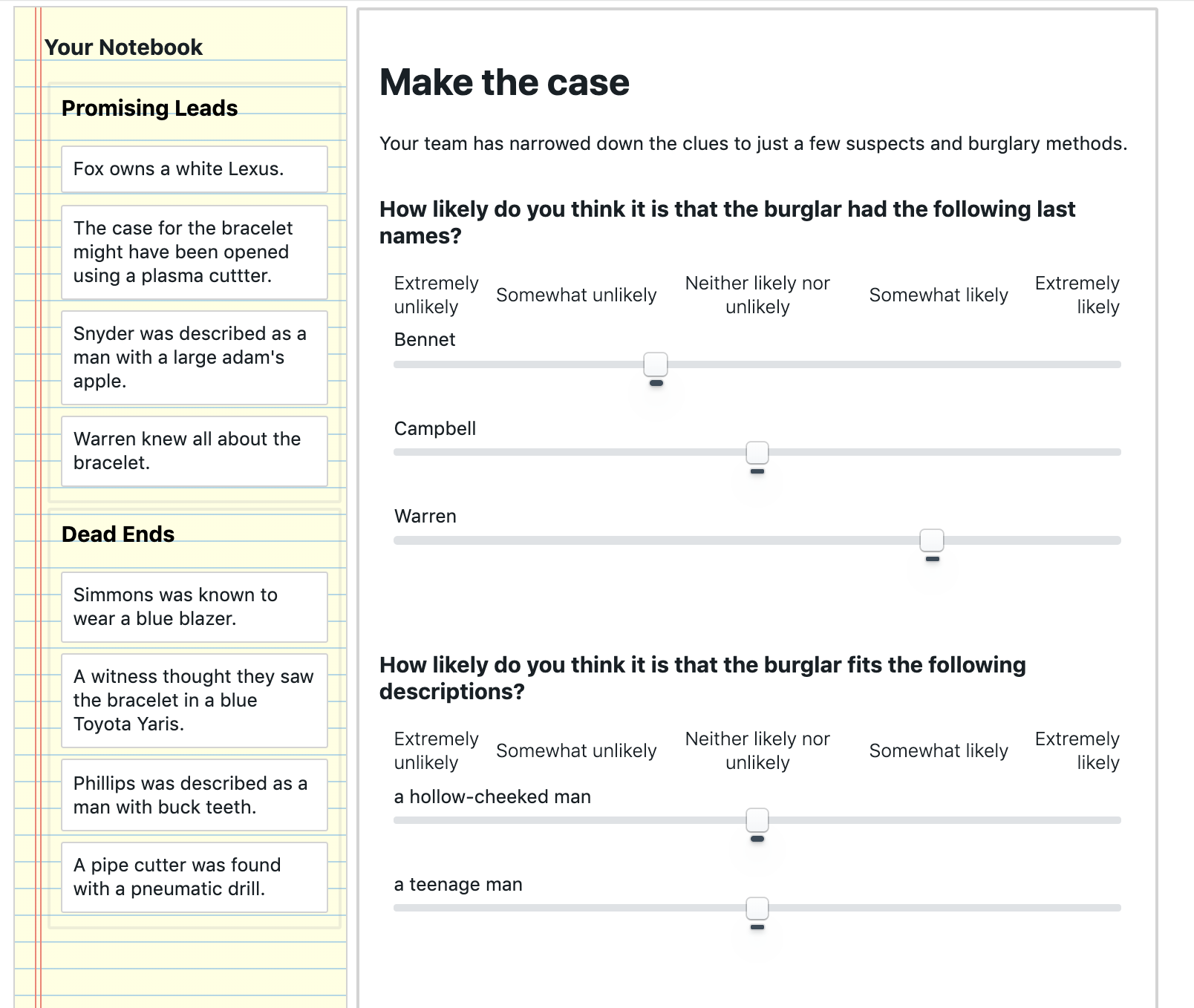


Fig. S6a: Post-game survey screen – Make the case for who committed the burglary

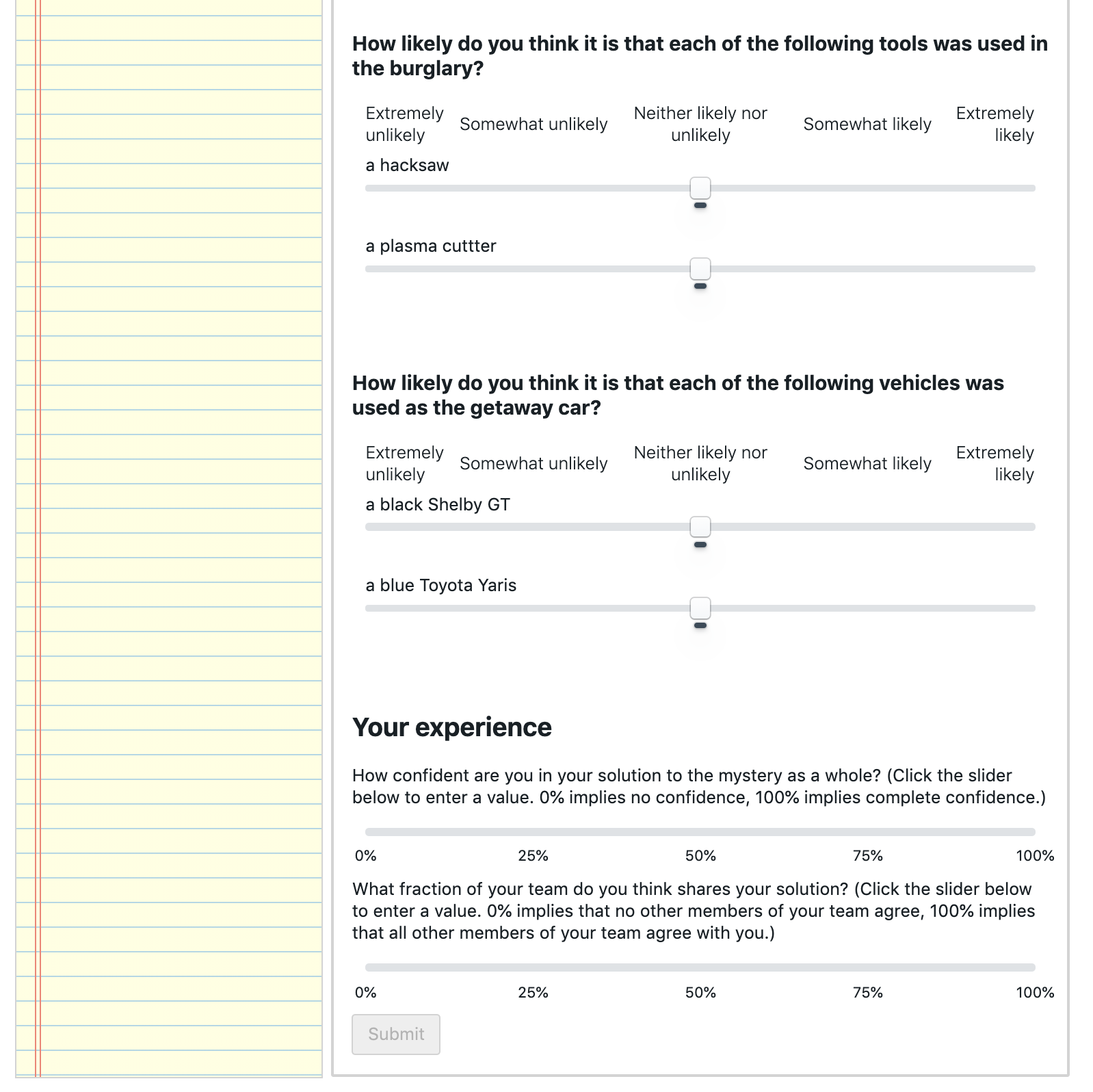


Fig S6b: Post-game survey screen – Assess confidence and consensus

1. After “Making the Case”, individuals are told that they were part of an experimental condition in which none of the clues were “False”, and they are rewarded $0.10 for each clue in the “Promising Leads” section of their notebook, along with $0.10 for each of the average number of clues their teammates categorized as a “Promising Lead”. Participants are given a completion code to collect their bonuses, and given an (optional) opportunity to report any problems with the game, and describe their strategy.

**Informational Manipulation**

1. This experiment manipulates the structure of clues within the mystery game, to create a treatment condition in which the clues interact strongly with each other, and a control condition that limits those interactions while preserving as much similarity with the treatment condition as possible.
2. Clues are constructed in three waves. The first wave is identical for treatment and control condition, and is illustrated in Fig. S7. Clues are created which link ‘hub’ concepts (including a crime scene and a stolen object) to ‘rim’ concepts (including three suspects, two articles of clothing, two physical descriptions, two tools, and two vehicles). For example “**Hayes** was seen at the **Daly Auction House**” or “The case for **the diamond** might have been opened using **a circular saw**”. A pool of rim concepts was constructed in pre-test to minimize any population bias towards one concept or another. For more details see section 18: *Clue generation procedure*. “Spoke” clues are independent of one another, as they only interact via association with the crime scene and stolen object – items that are known in advance to be relevant to the mystery. There are 11 rim concepts and 2 hub concepts, and so 22 spoke clues.

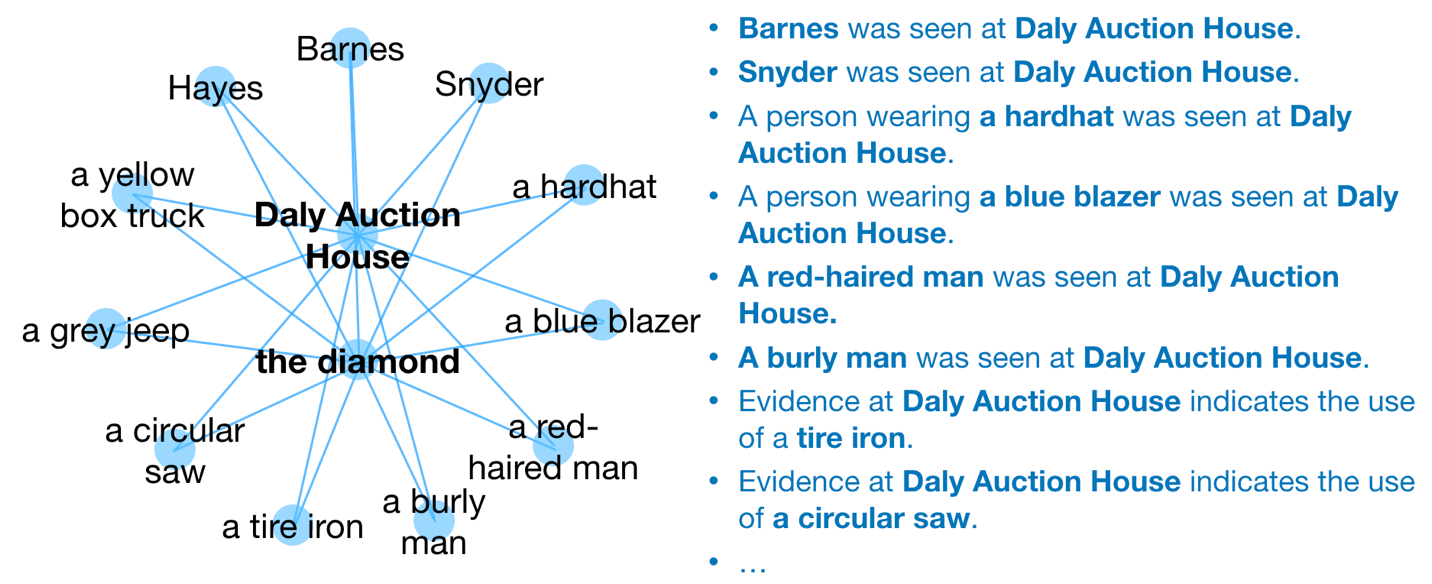


Fig. S7: “Spoke” clues connect rim concepts to hub concepts

1. In the treatment case, the second wave of clue construction creates “cross-link” clues, which connect each of the spoke clues to one another (e.g. “**Hayes** owns a **circular saw**”). These cross-link clues create interdependence between the spoke clues, and allow for clues to logically support one another (e.g. if I believe that “A **burly man** was seen at the **Daly Auction House**” and “**Barnes** is a **burly man**”, then I am more receptive to the idea that “**Barnes** was seen at the **Daly Auction House**”). A cross-link clue connects each of the 11 rim concepts to the other rim concepts, for a total of 55 unique cross-link clues, as shown in Fig. S8.

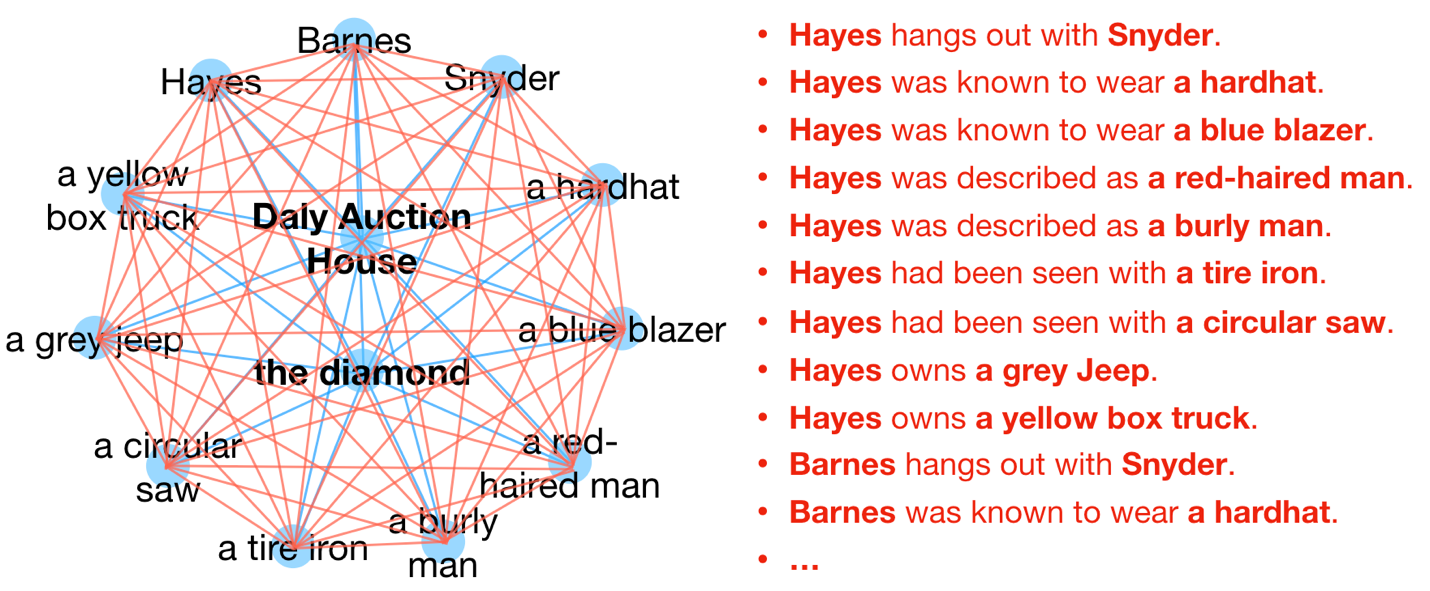


Fig S8. “Cross-Link” clues connect rim concepts to one another

1. In the control case, the second wave of clue construction creates “spur” clues that connect to the rim concepts, but do not connect to other clues (Fig S9). There are the same number of ‘spur’ clues in the control case as there are ‘cross-link’ clues in the treatment case: 55. By connecting to the rim concepts (rather than being disconnected altogether) these clues help separate the effect of interdependence manifest as logical relationships between clues from the effect of the frequency of each rim concept in the set of clues. The content of the spur clues was selected in pre-test to have a uniform impact on participants judgement of the rim element to which they connect.

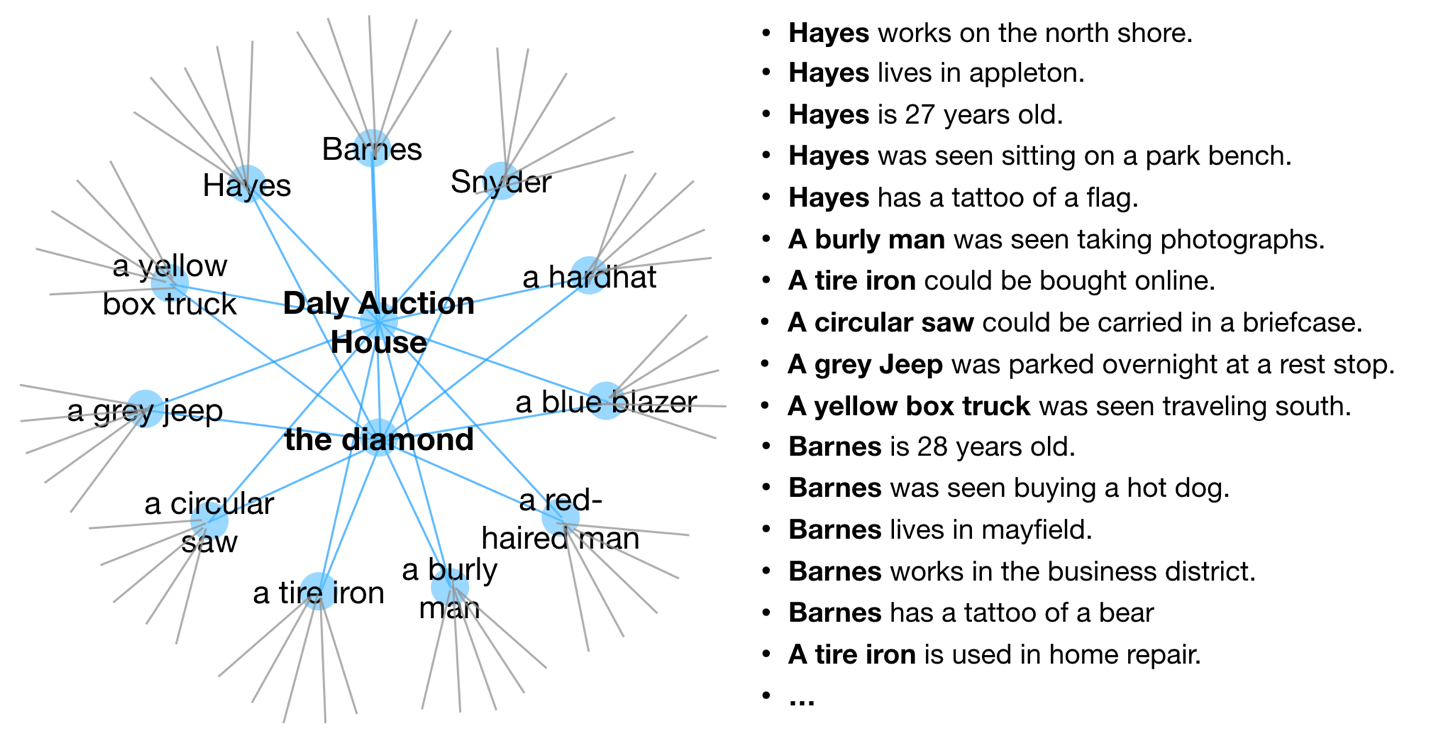


Fig S9: “Spur” clues fill the place of “cross-link” clues without creating links between rim concepts, while still allowing for multiple exposures to rim concepts.

1. The first and second waves of clue construction create 77 unique clues. As there are 20 individuals in each treatment within each game, 80 clues are needed to give each individual 4 starting clues. The third wave of clue construction fills the 3 remaining spaces with the clue connecting the crime scene to the stolen object (e.g. The **diamond** was stolen from the **Daly Auction House**.) This is redundant information, as all participants are told this at the start of the game.
2. Clues are randomly assigned to individuals at the start of the game. Each position in the control world is given the same “spoke” clues as their corresponding position in the treatment world, and a “spur” clue that shares one concept with the “cross-link” clue in the corresponding slot in the treatment game.
3. The clues to be used in the game, and their assignment to locations in network structures is included in the code supplement to this preregistration.

**Clue generation procedure**

1. Clues are constructed from a bank of concepts (11 Stolen Objects, 11 Crime Scenes, 33 Names, 22 descriptions, 22 articles of clothing, 22 tools, and 22 vehicles) and set of relationships (e.g. {Name} owns a {vehicle}, {A witness thought they saw {stolen object} in {vehicle}) that forms a complete network between all concepts. These are randomly shuffled such that different clues can be generated for each game.
2. The bank of concepts was constructed by starting with a pool of 403 candidate concepts including names, clothing, vehicles, etc. A pretest survey was conducted in which Amazon Mechanical Turk workers were asked to assess how likely a number of concepts was to have been used in a generic burglary. Individuals saw a subset of the concepts and were asked to give their gut reactions using a slider from Extremely Unlikely to Extremely Likely, as illustrated in Fig S10 below. In total, 139 participants rated each of 403 candidate concepts between 20 and 30 times. Participants in the pretest were paid $1.25 for a task which took each participant an average of about 4 minutes.

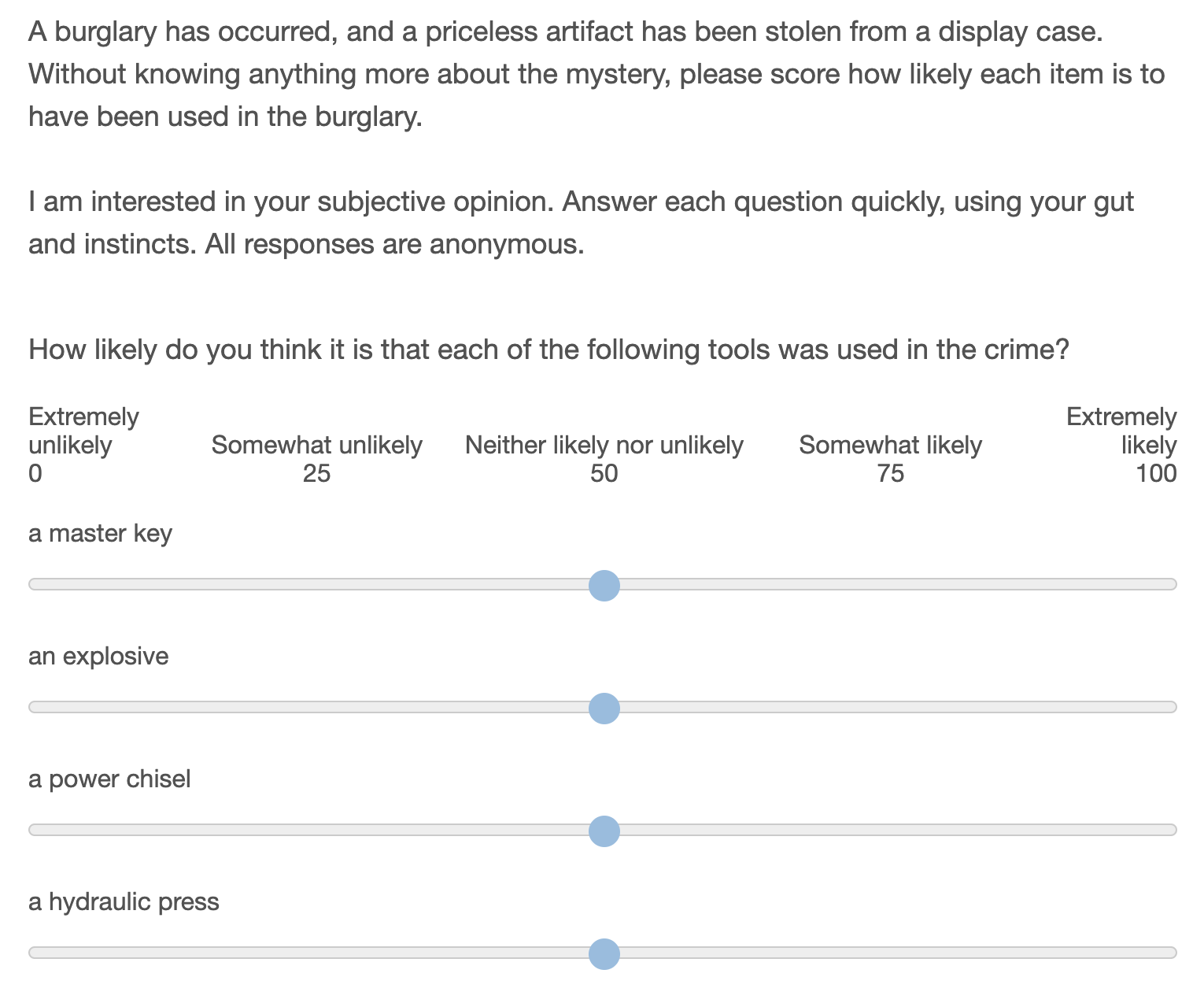


Fig S10: Pretest the perceived likelihood of each item being used in a crime

1. The pool of candidate names in the pretest represents the subset of the 200 most popular last names in the United States with a racial composition of between 50% and 80% ‘White’, as recorded in the 2000 US census. This selection is made to minimize the possibility of racial biases in the results. Additionally, names which are also common first names are excluded (e.g. “Stewart” or “Ross”) as are names which also serve as descriptors or adjectives in other clues (e.g. “Green”, “White”, or “Young”).
2. The remaining candidate concepts were written such that they would be as independent from one another as possible (e.g. I do not include both “a fat man” and “an overweight man” as these are synonymous, nor both “an old man” and “a man with grey hair” as these are perceived to go together.)
3. From the pretest results, I selected a subset of concepts that are perceived to be as likely as one another to be used in a burglary. (This helps to ensure that we do not see games in which all participants adopt “a set of lock picks” as a tool in the burglary, and reject “a machette”, just because lock picks are easier to imagine being used in a burglary.) The final selection was made by taking the subset of beliefs that minimized the difference in mean value of pretest survey responses when responses are normalized for each individual, and cross-checking against the means of the raw responses. 11 concepts were selected for each time the concept (name, vehicle, etc.) is used in a game.
4. A similar pretest survey was conducted to select ‘spur’ clue concepts from a pool of candidates.

**Clue Concept Pool**

1. The pool of concepts was extensively pretested to ensure that amongst the pool of experiment participants at the time the experiment was conducted, each element in a category was equally likely to be associated with a burglary.
2. “Core” Elements that may appear in both treatment and control games:

|  |  |  |  |
| --- | --- | --- | --- |
| **CrimeScene** | **StolenObject** | **Suspect** | **Clothing** |
| the art museum | the painting | Collins | a pair of overalls |
| the Pine Street Gallery | the statue | Hawkins | a wool hat |
| Kensington House | the relic | Mills | a blue denim jacket |
| the Asper Casino | the bracelet | Cooper | a tracksuit |
| the Danforth Hotel | the antique | Moore | a pair of skinny-jeans |
| Knight Secure Storage | the necklace | Bennet | a black scarf |
| the Daly Auction House | the watch | Mitchell | a motorcycle helmet |
| the Kentwood Mansion | the diamond | Stevens | a black leather jacket |
| the Dalhoff Estate | the opal | Wagner | a pair of ripped jeans |
| the Darrowby Country Club | the crystal | Edwards | a blue long sleeve shirt |
| DeRolfe Jewelers | the jewel | Rice |  |
|  |  | Roberts |  |
|  |  | Daniels |  |
|  |  | Warren |  |
|  |  | Sullivan |  |

|  |  |  |
| --- | --- | --- |
| **Appearance** | **Tool** | **Vehicle** |
| a long-haired man | a hacksaw | a yellow box truck |
| a pot-bellied man | a serrated knife | a blue Chevrolet Corvette |
| a partially-bald man | a set of hex keys | a green Mazda 3 |
| a grey-haired man | a masonry drill | a silver BMW |
| a short man | a circular saw | a silver VW Jetta |
| a well-groomed man | a blowtorch | a black Hummer |
| a man with sideburns | an impact wrench | a white Ford Fusion |
| a heavily-scarred man | a tire iron | a blue Toyota Yaris |
| a blonde-haired man | a pipe cutter | a white Toyota Avalon |
| a handsome man | a sledgehammer | a blue Honda Fit |

1. “Filler” elements that may appear in control games:

|  |  |  |  |
| --- | --- | --- | --- |
| **appearanceInjury** | **appearanceRemoved** | **appearanceReported** | **appearanceStreet** |
| a broken arm | a museum | waiting in a dark alley | Maple Avenue |
| a fractured kneecap | a public library | shouting at 11pm | Lincoln Boulevard |
| a concussion | a bar | sitting in a tree | Chestnut Street |
| a fractured rib | a restaurant | painting graffiti | Church Street |
| minor burns | a residence | vandalizing a vending machine | Hill Street |
| a drug overdose | a party | carrying a large bag | Ninth Avenue |

|  |  |
| --- | --- |
| **appearanceWanted** | **carBehavior** |
| Law enforcement is seeking | driving after midnight |
| Officers are asking questions about | with someone sleeping in the back seat |
| Private security companies have been warned to... | with darkly tinted windows |
| Airport security has been asked to look out for | parked in a lot for multiple nights |
| Police are interviewing witnesses about | taking the back streets |
| Information is wanted about | with its hood up on the roadside |

|  |  |  |  |
| --- | --- | --- | --- |
| **carBuy** | **carDamage** | **carEnterprise** | **carTicketed** |
| in a wholesale auction | a broken headlight | a club | an expired registration |
| at a police auction | a broken grill | a massage parlor | parking in a loading zone |
| from a classified ad | damaged suspension | a strip mall | driving without headlights |
| at an estate sale | a missing wing mirror | a laundromat | a broken tail light |
| from a used-car salesman | the airbags deployed | a delicatessen | illegal parking |
| from a junk-yard | a broken axle | a hotel | running a stop sign |

|  |  |  |
| --- | --- | --- |
| **clothingActivity** | **clothingDamage** | **clothingDiscoverer** |
| pacing back and forth | cut into pieces | a gym owner emptying abandoned lockers |
| entering a machine room | with tire marks on it | a dockworker moving shipping pallets |
| pulling an object out of a gutter | with frayed edges | a store worker breaking down cardboard boxes |
| looking through binoculars at night | burned in a fire | a journalist uncovering a story |
| getting into a taxi | discolored with bleach | a postal worker emptying a mailbox |
| climbing on a bridge | caked in mud | theater staff cleaning up after a movie |

|  |  |  |
| --- | --- | --- |
| **clothingFootage** | **clothingWith** | **suspectAge** |
| at a bus stop | a list of tools | 33 years old |
| in the woods | a home-made electronic device | 37 years old |
| in the park | a pair of rubber gloves | in their early 30's |
| at a campsite | an inter-city train schedule | in their mid 20's |
| on a bridge | the stub of a bus ticket | in their late 20's |
| on the golf course | an envelope containing GPS coordinates | in their late 30's |
|  |  | in their early 20's |
|  |  | 29 years old |
|  |  | 36 years old |

|  |  |
| --- | --- |
| **suspectConviction** | **suspectMeans** |
| drug possession | was trained as a welder |
| fraud | installs security systems |
| drug distribution | was trained as a goldsmith |
| running a Ponzi scam | has worked as an automotive repossession agent |
| embezzlement | has worked as a security guard |
| identity theft | worked at a pawn shop |
| perjury | has worked as an armored car driver |
| shoplifting | has worked for an import/export company |
| arson | worked for an alarm company |

|  |  |  |
| --- | --- | --- |
| **suspectMotive** | **suspectTattoo** | **toolDamage** |
| has paid hush-money to a former lover | a compass | showing signs of misuse |
| has family connections to organized crime | a bear | with minor damage |
| is deep in payday-loan debt | a heart | with burn marks |
| has an expensive drug habit | a flower | that had been damaged falling from a height |
| has a gambling addiction | a clock | covered in sawdust |
| wrote a revolutionary manifesto | a star | disassembled into pieces |
| had been involved in gang activity | a dog |  |
| has large gambling debts | a crown |  |
| has a heroin addiction | a flag |  |

|  |  |  |
| --- | --- | --- |
| **toolFound** | **toolUse** | **toolWith** |
| buried in debris | access maintenance crawlspaces | with a pair of work gloves |
| in an abandoned house | open an upper-story window | with safety features removed |
| in a trash compactor | bypass an alarm system | that had been painted black |
| in a garage | deactivate a motion sensor | with gunpowder residue |
| in a creek | disassemble an alarm panel | wrapped in newspaper |
| beside a road | circumvent a lock | wrapped in tape to make it quieter |

|  |
| --- |
| **toolrandom** |
| could be used by one person |
| was shown in news coverage of another burglary |
| could be concealed in a backpack |
| leaves distinctive marks if used carelessly |
| is often used by thieves |
| has been used in prior burglaries   1. Treatment condition clue templates |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **StolenObject\_1** | **Suspect\_1** | **Suspect\_2** | **Suspect\_3** | **Clothing\_1** | **Clothing\_2** |
| **CrimeScene\_1** | {StolenObject\_1} was kept in a case at {CrimeScene\_1} | {Suspect\_1} was seen at {CrimeScene\_1} | {Suspect\_2} was seen at {CrimeScene\_1} | {Suspect\_3} was seen at {CrimeScene\_1} | A person wearing {Clothing\_1} was seen at {CrimeScene\_1} | A person wearing {Clothing\_2} was seen at {CrimeScene\_1} |
| **StolenObject\_1** | - | {Suspect\_1} knew all about {StolenObject\_1} | {Suspect\_2} knew all about {StolenObject\_1} | {Suspect\_3} knew all about {StolenObject\_1} | A person wearing {Clothing\_1} had been seen lurking near {StolenObject\_1} | A person wearing {Clothing\_2} had been seen lurking near {StolenObject\_1} |
| **Suspect\_1** |  | - | {Suspect\_2} hangs out with {Suspect\_1} | {Suspect\_1} hangs out with {Suspect\_3} | {Suspect\_1} was known to wear {Clothing\_1} | {Suspect\_1} was known to wear {Clothing\_2} |
| **Suspect\_2** |  |  | - | {Suspect\_3} hangs out with {Suspect\_2} | {Suspect\_2} was known to wear {Clothing\_1} | {Suspect\_2} was known to wear {Clothing\_2} |
| **Suspect\_3** |  |  |  | - | {Suspect\_3} was known to wear {Clothing\_1} | {Suspect\_3} was known to wear {Clothing\_2} |
| **Clothing\_1** |  |  |  |  | - | {Clothing\_1} was found with {Clothing\_2} |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Appearance\_1** | **Appearance\_2** | **Tool\_1** | **Tool\_2** | **Vehicle\_1** | **Vehicle\_2** |
| **CrimeScene\_1** | {Appearance\_1} was seen at {CrimeScene\_1} | {Appearance\_2} was seen at {CrimeScene\_1} | Evidence at {CrimeScene\_1} indicates the use of {Tool\_1} | Evidence at {CrimeScene\_1} indicates the use of {Tool\_2} | {Vehicle\_1} was seen leaving {CrimeScene\_1} | {Vehicle\_2} was seen leaving {CrimeScene\_1} |
| **StolenObject\_1** | {Appearance\_1} had been lurking near {StolenObject\_1} | {Appearance\_2} had been lurking near {StolenObject\_1} | The case for {StolenObject\_1} might have been opened using {Tool\_1} | The case for {StolenObject\_1} might have been opened using {Tool\_2} | A witness thought they saw {StolenObject\_1} in {Vehicle\_1} | A witness thought they saw {StolenObject\_1} in {Vehicle\_2} |
| **Suspect\_1** | {Suspect\_1} was described as {Appearance\_1} | {Suspect\_1} was described as {Appearance\_2} | {Suspect\_1} had been seen with {Tool\_1} | {Suspect\_1} had been seen with {Tool\_2} | {Suspect\_1} owns {Vehicle\_1} | {Suspect\_1} owns {Vehicle\_2} |
| **Suspect\_2** | {Suspect\_2} was described as {Appearance\_1} | {Suspect\_2} was described as {Appearance\_2} | {Suspect\_2} had been seen with {Tool\_1} | {Suspect\_2} had been seen with {Tool\_2} | {Suspect\_2} owns {Vehicle\_1} | {Suspect\_2} owns {Vehicle\_2} |
| **Suspect\_3** | {Suspect\_3} was described as {Appearance\_1} | {Suspect\_3} was described as {Appearance\_2} | {Suspect\_3} had been seen with {Tool\_1} | {Suspect\_3} had been seen with {Tool\_2} | {Suspect\_3} owns {Vehicle\_1} | {Suspect\_3} owns {Vehicle\_2} |
| **Clothing\_1** | {Appearance\_1} was seen wearing {Clothing\_1} | {Appearance\_2} was seen wearing {Clothing\_1} | {Clothing\_1} was found with {Tool\_1} | {Clothing\_1} was found with {Tool\_2} | {Clothing\_1} was found in {Vehicle\_1} | {Clothing\_1} was found in {Vehicle\_2} |
| **Clothing\_2** | {Appearance\_1} was seen wearing {Clothing\_2} | {Appearance\_2} was seen wearing {Clothing\_2} | {Clothing\_2} was found with {Tool\_1} | {Clothing\_2} was found with {Tool\_2} | {Clothing\_2} was found in {Vehicle\_1} | {Clothing\_2} was found in {Vehicle\_2} |
| **Appearance\_1** | - | {Appearance\_2} was seen with {Appearance\_1} | {Appearance\_1} was seen with {Tool\_1} | {Appearance\_1} was seen with {Tool\_2} | {Appearance\_1} was seen driving {Vehicle\_1} | {Appearance\_1} was seen driving {Vehicle\_2} |
| **Appearance\_2** |  | - | {Appearance\_2} was seen with {Tool\_1} | {Appearance\_2} was seen with {Tool\_2} | {Appearance\_2} was seen driving {Vehicle\_1} | {Appearance\_2} was seen driving {Vehicle\_2} |
| **Tool\_1** |  |  | - | {Tool\_1} was found with {Tool\_2} | {Tool\_1} was found in {Vehicle\_1} | {Tool\_1} was found in {Vehicle\_2} |
| **Tool\_2** |  |  |  | - | {Tool\_2} was found in {Vehicle\_1} | {Tool\_2} was found in {Vehicle\_2} |
| **Vehicle\_1** |  |  |  |  | - | {Vehicle\_2} was found near {Vehicle\_1} |

1. Control condition clue templates

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **StolenObject** | **Suspect\_1** | **Suspect\_2** | **Suspect\_3** | **Clothing\_1** | **Clothing\_2** |
| **spoke1** | {StolenObject\_1} was kept in a case at {CrimeScene\_1} | {Suspect\_1} was seen at {CrimeScene\_1} | {Susect\_2} was seen at {CrimeScene\_1} | {Suspect\_3} was seen at {CrimeScene\_1} | A person wearing {Clothing\_1} was seen at {CrimeScene\_1} | A person wearing {Clothing\_2} was seen at {CrimeScene\_1} |
| **spoke2** | - | {Suspect\_1} knew all about {StolenObject\_1} | {Suspect\_2} knew all about {StolenObject\_1} | {Suspect\_3} knew all about {StolenObject\_1} | A person wearing {Clothing\_1} had been seen lurking near {StolenObject\_1} | A person wearing {Clothing\_2} had been seen lurking near {StolenObject\_1} |
| **spur1** | - | {Suspect\_1} has a tattoo of {suspectTattoo\_1} | {Suspect\_2} has a tattoo of {suspectTattoo\_2} | {Suspect\_3} has a tattoo of {suspectTattoo\_3} | A policeman saw someone in {Clothing\_1} {clothingActivity\_1} | A policeman saw someone in {Clothing\_2} {clothingActivity\_2} |
| **spur2** | - | {Suspect\_1} {suspectMotive\_1} | {Suspect\_2} {suspectMotive\_2} | {Suspect\_3} {suspectMotive\_3} | Someone wearing {Clothing\_1} was seen on security footage {clothingFootage\_1} | Someone wearing {Clothing\_2} was seen on security footage {clothingFootage\_2} |
| **spur3** | - | {Suspect\_1} is {suspectAge\_1} | {Suspect\_2} is {suspectAge\_2} | {Suspect\_3} is {suspectAge\_3} | {Clothing\_1} was discovered by {clothingDiscoverer\_1} | {Clothing\_2} was discovered by {clothingDiscoverer\_2} |
| **spur4** | - | {Suspect\_1} {suspectMeans\_1} | {Suspect\_2} {suspectMeans\_2} | {Suspect\_3} {suspectMeans\_3} | {Clothing\_1} was found with {clothingWith\_1} | {Clothing\_2} was found with {clothingWith\_2} |
| **spur5** | - | {Suspect\_1} has a prior conviction for {suspectConviction\_1} | {Suspect\_2} has a prior conviction for {suspectConviction\_2} | {Suspect\_3} has a prior conviction for {suspectConviction\_3} | Forensics identified {Clothing\_1} {clothingDamage\_1} | Forensics identified {Clothing\_2} {clothingDamage\_2} |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Appearance\_1** | **Appearance\_2** | **Tool\_1** | **Tool\_2** | **Vehicle\_1** | **Vehicle\_2** |
| **spoke1** | {Appearance\_1} was seen at {CrimeScene\_1} | {Appearance\_2} was seen at {CrimeScene\_1} | Evidence at {CrimeScene\_1} indicates the use of {Tool\_1} | Evidence at {CrimeScene\_1} indicates the use of {Tool\_2} | {Vehicle\_1} was seen leaving {CrimeScene\_1} | {Vehicle\_2} was seen leaving {CrimeScene\_1} |
| **spoke2** | {Appearance\_1} had been lurking near {StolenObject\_1} | {Appearance\_2} had been lurking near {StolenObject\_1} | The case for {StolenObject\_1} might have been opened using {Tool\_1} | The case for {StolenObject\_1} might have been opened using {Tool\_2} | A witness thought they saw {StolenObject\_1} in {Vehicle\_1} | A witness thought they saw {StolenObject\_1} in {Vehicle\_2} |
| **spur1** | {appearanceWanted\_1} {Appearance\_1} | {appearanceWanted\_2} {Appearance\_2} | {Tool\_1} {toolrandom\_1} | {Tool\_2} {toolrandom\_2} | {Vehicle\_1} was reported {carBehavior\_1} | {Vehicle\_2} was reported {carBehavior\_2} |
| **spur2** | A constable noticed {Appearance\_1} on {appearanceStreet\_1} | A constable noticed {Appearance\_2} on {appearanceStreet\_2} | {Tool\_1} was found {toolWith\_1} | {Tool\_2} was found {toolWith\_2} | {Vehicle\_1} was recently purchased {carBuy\_1} | {Vehicle\_2} was recently purchased {carBuy\_2} |
| **spur3** | {Appearance\_1} was reported {appearanceReported\_1} | {Appearance\_2} was reported {appearanceReported\_2} | {Tool\_1} could be used to {toolUse\_1} | {Tool\_2} could be used to {toolUse\_2} | {Vehicle\_1} was ticketed for {carTicketed\_1} | {Vehicle\_2} was ticketed for {carTicketed\_2} |
| **spur4** | {Appearance\_1} was treated for {appearanceInjury\_1} | {Appearance\_2} was treated for {appearanceInjury\_2} | An FBI agent found {Tool\_1} {toolFound\_1} | An FBI agent found {Tool\_2} {toolFound\_2} | An officer identified {Vehicle\_1} at {carEnterprise\_1} | An officer identified {Vehicle\_2} at {carEnterprise\_2} |
| **spur5** | {Appearance\_1} was forcibly removed from {appearanceRemoved\_1} | {Appearance\_2} was forcibly removed from {appearanceRemoved\_2} | A forensics report contained {Tool\_1} {toolDamage\_1} | A forensics report contained {Tool\_2} {toolDamage\_2} | {Vehicle\_1} was found with {carDamage\_1} | {Vehicle\_2} was found with {carDamage\_2} |

**Participant Recruitment and Compensation**

1. Participants were recruited from Amazon Mechanical Turk workers residing in the US or Canada over 18 years of age. Workers must have completed at least 100 HITs and have a 90% or better approval rating. Recruitment and compensation were handled using TurkPrime (cloudresearch.com), and the platform was also used to ensure that workers could only participate once.
2. For blocks 0-2, recruitment took place in the hour preceding a game launch via a timed “sign-up” HIT. This recruitment strategy had been shown to be effective during pilots, but did not scale well when multiple blocks were run during the same day.
3. For blocks 3-29, recruitment took advantage of a “panel” of workers who had previously indicated willingness to be notified of upcoming games. This panel was expanded through ongoing paid and unpaid recruitment HITs during the period the experiments were run. Panel members were notified via email at the beginning of each day when experiments would take place, and again 10 minutes prior to the launch of a game.
4. At the launch of a game time, 400 HITs were made available to any worker meeting the qualifications.
5. Participants were compensated $0.10 for accepting the game HIT, in addition to $1 for training, $1 for playing the game, and up to $2 in bonuses.
6. Participants who trained but were unable to play were eligible to attempt to play again. Those who completed training and entered the game were blocked from participating in future games via an exclusion qualification.
7. The game takes about 20 minutes to play, including training, waiting room, and follow-up. The average payout is approximately $4.00, for an hourly rate of approximately $12.00/hr. Participants who train but are unable to play take about 5 minutes before they are bumped, and earn $1.10, for an approximate hourly rate of $13.20. Fig. S11 shows the number of participants active in different parts of the game at different times for a 2-condition pilot test. There are necessarily some individuals who are dropped from training when the game launches, as the number of individuals who will show up to a game is unpredictable.

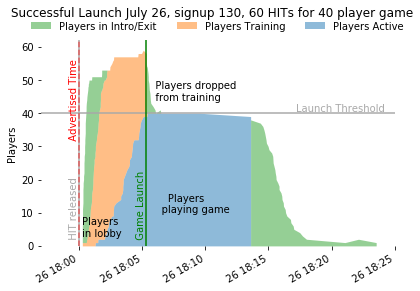
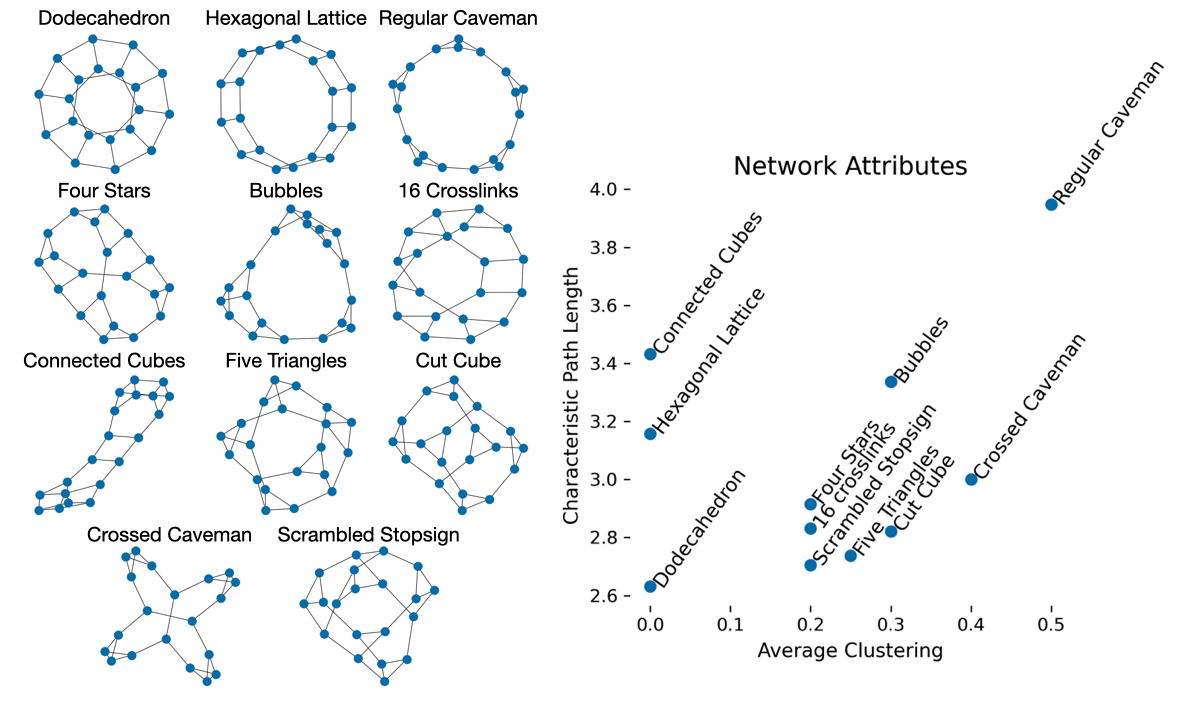


Fig. S11: Active players by stage

**Infrastructure**

1. Blocks 0-2 were hosted on a pair of Meteor Galaxy Quad servers at url `detective.meteorapp.com` with a single MongoDB Atlas M20 Database hosted on Google Cloud platform. The M20 db is necessary to handle the large number of IOPS generated in the game, far in excess of the 120iops provides by AWS servers.
2. Blocks 3-29 were hosted with training screens on a pair of Galaxy Quad servers at url `detective.meteorapp.com`. Upon completing training, individuals were redirected to a “game” server `detectivea.meteorapp.com`, also using a pair of Quad servers. When the block of games on server A filled and launched, newly arriving participants were redirected to `detectiveb.meteorapp.com`, a second game server hosting the next block in the series, and on to `detectivec` and `detectived`.
3. The chained approach to serving games allowed individuals to be assigned to treatment conditions upon completing training, rather than the default Empirica behavior of assigning to treatment before training. As the training was identical between conditions, this meant that when a game launched, there were fewer individuals rejected from the game because they happened to be assigned to an overflowing condition.
4. The chained approach also allowed the number of games to scale to accommodate a larger number of simultaneous arrivals. As each pair of quad servers (with its own database) was only capable of running a single block of 4 conditions at one time.

**Display considerations**

1. **Presentation of social information:** All at once  
   A ‘scrolling feed’ type information display has recency and primacy effects, and opens questions about how we should aggregate social information from multiple players. Showing all information at once, in the order that it is sorted by the neighbor, eliminates the effect of alternate ordering sequences.
2. **Number of neighbors:** 3  
   The number of neighbors is limited by the size of the screen and an individual’s ability to process information. The minimum number of neighbors for a non-trivial social network is 3, and is also a reasonable number for managing the cognitive load in the game.
3. **Number of starting clues:** 4  
   Fewer starting clues are preferred for minimizing cognitive load on individuals. With three neighbors, individuals see 16 clues all at once when they start playing. This takes about 30 seconds to read through and understand. The next increment (5 starting clues) gives 20 items for an individual to process at game start, which starts to be cognitively overwhelming.
4. **Number of players:** 20  
   Larger numbers of players are better for generalizability and seeing an effect size. Smaller numbers mean we can afford more replications. There needs to be enough players that the mean shortest-path-length is greater than two, to realistically represent multi-stage diffusion.
5. **Network shape:** Dodecahedron and regular connected caveman (k=5)  
   Eleven symmetric candidate networks were evaluated with n=20 and degree=3. Of this set, the Dodecahedral network minimizes the average shortest path between individuals with no network clustering, and represents a social network we should expect to exhibit low polarization *a priori*. A regular connected caveman network maximizes the characteristic path length and exhibits strong clustering, and so we expect to exhibit more polarization *a priori.* Descriptions of each of these networks are included in the preregistration code.
6. **Number of unique clues in the game:** 78
   1. From an information diversity perspective, more clues is better. With 4 starting clues and 20 players, we can have up to 80 unique clues in the game. 13 nodes yields 78 clues, and the two spots remaining can be filled with the given link between the crime scene and the stolen object.
7. **Number of times each clue is represented**: 1\*
   1. Each clue should be represented an equal number of times so as not to bias the network to one particular outcome.
   2. \*The ‘given’ clue that the object was stolen from the crime scene is included 3 times to fill out the 80 slots in the game.
8. **Length of game**: 8 minutes
   1. Pilot trials were conducted with durations of 5 and 8 minutes. It was observed that participants remained engaged for 8 minutes, and felt rushed with 5.
9. **Survey format**: Empty sliders
   1. Rather than force individuals to make a discrete choice between suspects/vehicles etc., a slider allows individuals to assess a degree of confidence in their assessment of the solution to the mystery.

**Choice of Measures**

1. **Self-report similarity**: Pearson Correlation  
   Correlation is a natural measure when we have a fixed number of continuous measures of each subject, as is the case in the self-report, and there is precedence for this use in recent literature *(4)*. It is useful to have a measure with a fixed range (-1,1) and which is readily interpretable.
2. **Behavioral similarity:** Phi coefficient
3. The phi coefficient corresponds to Pearson correlation when measures are binary, and has the same interpretable (-1,1) range. This is appropriate for a universe in which there are a finite number of beliefs measured, but would be less appropriate as the number of adopted beliefs becomes a very small fraction of the total number of possible beliefs.
4. **Polarization**
   1. Percent of Variance present in first principal component  
      This measure corresponds to the notion of “constraint” articulated by Dimaggio et al. *(6)*. In their paper they describe Chronbach’s alpha and the PCA measure both providing similar measures of constraint. I have chosen the PCA measure here as more interpretable and well known among computational social scientists.
   2. 5TH and 95th percentile similarities  
      There are a number of different measures in the literature that try to capture the notion that with polarization, the most similar individuals become more self-similar, and the least similar individuals move further away from one another. The fact that no single measure has emerged as the leader hints at problems with each. Variance *(see 1,11)* captures heterogeneity between individuals, but not clustering into camps. Kurtosis *(see 1,11)* is predicated on a bimodal distribution. The “gap” statistic *(4)* is one of dozens of ways of assessing the quality of a machine learning clustering algorithm. When the identities of camps are already known the difference of means between groups can be used *(9,11).*
      1. As I do not need to identify the groups themselves, or compare to external datasets, it is sufficient for me to merely report what each of these other measures is trying to approximate: the similarity that is found within groups, and that which is found across groups. As I am only interested in the relative differences between conditions (or for the same population over time) then I can arbitrarily designate a threshold for which comparisons will be considered ‘within-group’ or ‘across groups’. This provides a much more intuitive demonstration of increasing polarization than the measures found in literature.
      2. The closer the chosen thresholds are to the tails of the distribution, the more conservative the claim that the comparisons beyond this threshold are appropriately “within” or “across” groups. At the same time, we need enough samples included in the set to minimize noise due to the finite number of comparisons. In this 20-participant social network, the 95th and 5th percentiles correspond to 10 comparisons between individuals.

**Measurements**

1. Each drag event that results in a change in a player’s notebook (i.e. presence of a clue in a notebook section OR change of order within a notebook section) is logged. Logging information includes the ID of the clue being dropped, the source for the drag event (which exposing player or notebook the belief came from), the destination for the drag event (which notebook the clue is being dragged into), the position within the destination notebook that the clue will take (i.e. its numerical position in the notebook) and the time at which the drop event occurred.
2. The final state of all notebooks in the game is logged. Together with the initial state, this provides a check that all events are logged properly.
3. Each individual provides a self-report of the degree to which they believe each of the 11 “rim” concepts is to be connected to the crime, collected using an empty slider from “Extremely Unlikely” to “Extremely Likely”. Slider positions are captured as an integer value between 0 and 100.
4. Individuals report their confidence in their solution on a scale from 0 to 100 using a blank slider.
5. Individuals report the fraction of their team they think shares their solution on a scale from 0 to 100 percent, using a blank slider.

Measures for Mediation Effect:

1. While the mediation analysis yielded no significant results, the measures to be used in the analysis were fully specified in the preregistration:
2. Cox proportional hazard rate regressions were conducted to assess the effect of treatment and the proposed mediating variable on individuals' likelihood of adopting a clue to which they are exposed. The regressors used in the hazard rate analysis are listed below:
   1. "Logical support for a clue" is formalized as the number of length-2 pathways between the two concepts in the candidate clue that can be constructed from existing clues in the participant’s “Promising Leads”. A positive value of this regressor serves as a manipulation check.
   2. "Similarity of an individual to those exposing them to a clue" is the total number of clues that are shared between the individual and all of their exposing neighbors, and represents the mediating variable.
   3. "Familiarity with a clue’s concepts" is formalized as the number of clues within a participant’s “promising leads” that reference the rim concept of the candidate (spoke) clue, and represents a competing form of interaction to "logical support"
   4. "Social exposure" is formalized as the number of individuals exposing the candidate clue to the participant.
   5. "New clue volume" is formalized as the number of clues that an individual is exposed to, but has not yet categorized.
   6. "Number of current clues" is the number of clues the participant has already classified as a “promising lead”.
   7. "Prior rejection of a clue" is a binary variable indicating whether the individual has previously categorized the clue as a “dead end”.
   8. Timing dummies are included for the startup period (t<30s), two mid-game periods (30s <= t < 180s and 180s <= t < 420s) and the final minute (t >= 420s).
   9. A (gaussian) random effect is allowed for each participant, to account for the fact that some players are more active than others.
   10. The mediation of the treatment is computed as the fraction of the effect from the "Logical support for a clue" hazard upon each primary outcome variables that goes via the "Similarity of an individual to those exposing the clue" hazard. A value is computed for each regressor for each game, and the influence of these regressors on the outcome is computed using a structural equation model, accounting for the effect of network structure. Code for this analysis is found in the preregistration repository under `detective-game-interdependent-diffusion/analysis/Mediation Analysis.ipynb`.
3. The initial state of the game is not representative of an individual’s choice to adopt a belief and expose their neighbors to it, as the starting belief sets are randomly assigned. For all hazard rate regressions in the mediation analysis, I looked only at individual\*exposure events that happen after t0, the start of the game. Likewise the hazard rate regressions were limited to the 22 “core” clues.

**Missing data**

1. As the game is played in real-time, the effect of a participant ‘dropping out’ during game-play is equivalent to them holding their beliefs fixed for the remainder of the game. As it is impossible to distinguish these two behaviors during game-play, I identify a drop-out as any player failing to submit the post-game survey.
2. When an individual fails to complete the post-game survey, aggregate results for their condition will be calculated based upon the remaining players for use in assessing the primary hypotheses and the moderation analysis. Aggregate results for the paired comparison conditions were calculated as the average of all same-sized subsets of the comparison condition.

**Changes from preregistered procedure**

1. The original strategy of recruiting individuals with a pre-game “signup” HIT proved too cumbersome to execute with multiple one-shot games in a day. Instead, I drew on a pool of workers from US/Canada that had been recruited for other experiments in the past. This allowed me to recruit large numbers to show up at fewer times during the day, and to launch multiple simultaneous blocks. However as the panel was already built using US/Canada participants, and the specific locaitons were not tied to the panel, this required me to expand the pool of participants from US only to include Canada. I felt that this addition would not significantly influence the likelihood of success of the experiment or it’s generalizability, as the phenomenon under study is not expected to behave differently in different populations, and the material in the study does not require any special outside knowledge.
2. The analysis code for comparing end-of-game measures was designed originally to account for dropouts in a game (say, a treatment game) by averaging over same-sized subsets of the final survey results in the paired (control) game. However, with the addition of the second social network condition, the code as preregistered did not account for the fact that comparisons were now being made between a control game with one network and a control game with a second network. The code was expanded to make the correct comparisons.
3. While the interaction analysis presented above was included in the text of the preregistration, the code for computing the measure was mistakenly omitted. Additional code was written for this analysis.