Inferring behaviour from partial social information plays little or no role in the cultural transmission of adaptive traits: supplementary information

Mark Atkinson, Kirsten H. Blakey, and Christine A. Caldwell

Psychology, University of Stirling, Stirling, United Kingdom

1 Experiment 1: exposure to minimal partial information (as preregistered)

Here we report the version of Experiment 1 as preregistered with the Open Science Framework (`https://osf.io/9heqv`). It primarily differs from that reported in Section 2 in having two Observer conditions. We investigated whether participants could make and use valid inferences about others' past experience of a problem from exposure to a single attempt by comparing performance in these two Observer conditions, as opposed to comparing Observer performance to the expected performance if they were making selections randomly, as we report in Section 2.

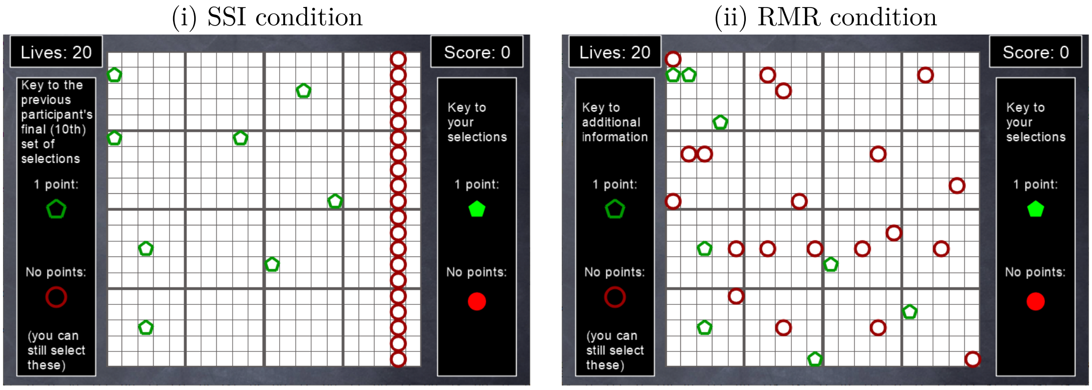
1.1 Methods

The methodology broadly followed that described in Section 2.1. However, participants were recruited in groups of three, with the first participant of each group assigned to the Demonstrator role, and the other two assigned Observer roles. One Observer participant was assigned to a Structured Social Information (SSI) condition, and this is the Observer condition we refer to throughout Section 2. The other Observer participant was assigned to a Random Matched Reveal (RMR) condition.

As described in Section 1.1, participants in the SSI condition were told that another participant had taken part in the task, and that their task was the same, i.e. that they had had ten expeditions for each of the three problems, and were aiming to get as high a score as possible. SSI participants were then given full information about the selections that were made by the Demonstrator on the Demonstrator's final expedition for each problem.

In the RMR condition participants also completed the same problems as their Demonstrator, i.e. the location of scoring tiles were the same (as was the case in the SSI condition). However, rather than being given the actual selections made on the Demonstrator's final expedition, they were instead given information about a randomly selected set of tiles, which were nonetheless matched in terms of the number of hits and misses. For example, if the Demonstrator selected three scoring tiles and 20 non-scoring tiles, then three scoring tiles and 20 non-scoring tiles were randomly selected and revealed to the Observer in the RMR condition. The information received in the RMR condition was therefore comparable to that in the SSI condition in that it informed the participant about the same number of hits and misses, but it did not include any structural information about the Demonstrator's behaviour, and so there was no possibility that an RMR participant could infer anything about the Demonstrator's experience of the task from the information observed. RMR participants were informed that the tiles had been randomly selected to be revealed as clues.

See SI Fig. 1 for an illustration of the differences between the SSI and RMR conditions. Aside from their location, the clues were displayed in otherwise exactly the same ways across both conditions. The Observers in both conditions (SSI and RMR) could make use of this information however they wished.



SI Fig. 1: **Comparison of the SSI and RMR Observer conditions.** Panel (i) shows an example of how the task would appear to an SSI participant at the start of an expedition (this is the same as panel (iii) in Fig. 1). Panel (ii) shows of an example of how the task would appear at the start of an expedition to the RMR participant in the same group: there are the same number of hits and misses revealed as for the SSI, and the underlying reward structure is the same, but otherwise the hits and misses have been selected randomly. In the SSI condition, a participant could potentially detect structural regularities in the Demonstrator's behaviour, and so make valid inferences about the Demonstrator's behaviour over and above the information they received directly. In the RMR condition, this would not have been possible.

In our analyses, we considered two key variables: for all participants (both Demonstrators and Observers in both conditions), we assessed score; for the Observers, we compared their selection of redundant selections.

1.2 Participants

In addition to the participants reported in Section 2.1.1, we recruited an additional 20 adult participants for the RMR Observer condition.

Overall, 60 adult participants (mean age 21.0, range 18–51; 35 female and 25 male) were recruited at the University of Stirling and took part in exchange for either course credit (54 participants) or £3 in cash (6 participants). Data from 1 additional participant was collected, but was excluded due to an experimenter error leading to not all relevant information being saved. Another participant was recruited in their place.

1.3 Results

1.3.1 Selection of scoring tiles

The mean total scores (for all three problems) were 133.6 (sd = 39.33) for the Demonstrators and 273.45 (sd = 50.03) for the Observers: 278.2 (sd = 48.80) for the Observers in the SSI condition and 268.7 (sd = 52.05) for those in the RMR condition.

We constructed a logit-linked mixed model with score as dependent variable, and participant role, expedition number (centred), and their interaction as fixed effects. Participant role/Condition was Helmert contrast coded to allow two contrast types to be investigated: SSI versus RMR, followed by

Observer versus Demonstrator. Participant identity nested within trio membership, and problem number were included as random intercepts.

Demonstrators scored lower than the Observers (b = -0.270, SE = 0.013, z = -20.157, $p < .001$). Score increased with expedition number (b = 0.082, SE = 0.004, z = 21.576, $p < .001$). There was no difference between the two Observer conditions (b = -0.018, SE = 0.021, z = -0.862, p = .388) and no interaction effect between condition and expedition number for the two Observer conditions (b = -0.005, SE = 0.004, z = -1.180, p = .238). There was an interaction effect involving expedition number when comparing the Demonstrator and Observer participant roles (b = 0.039, SE = 0.003, z = 12.899, $p < .001$), however, with the effect of expedition number being stronger in the Demonstrator participant role.

1.3.2 Selection of redundant selections

In the SSI condition, the mean number of redundant selections was 236.1 (sd = 26.17). In the RMR condition, it was 266.8 (sd = 27.12). These raw scores do not account for there being a difference in the number of *potential* redundant selections in the SSI and RMR selections, however. In the SSI condition, the information they received was the exact set of selections made by the Demonstrator in their final search of the grid for that problem. Therefore the potential redundant selections which the SSI Observer could make were all the misses that the Demonstrator made in the first nine expeditions for that problem which were not also made in the tenth expedition. In the RMR condition, though they received the same number of hits and misses as the Demonstrator made in their final search of the grid for that problem, the hits and misses observed by the RMR Observer were randomly selected. Therefore, compared to the SSI Observer condition, there would very likely be a greater number of potential redundant selections which could have been made. Some potential selections which would have been available to the SSI participant may have been unavailable to the RMR participant by the random allocation of misses in the RMR condition (e.g. consider the very top left tile in SI Fig. 1 for each panel: if the Demonstrator selected this tile within their first nine expeditions, it would be a potential redundant selection for the SSI participant, but not for the RMR participant). However, there would very likely be a greater number of potential redundant selections from the last Demonstrator expedition which would be available to the RMR (e.g. the misses from the penultimate column in panel (i) in SI Fig. 1 which are not given in panel (ii)), but which would not be redundant selections if made by the SSI, as they would have been given to the SSI participant as partial information of Demonstrator performance. To put it another way, while the SSI participant received information about 20 Demonstrator misses, therefore making theoretically up to 180 potential redundant selections available to them, the RMR participant received information of 20 misses, each of which may or may not have been selected by the Demonstrator in any round: theoretically the RMR would have had up to 200 potential redundant selections available to them. The consequence of this is that even if the participants were making their selections entirely at random, we would expect there to be a greater number of redundant selections to be made by an RMR participant compared to an SSI. We therefore perform two analyses to assess redundant selections. In the first, we consider the number of redundant selections, to assess the effect of expedition number. In the second, we consider redundant selections as a proportion of potential redundant selections.

In our first analysis, we constructed a logit-linked model with redundant selection as dependent variable, and condition, expedition number (centred), and their interaction as fixed effects. Participant identity and problem number were included as random intercepts.

Redundant selections decreased with expedition number (b = -0.021, SE = 0.004, z = -5.188, $p < .001$), but there was no interaction between participant role and expedition number (b = -0.003, SE = 0.004, z = -0.787, p = .431). There was an effect of participant role (b = -0.094, SE = 0.024, z = -4.006, $p < .001$), though as discussed above, under this analysis we would have expected this even if all participants were making their selections at random.

In our second analysis, we consider the set of unique redundant selections as a proportion of the redundant selections which could have been selected in principle.

Participants in the SSI condition on average (mean of the mean of each individual participant's three problems) selected 45.1% (sd = 7.1%) of the redundant selections which could have been selected in principle at least once. For participants in the RMR condition, this was 44.6% (sd = 6.3%).

We constructed a mixed model with redundant selections selected at least once as a proportion of the redundant selections which could have been selected in principle as dependent variable and with condition (sum coded) as fixed effect. We included a random intercept for participant identity.

There was no effect of condition (b = 0.002, SE = 0.011, t(119) = 0.207, p = .836).

1.4 Discussion

As predicted, Observers scored higher than the Demonstrators, indicating that Observers were making use of the information available to them when making their selections (i.e. by selecting the tiles they were informed were hits and/or avoiding the tiles they were informed were misses). Both Demonstrators and Observers also scored higher in later expeditions for each problem, indicating that they were using their own past experience of the problem when making their selections (i.e. by remembering where they searched before and remembering which of their previous selections were hits and/or which of their previous selections were misses).

Contra our predictions, there was no evidence that participants in the SSI condition achieved different scores to participants in the RMR condition, nor was there any evidence of any different effect of expedition number for each of these conditions. There was also no evidence that, once the potential number of redundant selections was accounted for, SSI participants avoided redundant selections to a greater extent than RMR participants, nor was there any evidence of a difference in how expedition number affected redundant selections between conditions. There is therefore no evidence here that the participants who received information in which they could potentially have detected structural regularities and so infer Demonstrator behaviour did so.

As discussed in Section 2, there is also no evidence that the SSI Observer participants avoided redundant selections more than they would have been expected to do by chance. Please see Section 2.3 for further discussion.

2 Experiment 1 exploratory analyses

2.1 Demonstrator reselection of unrewarded tiles

Mean proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition was 24% (sd = 8.4%) in Problem 1, 22% (sd = 7.8%) in Problem 2, and 22% (sd = 8.9%) in Problem 3.

To investigate the effect of problem number on these reselections of unrewarded tiles, we constructed a linear model with proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition as dependent variable. Problem number (centred) was the only fixed effect and participant identity was included as random intercept.

There was no effect of problem number (b = -0.008, SE = 0.011, t(59) = -0.73, p = .468).

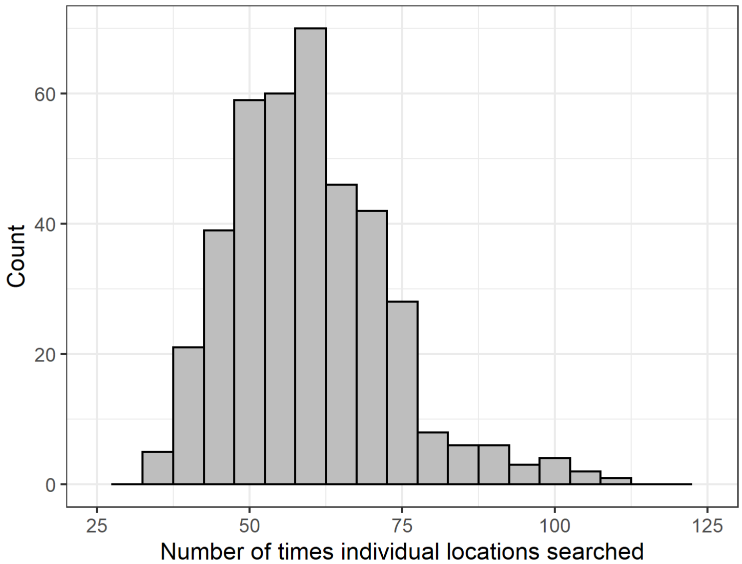
2.2 Effect of problem number on redundant tiles

We constructed an additional linear mixed model to also investigate the effect of problem number on the number of redundant selections made by the Observers. The fixed effects were the actual number of redundant selections compared to the expected values (again treatment coded with the expected values as the baseline), problem number (centred), and their interaction. Pair membership was included as random intercept.

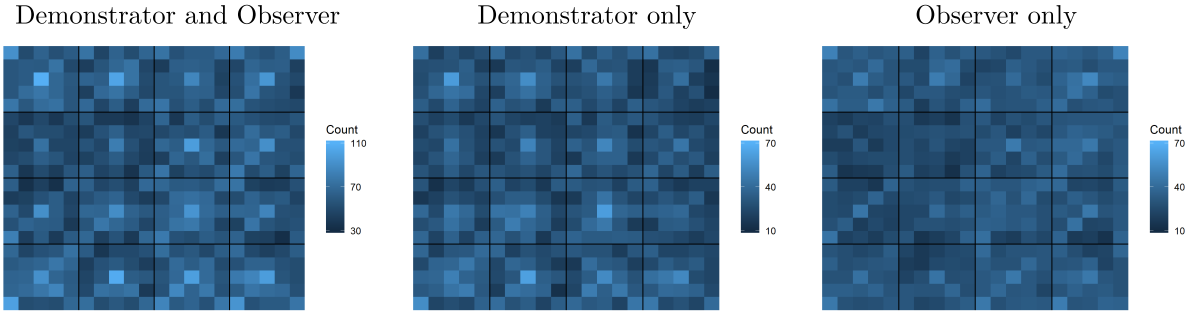
Replicating the result in Section 2.2.2, the actual number of redundant selections made at least once by the Observers was more than would have been expected if the number of unique Observer selections were made randomly (b = 20.001, SE = 2.588, t(117) = 7.728, $p < .001$). There was no effect of problem number (b = 2.348, SE = 2.241, t(117) = 1.047, p = .297), or the interaction between redundant selections and problem number (b = -2.848, SE = 3.170, t(117) = -0.898, p = .371).

3 Experiment 1: distribution of searches by location in grid

To illustrate how (non-)uniformly grid locations were searched in Experiment 1, we pooled all the unrewarded selections made by all the participants, counting how often tiles in individual locations were selected. Counts of the number of times individual locations were searched is illustrated in SI Fig. 2 with heat maps for the individual locations in SI Fig. 3.



SI Fig. 2: **Histogram for the number of times individual locations were searched in Experiment 1.** The minimum number of times a location was searched was 33; the maximum was 109.



SI Fig. 3: **Heat maps for the number of times individual locations containing unrewarded tiles were searched in Experiment 1.** Lighter colouring indicates a location was searched more times, e.g. the very bottom left location was searched more often than the immediately surrounding locations.

4 Experiment 2 exploratory analyses

4.1 Demonstrator reselection of unrewarded tiles

Mean proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition was 15% (sd = 6.7%) in Problem 1, 14% (sd = 8.6%) in Problem 2, and 15% (sd = 10.6%) in Problem 3.

To investigate the effect of problem number on these reselections of unrewarded tiles, we constructed a linear model with proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition as dependent variable. Problem number (centred) was the only fixed effect and participant identity was included as random intercept.

There was no effect of problem number ($|b| < 0.001$, SE = 0.010, $t(59) = -0.042$, $p = .967$).

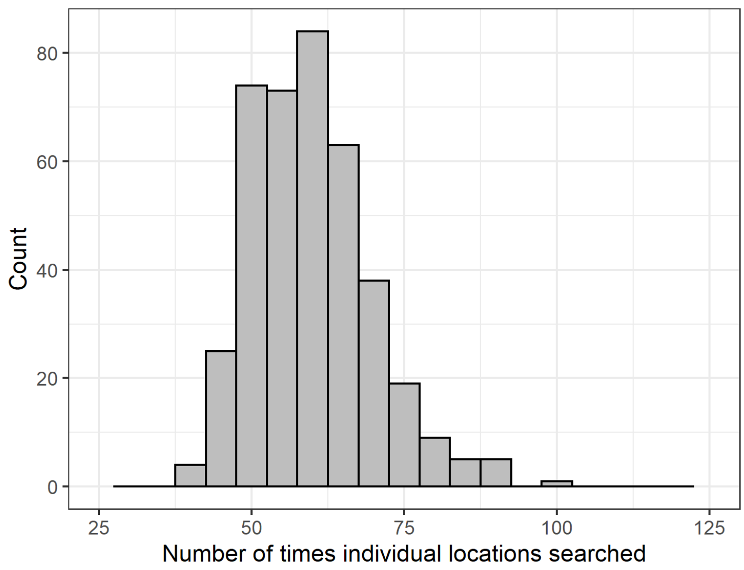
4.2 Effect of problem number on redundant tiles

We constructed an additional linear mixed model to also investigate the effect of problem number on the number of redundant selections made by the Observers. The fixed effects were the actual number of redundant selections compared to the expected values (again treatment coded with the expected values as the baseline), problem number (centred), and their interaction. Pair membership was included as random intercept.

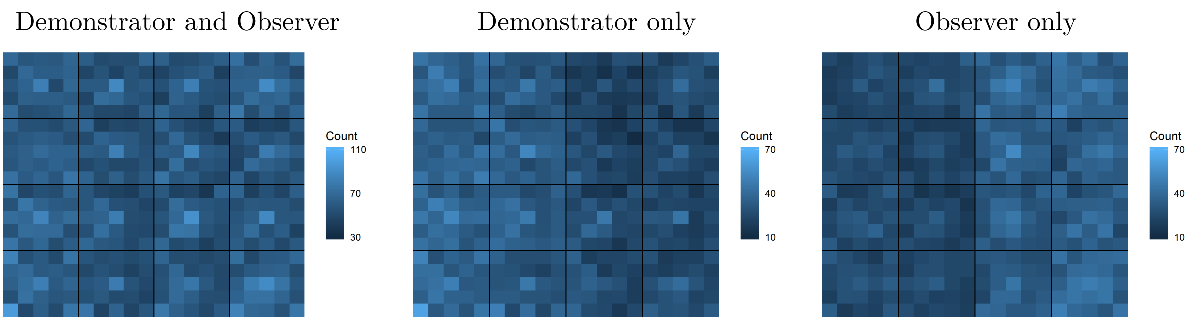
Replicating the result in Section 3.2.2, the actual number of redundant selections made at least once by the Observers was less than would have been expected if the number of unique Observer selections were made randomly ($b = -7.029$, SE = 3.008, $t(117) = -2.336$, $p = .021$). There was no effect of problem number ($b = 3.905$, SE = 2.605, $t(117) = 1.499$, $p = .137$), or the interaction between redundant selections and problem number ($b = -0.130$, SE = 3.685, $t(117) = -0.035$, $p = .972$).

5 Experiment 2: distribution of searches by location in grid

As for Experiment 1, to illustrate how (non-)uniformly grid locations were searched in Experiment 2, we pooled all the unrewarded selections made by all the participants, counting how often tiles in individual locations were selected. Counts of the number of times individual locations were searched is illustrated in SI Fig. 4 with heat maps for the individual locations in SI Fig. 5.



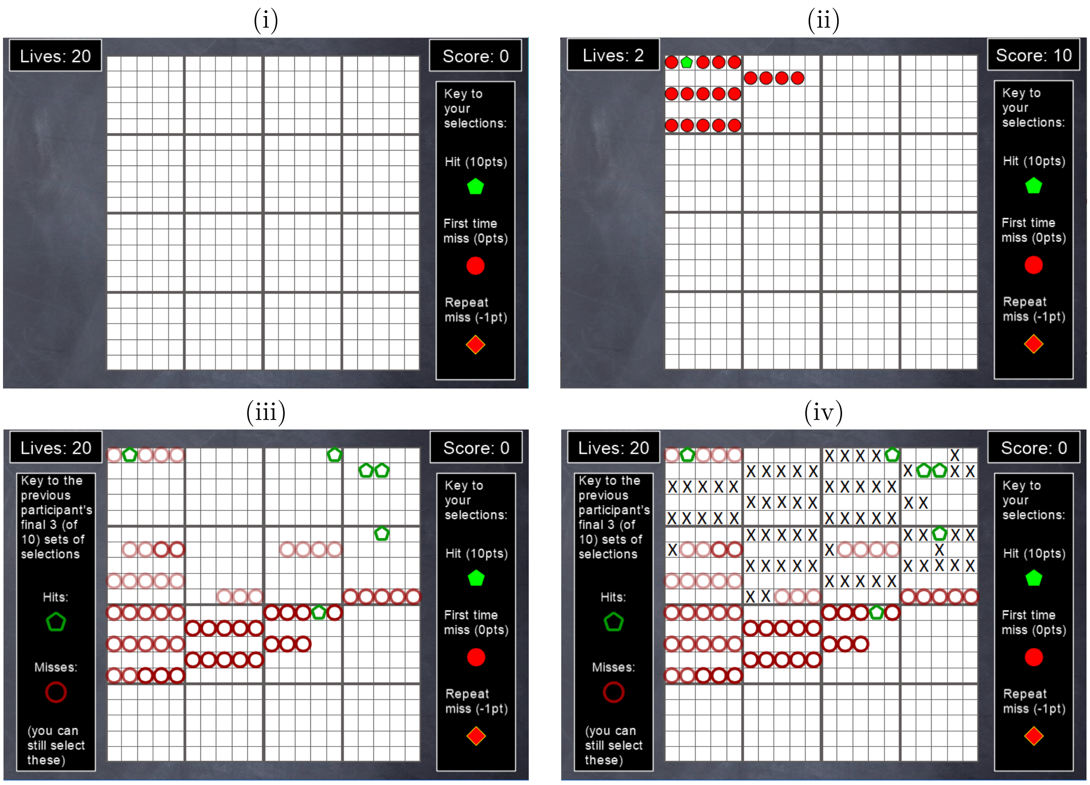
SI Fig. 4: **Histogram for the number of times individual locations were searched in Experiment 2.** The minimum number of times a tile was selected was 38; the maximum was 98.



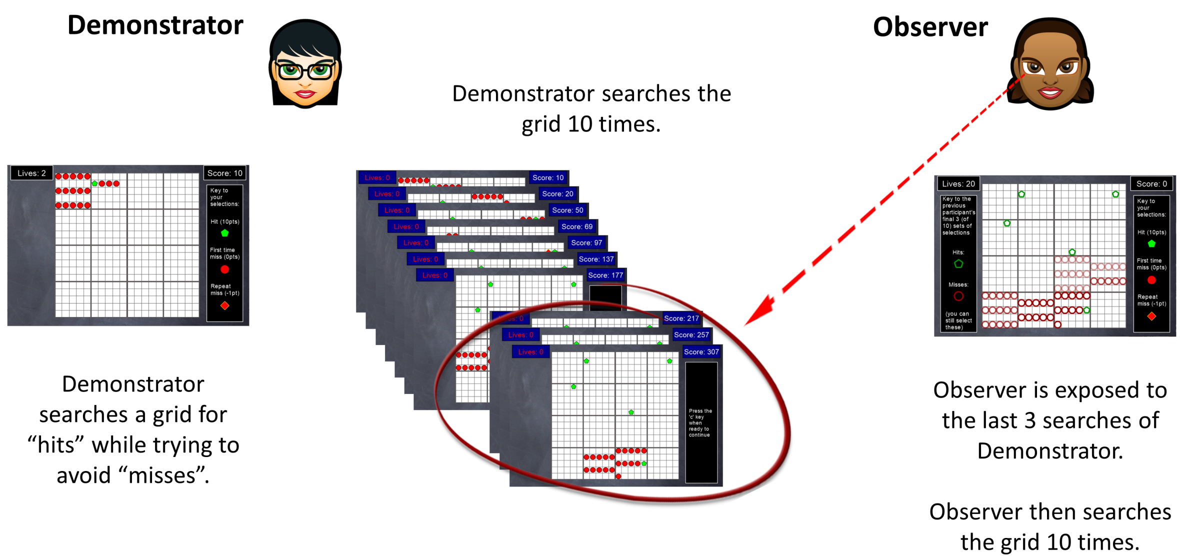
SI Fig. 5: **Heat maps for the number of times individual locations containing unrewarded tiles were searched in Experiment 2.** Lighter colouring indicates a location was searched more times, e.g. the very bottom left locations was searched more often than the immediately surrounding locations.

6 Experiment 2 additional figures

See SI Fig. 6 for example screenshots from Experiment 2. As for the illustrative example of Experiment 1, note that such a transparently systematic search of the grid was not typical for the majority of participants. See SI Fig. 7 for an illustration of the relationship between the Demonstrator and Observer for a given pair for participants.



SI Fig. 6: **Example screenshots from hypothetical participants in Experiment 2.** Panel (i): Demonstrator at the start of an expedition. Panel (ii): Demonstrator partway through an expedition. Panel (iii): Observer at the start of an expedition. More transparent outlines indicate that the information comes from earlier Demonstrator expeditions. Panel (iv): Observer at the start of an expedition (as panel (iii)), with additional crosses illustrating potential redundant selections. Above-chance avoidance of these crosses (which were not made visible to the participant at any point) would suggest that the Observer was able to infer the search behaviour of the Demonstrator (at least to some extent) and avoid misses that the Demonstrator had already selected, even though the Observer did not have direct access to this information. Note that such transparently systematic Demonstrator behaviour was not typical for the majority of our participants.



SI Fig. 7: **Illustration of the relationship between Demonstrator and Observer in Experiment 2.**

7 Comparison of scoring tile selections in Experiments 1 and 2

We combined our datasets from Experiments 1 and 2, and constructed a logit-linked linear mixed model with scoring tile selection as binary dependent variable. The fixed effects were participant role (Demonstrator v Observer, treatment coded with Demonstrator as the baseline), expedition number (centred), experiment number (sum coded: Experiment 1 = -1; Experiment 2 = 1), and all interactions. Pair membership was included as a random intercept effect.

Consistent with the individual analyses for Experiments 1 and 2, selection of scoring tiles increased with expedition number ($b = 0.158$, $SE = 0.006$, $z = 27.464$, $p < .001$), and there was an effect of participant role ($b = 0.886$, $SE = 0.020$, $z = 44.449$, $p < .001$), indicating that there were more scoring tiles selected by Observers than by Demonstrators. An interaction between expedition number and participant role indicated that the effect of expedition number was more pronounced in the Demonstrators ($b = -0.104$, $SE = 0.007$, $z = -14.918$, $p < .001$).

There was no main effect of experiment number ($b = -0.037$, $SE = 0.045$, $z = -0.833$, $p = .405$), although there was an interaction between experiment number and participant role ($b = 0.075$, $SE = 0.020$, $z = 3.745$, $p < .001$), indicating that the effect of participant role (Observers selecting more scoring tiles than Demonstrators) was more pronounced in Experiment 2. There were no other significant interaction terms ($|b| \leq 0.007$, $SE \geq 0.006$, $|z| \leq 0.945$, $p \geq .345$).

8 Comparison of Demonstrator reselection of unrewarded tiles in Experiments 1 and 2

We combined our datasets from Experiments 1 and 2, and constructed a linear mixed model with proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition as dependent variable. The fixed effects were experiment number (sum coded: Experiment 1 = -1; Experiment 2 =1), problem number, and their interaction. Participant identity was included as a random intercept effect.

There was less reselection of unrewarded tiles in Experiment 2 ($b = -0.044$, $SE = 0.013$, $t(117) = -3.450$, $p < .001$). Consistent with the individual experiment analyses involving problem number above, there was no effect of problem number ($b = -0.004$, $SE = 0.008$, $t(117) = -0.577$, $p = .565$), and there was no effect of the interaction between experiment number and problem number ($b = 0.004$, $SE = 0.008$, $t(117) = 0.522$, $p = .603$).

9 Experiment 3 exploratory analyses

9.1 Demonstrator reselection of unrewarded tiles

Mean proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition was 20% (sd = 7.3%) in Problem 1, 13% (sd = 5.8%) in Problem 2, and 15% (sd = 7.8%) in Problem 3.

To investigate the effect of problem number on these reselections of unrewarded tiles, we constructed a linear model with proportion of Demonstrator unrewarded selections (after the first expedition) which were tiles that they had already selected and found to be non-scoring on at least one previous expedition as dependent variable. Problem number (centred) was the only fixed effect and participant identity was included as random intercept.

Demonstrator reselection of unrewarded tiles reduced with problem number ($b = -0.027$, $SE = 0.008$, $t(59) = -3.158$, $p = .003$).

9.2 Effect of problem number on redundant tiles

We constructed an additional linear mixed model to also investigate the effect of problem number on the number of redundant selections made by the Observers. The fixed effects were the actual number of redundant selections compared to the expected values (again treatment coded with the expected values as the baseline), problem number (centred), and their interaction. Pair membership was included as random intercept.

Replicating the result in Section 4.2.2, the actual number of redundant selections made at least once by the Observers was more than would have been expected if the number of unique Observer selections were made randomly ($b = 13.497$, $SE = 2.820$, $t(117) = 4.785$, $p < .001$). There was also an effect of problem number ($b = 5.094$, $SE = 2.443$, $t(117) = 2.086$, $p = .039$), but no interaction between redundant selections and problem number ($b = 4.031$, $SE = 3.454$, $t(117) = 1.167$, $p = .246$).

10 Experiment 4 exploratory analysis: effect of problem number on mutually redundant selections

We constructed an additional linear model to also investigate the effect of problem number on the number of mutually redundant selections. The fixed effects were the actual number of mutually redundant selections compared to the expected values (again treatment coded with the expected values as the baseline), problem number (centred), and their interaction.

Replicating the result in Section 5.2.2, There was no evidence of a difference between the actual number of mutually redundant selections and that expected if the number of unique unrewarded selections were made randomly ($b = -0.578$, $SE = 0.586$, $t(116) = -0.987$, $p = .326$). There was no effect of problem number ($b = 0.448$, $SE = 0.507$, $t(116) = 0.883$, $p = .379$), or the interaction between mutually redundant selections and problem number ($b = -1.123$, $SE = 0.717$, $t(116) = -1.566$, $p = .120$).