**Open Science Foundation Experiment Preregistration**

[Study Information](#_3o9ssh2swz09)

[Design Plan](#_4mzf79vx2q6j)

[Sampling Plan](#_hu8o0vkz41nk)

[Variables](#_pec3rgxfolor)

[Analysis Plan](#_3mtn7m44krsg)

[Other](#_6wujw18ggcuz)

### Study Information

1. Title
   1. **The effect of belief interaction on group polarization in networked experiments**
2. Authors
   1. James P. Houghton
      1. MIT Sloan School of Management
3. Description
   1. Simulations of social contagion *(1-5)* predict that when beliefs interact with one another as they diffuse:
      1. New sociological processes emerge at the individual, dyad, group, and community level
      2. Particular sets of beliefs come to be widely shared amongst subsets of the population (i.e. “worldviews” emerge)
      3. The most similar individuals become yet more similar, and the least similar individuals increase their differences (i.e. “camps” become more self-consistent, and differentiated from one another)
      4. Individuals’ beliefs can be more easily described as positions along a “left-right” axis (i.e. a “political spectrum” emerges)
   2. 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.
   3. 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.
   4. This laboratory experiment places human actors anonymously in an artificial social network, and gives them 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.
   5. It is impossible to conduct a perfect experiment in which beliefs are allowed to interact in the treatment world, and the same identical beliefs are constrained to not interact within the minds of individuals in the control world. However, by manipulating the structure of the belief set, we can encourage 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 we can substitute filler beliefs in place of the cross-linking beliefs, and maintain independence between at least the core subset of beliefs. We can then test the above predictions on this core subset of beliefs. This forms a conservative estimate of the effect of belief interaction on diffusion outcomes.
4. Hypotheses
   1. Manipulation Check (M0): Under treatment conditions in which beliefs interact, an individual’s hazard of adopting a belief to which they are exposed (i.e. a “candidate” belief) is increased when the concepts referred to by the candidate belief are linked by other beliefs in the individuals existing belief set. Under control conditions, no such links are possible, and so no effect can be discerned.
   2. Hypothesis 1 (H1): Under control conditions, an individual’s hazard of adopting a candidate belief is increased by their familiarity with the concepts the belief refers to (H1a). Under treatment conditions, other signals also influence adoption decisions, and so familiarity will have less impact than under control conditions (H2b).
   3. Hypothesis 2 (H2): An individual’s adoption hazard is increased when they hold other beliefs in common with exposing alters (H2a), signaling the action of “agreement cascades” in which similarity between individuals is reinforced through subsequent behavior. This effect will be stronger under the treatment condition than under the control condition (H2b).
   4. Hypothesis 3 (H3): Under treatment conditions, the similarity amongst the most similar individuals (i.e. within-camp similarity) will be larger than under control conditions. We will see this effect both using participants’ final belief sets as a “behavioral” measure of this similarity (H3a), and post-game surveys as a “self-reported” measure (H3b). The magnitude of this effect will be positively correlated with the magnitude of the “agreement cascade” mechanism (H3m).
   5. Hypothesis 4 (H4): Under treatment conditions, the similarity amongst the least similar individuals (i.e. across-camp similarity) will be decreased below control conditions. We will see this effect both using participants’ final belief sets as a “behavioral” measure of this similarity (H4a), and post-game surveys as a “self-reported” measure (H4b). The magnitude of this effect will be positively correlated with the magnitude of the “agreement cascade” mechanism (H4m).
   6. Hypothesis 5 (H5): The extent to which individuals fall along a “left-right” axis will be greater under treatment conditions than under control conditions. We will see this effect both using participants’ final belief sets as a “behavioral” measure of this similarity (H5a), and post-game surveys as a “self-reported” measure (H5b). The magnitude of this effect will be positively correlated with the “agreement cascade” mechanism (H5m).
   7. Hypothesis 6 (H6): Individuals will report more confidence in their beliefs (H6a) and more consensus in the population (H6b) under treatment conditions than under control conditions.
   8. Hypothesis 7 (H7): Hypotheses 3-5 continue to hold in a social network that contributes to polarization (H7a) although the net effect of interdependence will be weaker in an already polarizing network structure (H7b).

### Design Plan

1. Study type
   1. This study uses an online multiplayer game as part of a controlled laboratory experiment. Subjects are randomly assigned to be part of a ‘treatment’ condition or a matched ‘control’ condition. The difference between conditions is compared across 30 different matched pairs. The study is replicated using two different social networks, one designed to minimize factors expected to contribute to polarization (characteristic path length and clustering) and the other designed to maximize these factors.
2. Blinding
   1. Participants in the study do not know how the experiment is manipulated to create a treatment and a control condition, nor do they know the outcomes of interest in the study. However, the nature of the manipulation is such that if participants were aware of the study design they would be able to identify whether they were assigned to treatment or control condition.
   2. All interaction between the experimenter and the participants is mediated by a uniform web interface. Data analysis is necessarily conducted with knowledge of the treatment condition of each participant. Both treatment and control conditions are subject to the same analysis. The analysis code is submitted with this preregistration and has no interpretive component.
   3. All participants are anonymously recruited and compensated.
3. Study design
   1. Within each instance of this experiment (block), 80 individuals will use their browser to simultaneously access the study website. They will be randomly assigned to positions within four separate social networks of 20 individuals, such that any individual has an equal chance of being assigned to any position in either network.
   2. Two of these networks are “dodecahedral” networks with short average path lengths and no clustering. Two of the networks are “regular connected caveman” networks (5 clusters of 4 participants) with long average path lengths and relatively high clustering.
   3. One network of each type will receive a “treatment” set of beliefs, while the other receives a matched “control” set, identical in every way except for the manipulation (described below).
   4. 30 blocks will be conducted over the course of one week, at various times of day.
   5. Individuals are restricted to playing at most one game.

### Sampling Plan

1. Existing data
   1. Several pilot tests have been conducted during development to ensure that the experimental platform operates as expected, and that the data collection process is correct. The results of these pilot tests have been used to test the analysis code included within this preregistration. The data from these pilot tests will be excluded from analysis of the final experiment, but made available upon publication.
2. Data collection procedures
   1. Participants will be recruited from Amazon Mechanical Turk workers residing in the US over 18 years of age. Workers must have completed at least 100 HITs and have a 90% or better approval rating. Recruitment and compensation will be handled using TurkPrime (cloudresearch.com), and the platform will also be used to ensure that workers only participate once.
   2. Between 1 and 2 hours prior to the launch of a game, a “Qualification HIT” will be available to 120 persons, who will be asked to sign up to play by submitting the code “I will show up at <launch time>”. Participants are compensated $0.05 for signing up. (Fig 1)

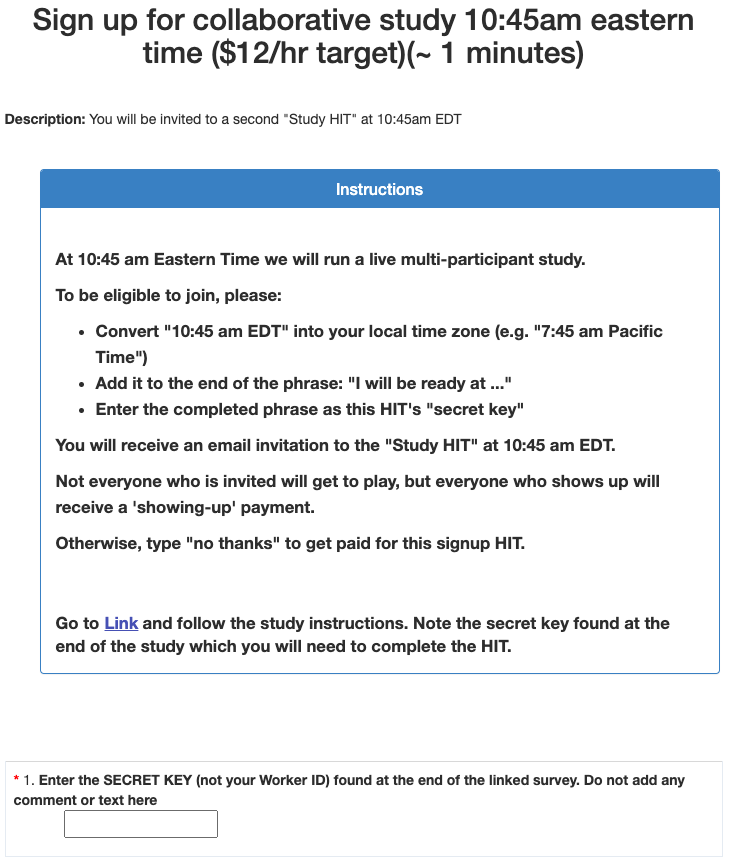


Fig.1 Signup Task

* 1. At the launch time, HITs will be made available to all who signed up to play, and these individuals will receive an automated email instructing them to proceed to the game. Participants are compensated $0.10 for showing up at the launch time. Pilot tests indicate that the recruitment strategy has about 75% yield.
  2. Participants will be shown a consent screen (instructing them about how the HIT operates and what they will need to complete, along with a consent statement.

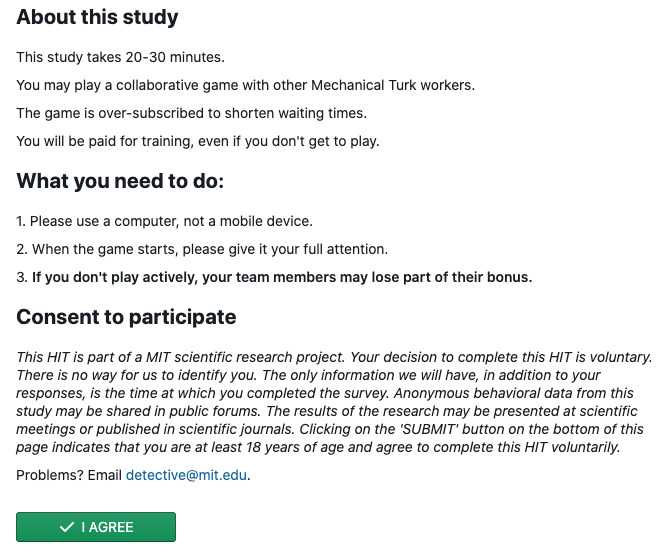
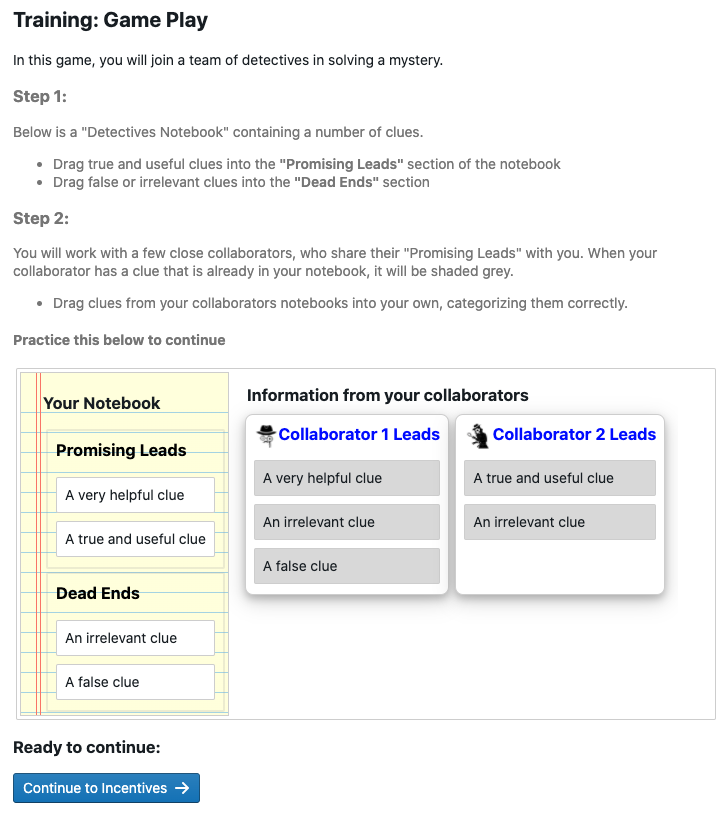


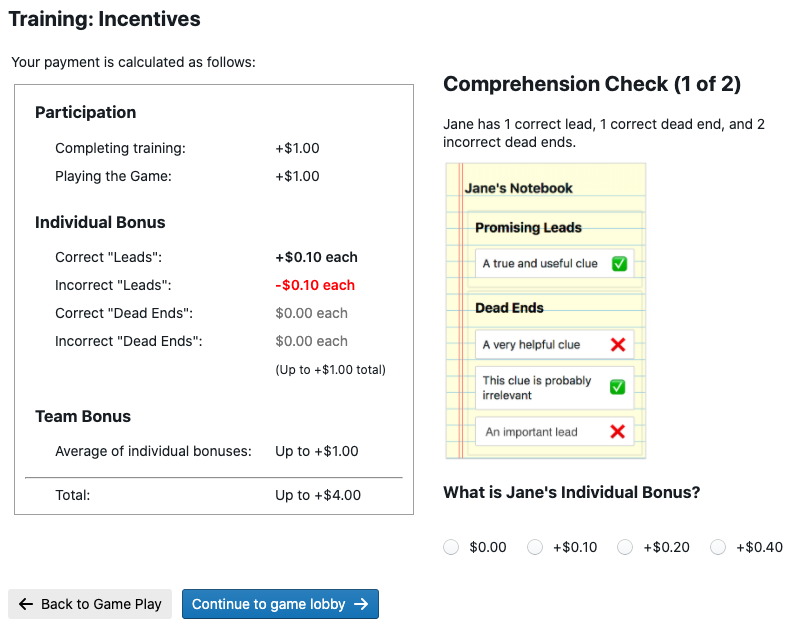
Fig 2. Consent to participate

* 1. The first training screen (Fig. 3) 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. 3: Training screen 1 after participant completion*

* 1. The second training screen (Fig. 4) 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 2 - 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. 5) 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.

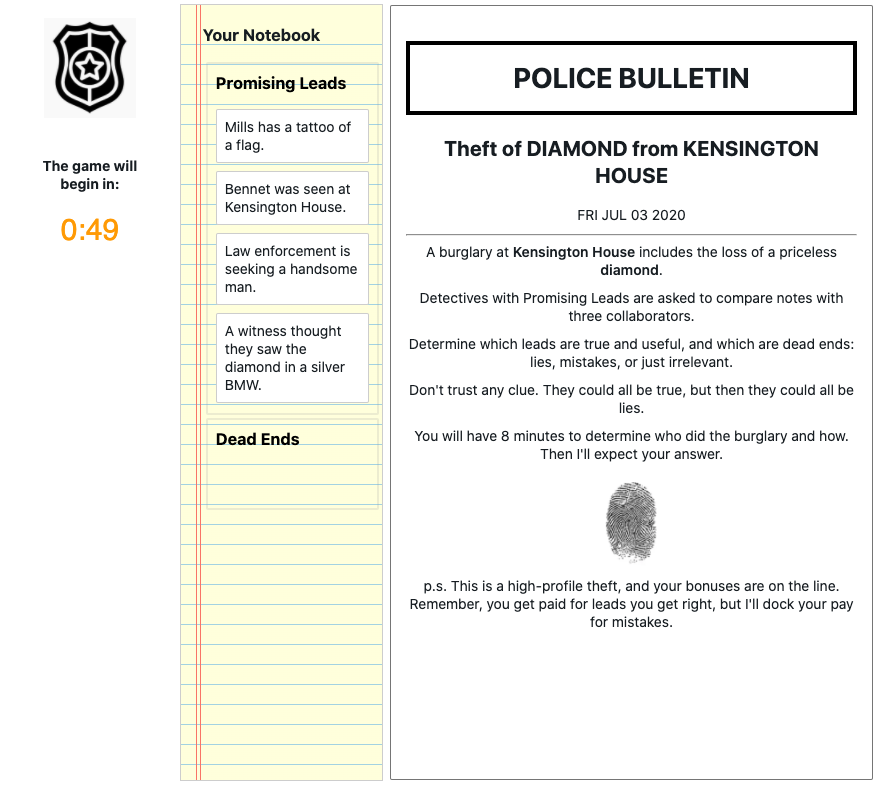


Figure 5: Exposition for the mystery

* 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 6a and 6b respectively.

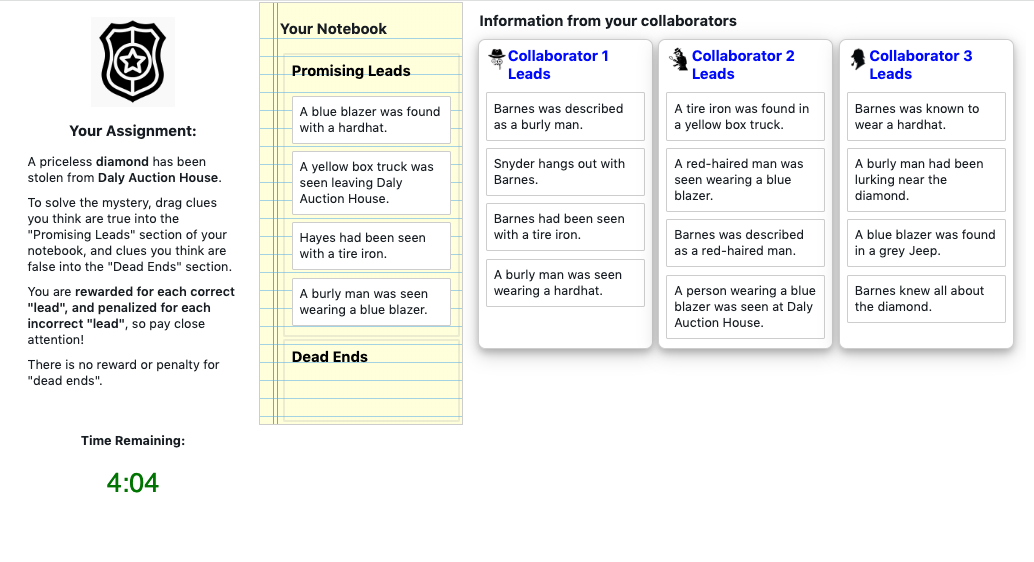


Fig 6a: Game screen – Treatment case

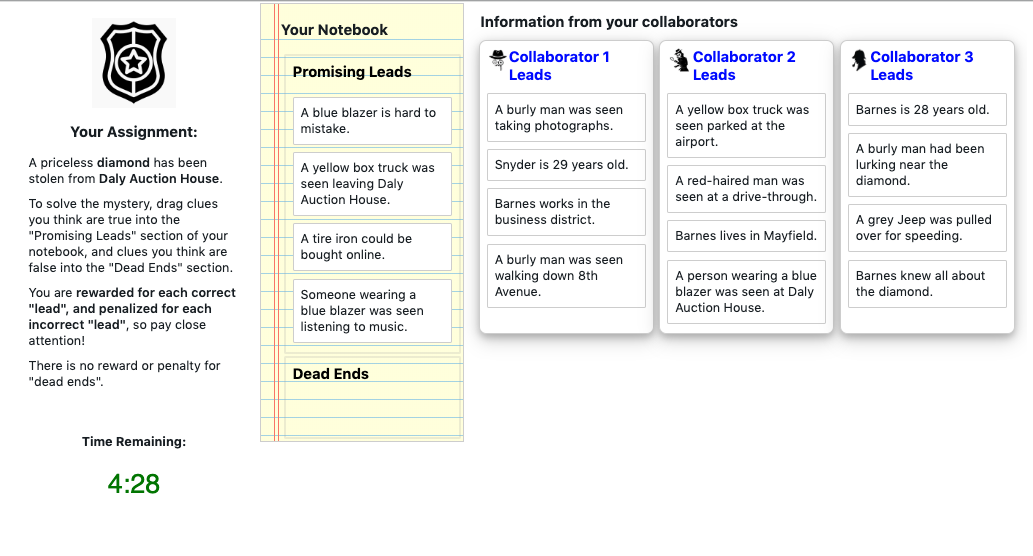


Fig 6b: Game screen – Matching control case

* 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. 7a. 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.

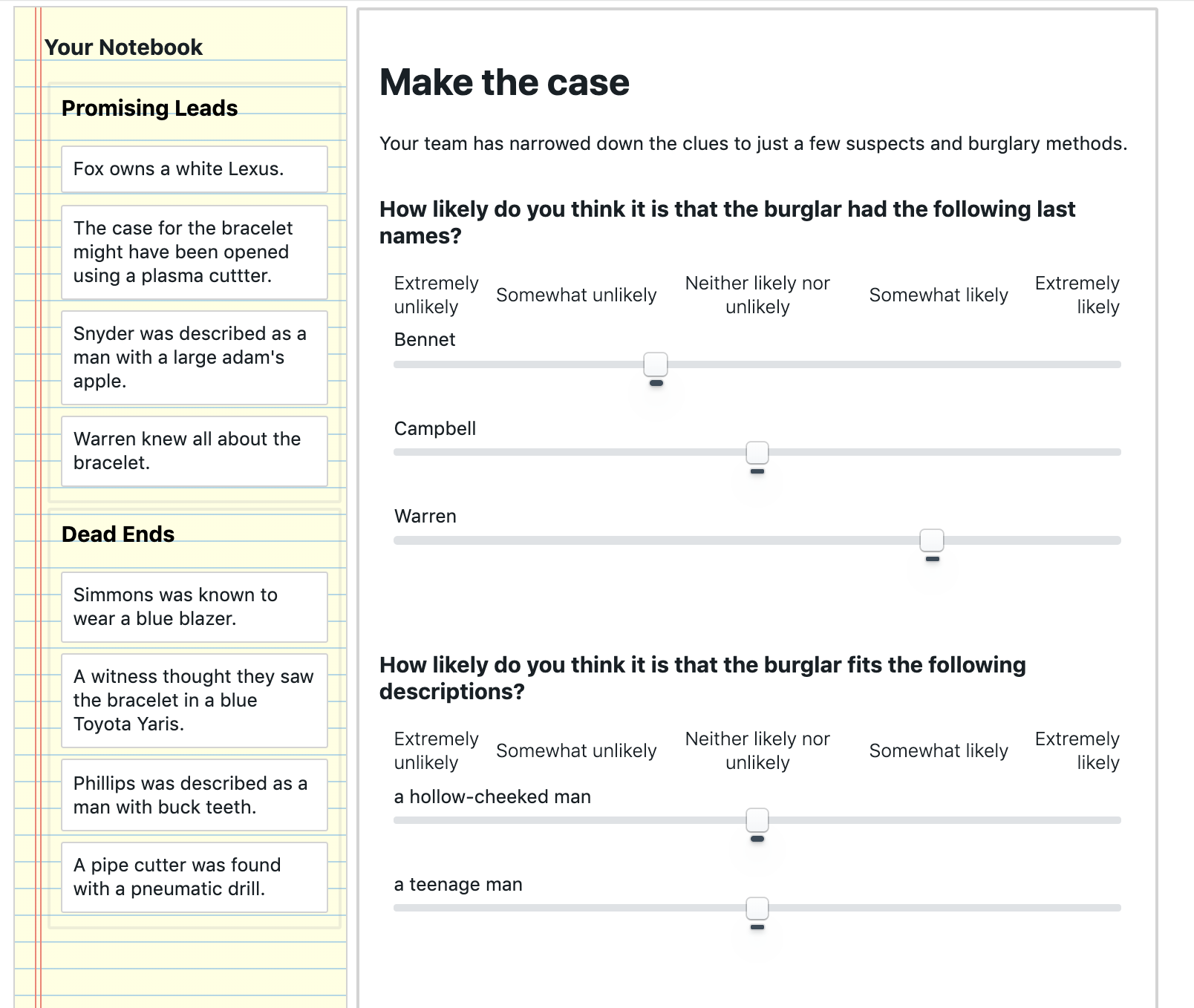


Fig. 7a: Post-game survey screen – Make the case for who committed the burglary

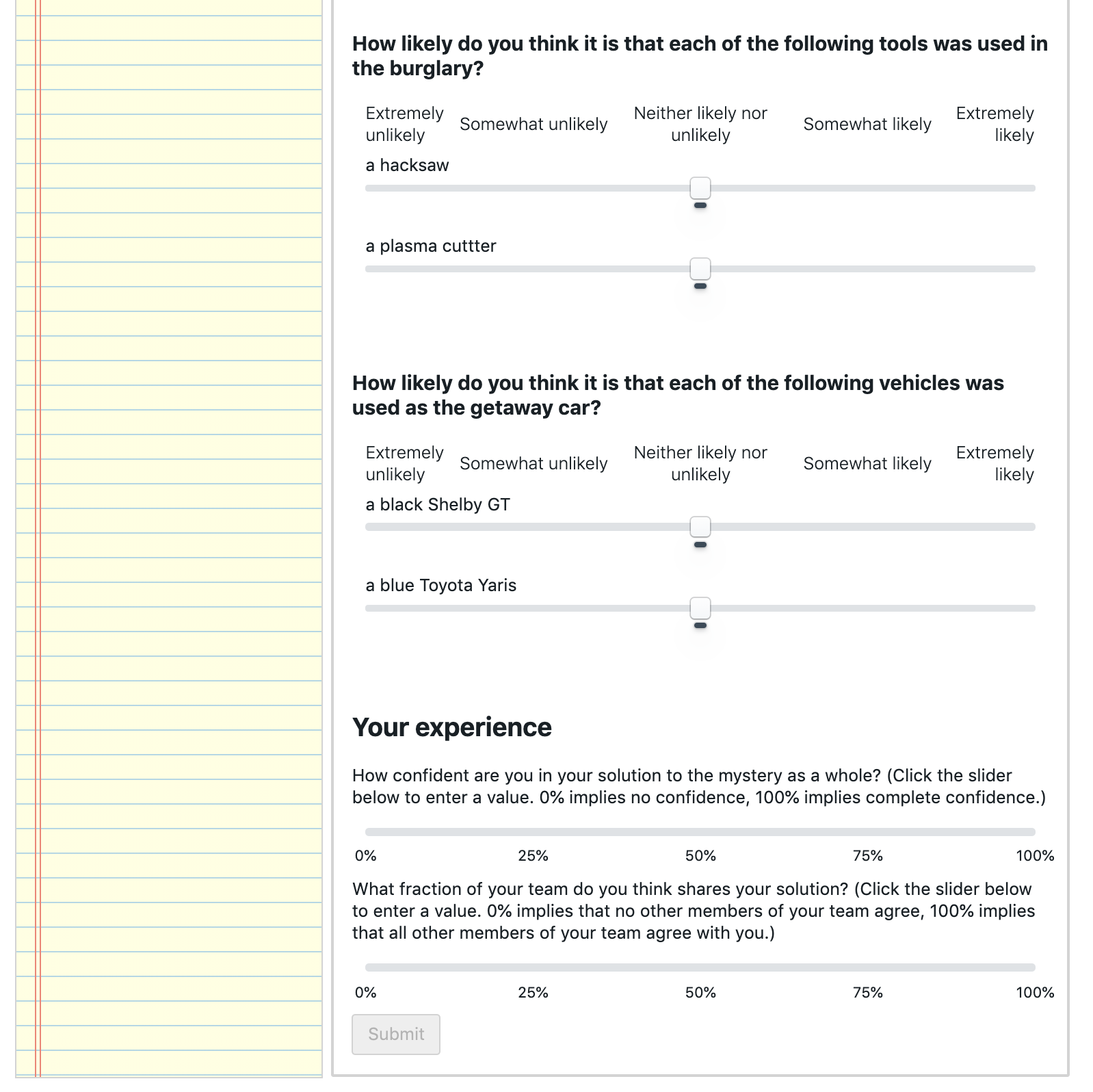


Fig 7b: Post-game survey screen – Assess confidence and consensus

* 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.
  2. 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. 8 shows the number of participants active in different parts of the game at different times for one pilot test.

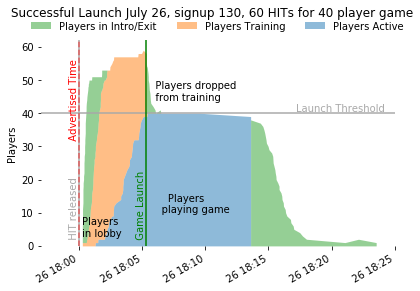


Fig. 8: Active players by stage

* 1. Many players in the pilot tests report (in the open response questions and on online worker forums) that the game “works smoothly”, and is “kinda fun” and “not boring”.

1. Sample size
   1. 30 games will be conducted, each with 80 participants divided into four groups, occupying independent social networks of 20 players. One pair of groups will join a dodecahedral social network, and the other a connected caveman network. Within each network condition, one group will be assigned the treatment condition and the other a matched control condition.
   2. For the manipulation check and H1-H2, the unit of analysis is the individual\*exposure event. Each game yields approximately 400 individual\*exposure events within the sample period, for an expected total of approximately 12,000 individual\*exposure events. These do not include exposures that are due to the initial seeding of the game, as these do not represent choices made by the exposing individual.
   3. For H3-H5 the natural unit of analysis is the team\*game event. A single measure for each of the social networks in the game is calculated, for a sample size of 30 games in each condition.
   4. H6 asks about individual reactions to the game. However, these reactions cannot be assumed to be independent, and so an average value is calculated for each team\*game event, and these are compared, for 30 samples in each condition.
2. Sample size rationale
   1. Each block of four games costs approximately $450 to run (including recruitment and bonus overhead). Budget constraints limit samples to 30 games.

### Variables

1. 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. 9. 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.

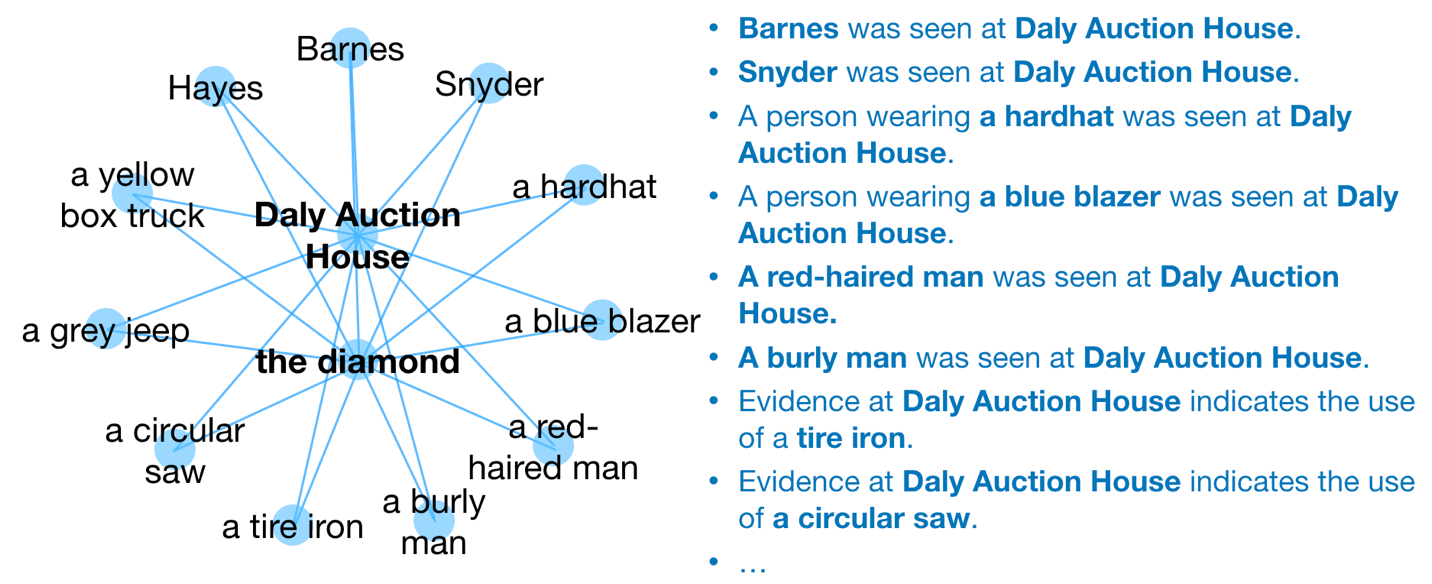


Fig. 9: “Spoke” clues connect rim concepts to hub concepts

* 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. 10.

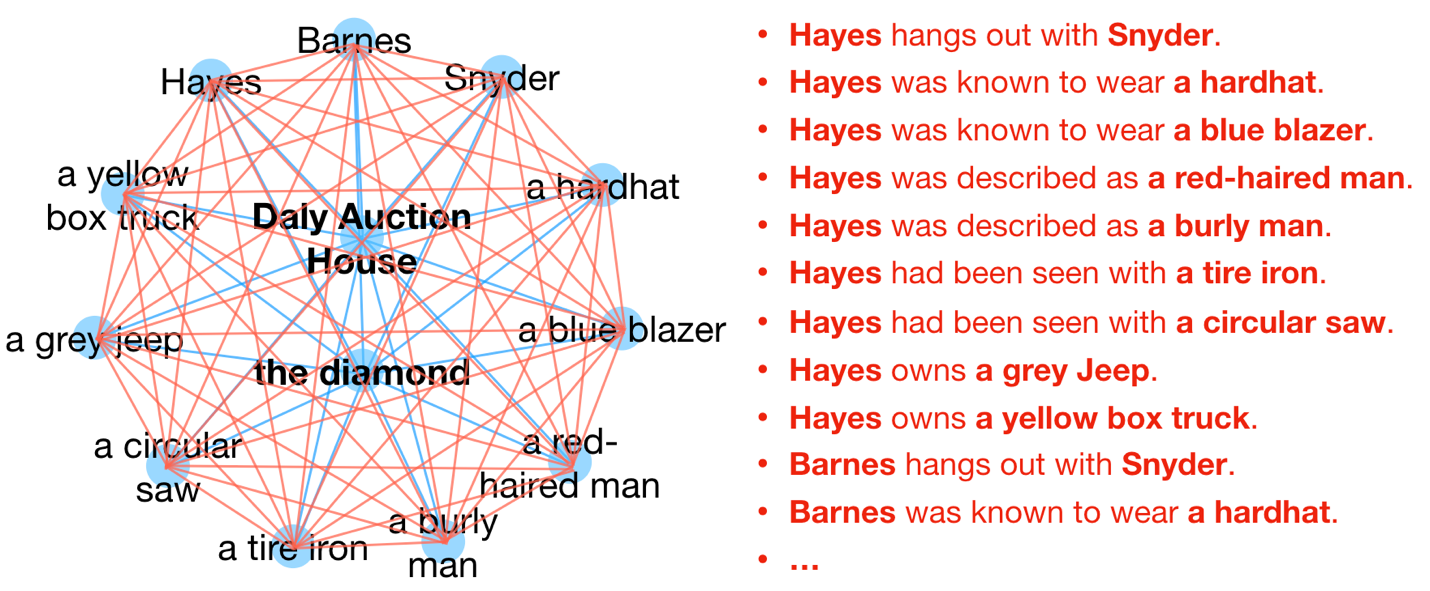


Fig 10. “Cross-Link” clues connect rim concepts to one another

* 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 11). 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.

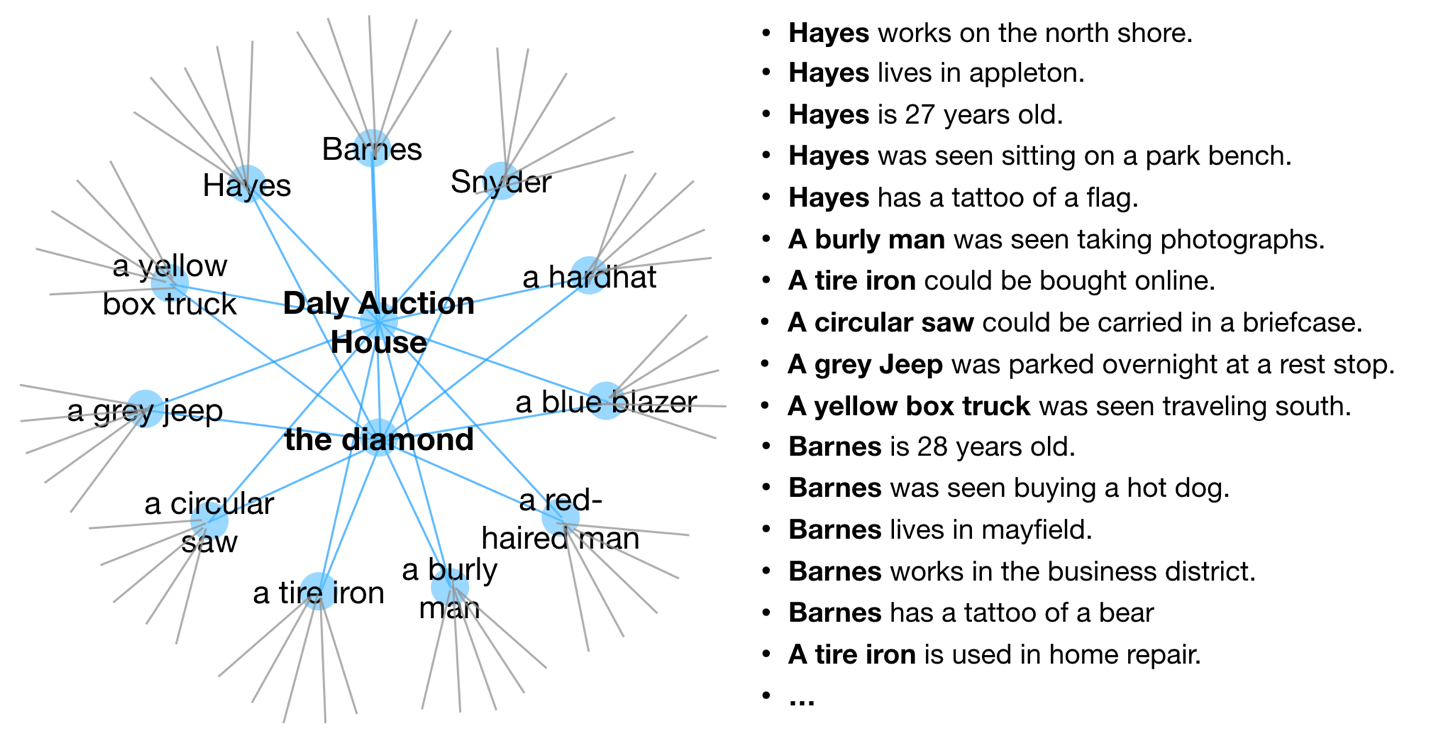


Fig 11: “Spur” clues fill the place of “cross-link” clues without creating links between rim concepts, while still allowing for multiple exposures to rim concepts.

* 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.

1. Measured variables
   1. The initial state of each player’s notebooks is logged. Note: this is the manipulation design.
   2. 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. it’s numerical position in the notebook) and the time at which the drop event occurred.
   3. The final state of all notebooks in the game is logged. Note: the state of the game can be reconstructed for all times in the game from the initial (design) state of the notebooks, and the action log. The final state provides a check that all events are logged properly.
   4. I record each individual’s estimate of how likely each of the 11 “rim” concepts is to be connected to the crime. Participants click on a slider (with no predefined value) to indicate their level of belief from “Extremely Unlikely” to “Extremely Likely”. Slider positions are captured as an integer value between 0 and 100. (Fig. 5a shows the precise wording for these questions.)
   5. Individuals enter a value for their confidence in their solution on a scale from 0 to 100 using a blank slider. (Fig. 7b shows the precise wording for this question.)
   6. Individuals enter a value for the fraction of their team they think shares their solution on a scale from 0 to 100 using a blank slider. (Fig. 7b shows the precise wording for this question.)

### Analysis Plan

Three types of data will be used in analysis: 1) time-series logs of all events in the game, 2) final-state values of each participant’s detective notebooks, and 3) post-game survey responses. These will be used individually and in combination to test the hypotheses above.

1. Statistical models
   1. Time-series data analysis
      1. I will first reconstruct the state of the system at all times during gameplay, and calculate the hazard of adopting a candidate belief conditional on exposure. I will use a time-varying Cox proportional hazard regression to estimate the effect of various factors on adoption. See code appendix for details. In this survival analysis, the “birth” event corresponds to an individual’s first exposure to a clue that they have not classified as a “promising lead” (i.e. a “candidate” clue), occurring at time “tE”. The “death” event corresponds to the individual’s adoption of the clue as a “promising lead”, occurring at time “tA”. The data is right-censored as not all exposures lead to adoptions within the time window.
      2. To ensure an apples-to-apples comparison between the treatment and control case, hazard calculations are conducted for “spoke” clues only, as these are shared between the two conditions. Limiting the analysis to “spoke” clues weakens the statistical power of the analysis, but strengthens its validity as a test of the effect of belief interactions.
      3. 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”. This is designed to be zero for all cases in the control condition.
         2. **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.
         3. **Social exposure** is formalized as the number of individuals exposing the candidate clue to the participant.
         4. The **new clue volume** is formalized as the number of clues that an individual is exposed to, but has not yet categorized. In the game interface these are discernable (via subtle shading) from clues that have been categorized, and so serves as a useful signal of the outstanding ‘workload’ on the participant. We should expect that under either treatment or control condition, a larger volume of new clues should decrease the speed with which any individual clue is adopted.
         5. The **number of current clues** is merely the number of clues the participant has already classified as a “promising lead”. We might expect that when they have fewer “promising leads” (at the start of the game) the participant may be quicker to adopt.
         6. The **prior rejection of a clue** is a binary variable indicating whether the individual has previously categorized the clue as a “dead end”. While we expect this categorization to have a strong inhibitory effect on adoption, clues are reassessed through the course of the game. (This is why I choose not to merely exclude clues that have been declared “dead ends” from the set of considered exposures.)
         7. The **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.
         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).
      4. A (gaussian) random effect is allowed for each participant, to account for the fact that some players are more active than others.
      5. The manipulation check (M0) and Hypothesis 1 (H1) will be assessed using a time-varying cox proportional hazard regression on all of the above regressors except for ‘similarity’. I assume that a participant’s hazard of adoption is independent of other participants conditional upon these regressors, and so I will perform a single regression using all individual\*exposure events from all games.
      6. Hypothesis 2 (H2) will be assessed using a time-varying cox proportional hazard regression the above regressors except for ‘logical support’ and ‘familiarity’. I assume that a participant’s hazard of adoption is independent of other participants conditional upon these regressors, and so I will perform a single regression using all individual\*exposure events from all games.
   2. End-state and survey data analysis
      1. For hypotheses H3a and H4a, I formalize the similarity between pairs of individuals as the Phi coefficient (binary form of the Pearson correlation) between the binary final state adoption vectors of each individual for all “spoke” clues.
      2. For hypotheses H3b and H4b, I formalize the similarity between pairs of individuals as the Pearson correlation between the vector of responses to survey questions described in Fig. 5a.
      3. I assume that pairs of individuals scoring at the 95th percentile similarity can be conservatively categorized as being within the same ideological “camp”, and that pairs of individuals scoring in the 5th percentile similarity can be conservatively categorized as being in different ideological “camps”.
      4. To assess hypothesis H3a+b and H4a+b, I will use a one-tailed paired T-test for the difference between matched treatment and control conditions on the 95th and 5th percentile similarities respectively.
      5. For hypothesis 5, I formalize the extent to which individuals fall along a “left-right” axis as the percent of variation in the population that can be explained by the first principal component in a principal component analysis. I estimate the variance in this component using the Python library Scikit-Learn’s PCA module, which uses singular value decomposition algorithm to identify components. See code appendix for details.
      6. For H5a, the PCA input feature space is the binary presence of the 22 spoke clues within individuals final “promising leads”. For H5b, the PCA input feature space is the 11 survey response questions, each of which take values between 0 and 100.
      7. To assess H5a and H5b, I will use a one-tailed pairwise T-test for the difference in the percent of variation explained by the first principal component in each feature space.
      8. To assess the mediation claims in H3m, H4m, and H5m, I will replicate the hazard rate regression described above for each game, generating 30 values for the effect of ‘coverage’ in each condition. I define a structural equation model with a direct pathway between treatment and each of the outcome variables, and an indirect pathway via the ‘coverage’ mediator. I will also include a fixed effect for each treatment/control game pair such that the results of the mediation analysis are applied to the delta between treatment and control cases, similarly to the paired T-tests above. I implement the SEM using the R package “Lavaan”. See code appendix for details.
      9. To assess Hypothesis 6a+b, I will average the results of the questions given in Fig. 5b across all players within a condition in each block, and use a one-tailed pairwise T-test for the difference between averaged responses to the given questions.
2. Inference criteria
   1. All significance tests will be assessed at p<.05 and all tests will be reported.
   2. I assume that the manipulation check is satisfied if I observe a significant and positive value for the log hazard ratio (LHR) for the influence of logical support interacted with a treatment dummy.
   3. H1a is satisfied (i.e. not rejected) if the LHR for the effect of familiarity on adoption is significant and greater than 0. H1b is satisfied if the LRH for the effect of familiarity interacted with a treatment dummy is significant and less than 0.
   4. H2a is satisfied if the LHR for the effect of coverage on adoption is significant and greater than 0. H2b is satisfied if the LHR of the effect of coverage interacted with the treatment dummy is greater than 0.
   5. H3a is satisfied if the net increase in 95th percentile between-participant phi coefficients compared to a randomized baseline (preserving the number of adoptions of each clue and the number of clues for each participant) is greater under treatment conditions than under control conditions. H3b is satisfied if the 95th percentile value of between-participant correlations in treatment conditions is greater than under the control conditions.
   6. H4a is satisfied if the net decrease in 5th percentile between-participant phi coefficients compared to a randomized baseline (preserving the number of adoptions of each clue and the number of clues for each participant) is greater under treatment conditions than under control conditions. H3b is satisfied if the 5th percentile value of between-participant correlations in treatment conditions is lower than under the control conditions.
   7. H5a is satisfied if the net increase in the percent of variation explained by the first principal component compared to a randomized baseline (preserving the number of adoptions of each clue and the number of clues for each participant) is greater under treatment conditions than under control conditions. H5b is satisfied if the percent of variation explained by the first principal component is greater in treatment conditions than control conditions.
   8. Mediation hypotheses H3m, H4m, and H5m are satisfied if the indirect effect estimated in the SEM is significant in each case.
   9. H6a and H6b are satisfied if the difference between the game-treatment averaged responses to the questions in Fig. 5b under treatment conditions vs matched control conditions is greater than 0 when averaged across all games, and a one-tailed pairwise T test is significant.
   10. H7 is satisfied by the replication of tests for H3-5 on the connected-caveman network structure.
3. Data exclusion
   1. 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 M0, H1 and H2, I will only look at individual\*exposure events that happen after t0, the start of the game.
   2. The software platform will not allow participants to receive their payment code without submitting answers to all questions, and so no data exclusion will be performed other than described by the drop-out policy below.
4. 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 will identify a drop-out as any player failing to submit the post-game survey.
   2. Individuals who drop out of the game will still be included in hazard-rate calculations for H1-2.
   3. 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 H3-7. Aggregate results for the paired comparison condition will be calculated as the average of all same-sized subsets of the comparison condition.

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. J. Houghton. Interdependent Diffusion: The social contagion of interacting beliefs. MIT working paper. (2019)
6. DiMaggio, P., Evans, J. & Bryson, B. dimaggio.pdf. *Am J Sociol* **102**, 690–755 (1996).
7. 1.Becker, J., Porter, E. & Centola, D. The wisdom of partisan crowds. *Proc National Acad Sc*i **116**, 10717–10722 (2019).