# Introduction

* Definition of uncertainty and the difference between this and risk
* Definition of heuristics
* Heuristics as a useful tool under uncertainty
* Previous research on heuristics under uncertainty in binary choice paradigms
* Limitations of choice-based classification
* Hook from our study: how prevalent these different strategies are in naïve observers and how can we describe their sampling patterns.

# Methods

## Participants

We collected a total of 81 participants from the student population of the University of Massachusetts, Amherst. These subjects participated in two separate experiments (E1 and E2). From them 40 participated in E1, 34 were females (mean age = 20.71 SD = 1.82) and 6 were males (mean age = 20.83, SD = 1.72). For E2, we had a total of 41 participants, from which 32 were females (mean age = 19.84 SD = 1.27) and 9 were males (mean age = 19.78, SD = 1.20). We compensated all of them with a flat rate of class credit or $12, as well as a monetary bonus of $3. All participants provided informed consent in line with the Institutional Review Board from the University of Massachusetts, Amherst.

## Materials

After signing the consent form, participants performed the matrix reasoning subscale from WASI (Wechsler, 1999), to collect information on fluent intelligence, in order to assess if it had any influence in strategy selection or their performance. After this, we presented the participants with a multicue forced-choice probabilistic paradigm, with a mostly non-compensatory environment. This was based on the work from Oh et al. (2016). The task consists on repeatedly exposing the participants to different non verbal compound cues with an unknown distribution of probabilistic weights. The validities from each piece of information are learned through trial an error. We selected this design since it allows to emulate the uncertainty underlying the outcomes of their choices, a large number of unique information pieces to avoid memorization, and to minimize the influence of semantic expertise in any field on the performance of our subjects.

We designed the task in Matlab, using the Psychophysics Toolbox extension (Brainard (1997)). We monitored eye-movements with a RED250mobile eye tracker from SMI, using the SMITE toolbox Niehorster and Nyström (2020) to link it to our task.

### Stimuli.

We framed our experiment as a race between two different pilots, which had four different pieces of equipment (domains) with 2 possible brands on each (binary states). The compound cues we used were two 2x2 grids, i.e., squares with 4 cells (each cell corresponding to a domain), presented at the right and left side of the screen (figure 1). Each domain contained a different item, presenting one of two possible states. These states, unknown to the participants, had complementary distribution of probabilistic weights (figure 2). These weights are the probabilities of winning from each state. The distance between a given pair of weights (e.g. 0.95-0.05 = 0.9) represents an objective measure of the validity (importance) of each domain towards predicting the winning option.

We used all possible combinations of cue states which gave us a total of 16 unique grids (4^2 = 16). Also, we used all non repeating pairs of cues, giving us a total of 120 unique stimuli ([16^2/2] - [16/2]). Mirroring these stimuli (i.e. switching the position from left to right) gives us a full set of 240 pairs of cues.

The position of equipment and brands was the same for all the participants. However, there are 24 different combination of the underlying weights, which we allocated semi-randomly across subjects.

## Procedure

In both experiments, the task consisted on three phases: a training phase (TP), with a full set of 240 trials divided in 4 blocks of 60 trials, and two testing phases, each with a set of 120 unique trials, divided in 2 blocks. During the first testing phase (T1) we kept the same contingencies from TP (Figure 2). For the second testing phase (T2) we introduced changes in the weights across domains in both experiments. For E1 we assigned to all the different domains a distribution of 0.80/0.20, preserving the winning cue states (Figure 3). On the other hand, for E2 we inverted the distribution (Figure 4), making the most informative cue (0.95/0.05) have the contingencies from the least informative cue (0.5001/0.4999)

We included T2 as a way to test whether participants used decision heuristics that prioritize certain pieces of information over others or if they used compensatory strategies that considered all the domains. This change in contingencies would negatively affect the performance of subjects using non compensatory strategies, since they would disregard pieces of information that are relevant under the new environment. Subjects using compensatory strategies should not show a change in their performance, since the new scenario will still favor integration of more pieces of information.

After completing the WASI subscale, we explained the task to the subjects. We told them that they would have to chose between two pilots that had different equipment which one they think would win the race. We also mentioned that, at the beginning of the task they had no information about which brand or equipment gave more chances of winning to a pilot, but that they would learn with trial an error. We instructed them to pay attention of which combinations of equipment made them win more often. Finally, we explained them that their goal was to obtain as many points as possible, and how they could earn them during the task.

Each trial they won got them 1 point, while fast responses in winning trials gave them 2 points. Any trial were they chose the losing pilot, where they failed to respond within 15 seconds or where their response was too slow yielded no points. The outcome of each trial was determined by getting the probability of winning from the grid selected by the subject which was determined using the following equation:

Then, we compared this number with a random draw from a uniform distribution from 0 to 1. If the value from the grid selected was larger than this random number, they could win the trial, if their RT followed the conditions described above.

We included the speed threshold to emulate the time pressure faced commonly in decision making scenarios, however, to avoid punishing unfairly subjects which had slower RT, we determined the threshold as +1 SD from their mean RT for the slow responses and -1 SD for fast responses. Finally, we determined a 400 points threshold for receiving the monetary bonus at the end of the task. This threshold was determined estimating 70% winning trials with 30% of them being fast responses.

The participants chose between the two grids, using the the left or right arrow in a keyboard. We provided feedback showing the points earned in each trial and a summary of the points acquired during each block, with a total sum showed at the end of the task.

## Data analysis

We determined subjects’ performance based on the number of accurate choices they made. We defined a correct choice as a trial where the participant selected the grid with higher chances of winning, regardless of whether they won or not the trial. For the eye-tracking data, we identified fixations using the I2CM algorithm from Hessels, Niehorster, Kemner, and Hooge (2017). We determined the location of the fixations by defining eight squared AOIs of 200x200 pixels centered in each domain from both cues.

### Model classification.

We used the participants’ responses in each of the different cue configurations as input to classify them across different decision models. We used only the trials during T1, given that at this points, we assumed that participants should have a more crystallized strategy to solve the problem. We followed a similar approach as Oh et al. (2016), creating a total of 15 models which account for all the combinations of possible information usage (Figure 5, 1-15). Additionally, we accounted for two additional strategies widely referred in the literature as feasible in decision paradigms like this: Tallying (equal weighting of information) and Take-The-Best (TTB) which is a type of lexicographic search. The tallying strategy assumes that all the domains would be of equal informational value for the decision maker, and integrates all the pieces of information as a single predictor, i.e., it sums all the positive attributes and chooses the option with a higher number of these features. On the other hand, TTB ranks the domains according to how predictive they are of the outcome, and searches serially starting from the most important feature, stopping as soon as there is a difference between options, choosing the one with the winning state of the attribute. Additionally, we created models that accounted for an incomplete use of these strategies, i.e., where only part of the domains were consulted and included in the decision process (Figure 5, 16-21).

All these were logistic models, with different number of predictors (one for each domain used) or with different predicted outcome based on the domain arrangement (serial search). We used variational Bayesian inference, with the help of the MATLAB tool from Drugowitsch (2019) using model 15 (fully compensatory strategy using all domains) as the reference to compare all others. We assigned participants to the different decision groups, based on the model with the highest support (with a BF >3 compared to the reference model) given each subject’s responses.

### Parameter estimation and contrasts.

Finally, to estimate the differences in accuracy and search patterns among these decision groups, we performed bayesian parameter estimation using R (Version 4.0.2; R Core Team, 2020) and the R-packages *brms* (Version 2.13.0; Bürkner, 2017, 2018), and *rstan* (Version 2.19.3; Stan Development Team, 2020). All the comparisons reported use a ROPE of 0.1 on a standardized scale for log odds. For the accuracy in the different phases, we used a hierarchical logistic model, whereas the fixation analyses are built as hierarchical multinomial models.

All models estimated random effects for each participant, nested within groups. Each model was fitted using 4 chains with enough iterations to reach 10,000 ESS in the group level random parameters, in order to have reliable estimation of the standard deviation of the posterior. Given the complexity of the models, we assigned a small step size for the MCMC algorithm (0.99) to get rid of divergences while exploring the parameter space. Correspondingly, we also adjusted the treedepth (12-14) to allow the model to continue exploring for enough time.

For all the priors for the covariance matrices between random intercepts and slopes, we used a mildly regularizing prior of 2 for the parameter. For the variability among participants and groups, we used a mildly informative half Gaussian prior centered at 0, with a of 2.

Finally, we used the posterior samples transformed to standardized difference (Cohen ) to perform the follow up comparisons.

# Results

## Strategy classification

The results from the model selection showed heterogeneity in the strategy usage across the two experiments (figure  7). Two strategies are predominant among participants in both experiments: relying solely on the information from the most informative domain (1st only) and an incomplete integration of an equal weights to all the domains (Partial Tallying). Besides these two, we decided to examine the other strategies that would give participants good performance in this task. The first candidate for this was the lexicographic strategy Take-the-Best, for which there is widespread knowledge of its efficiency in non-compensatory binary decision paradigms like this (**references**). We also included the incomplete version of this strategy, combining both into a “Serial search” decision group. Finally, we decided to also include participants whose choices showed a strong reliance in the second most informative domain (2nd only), since this strategy should yield still reasonable performance on T1. On figure 8 we display these decision strategies and their frequency from across participants of both experiments.

## Performance

To assess participant’s accuracy in selecting the cue with the highest chances of winning, we ran a hierarchical logistic model, labeling correct and incorrect choices as 1 and 0’s, respectively. We used both the experiment and phase as separate predictors, with random effects and intercepts for each participant, nested within decision groups. For our priors, we used a Gaussian distribution for the effects and the intercept, centered at 2, with a sigma of 2. These values were selected, given the fact that we know that these strategies should perform well above chance level (0 in log odds), but we gave enough leeway for the model to explore different values.

We extracted samples from the posterior for the accuracy of the different strategies in each phase for both E1 and E2. We can see that the changes in contingencies across T1 and T2 affected the decision groups differently (figure 9). To look this in more detail, we looked at the difference in performance between the two phases (figure 16 and table 1). These comparisons show that, during E1, only the Partial Tallying group showed a meaningful reduction on their performance (PD > 0.999, ROPE overlap = 0.50%). Additionally, participants on the Serial Search group also showed a negative trend in their performance (PD = 0.995, ROPE overlap = 2.60%) but there is too much uncertainty on its magnitude. E2 shows a stronger effect across decision groups, decreasing strongly the performance of 1st Only (PD > 0.999, ROPE overlap = 0.00%), Partial Tallying (PD > 0.999, ROPE overlap = 0.00%), and Serial Search (PD > 0.999, ROPE overlap = 0.00%). On the other hand, participants classified as Tallying users show no meaningful change in their performance between phases (PD = 0.633, ROPE overlap = 65.30%), which matches their expected performance. Finally, despite the negative trend, we did not observe a meaningful decrease in accuracy for participants in the 2nd only strategy (PD = 0.941, ROPE overlap = 24.50%)

Moreover, when we compared the performance at T1 across strategy groups (figure 18) we found that Serial Search users tend to outperform more frugal strategies (1st only vs Serial search: PD = 0.997, ROPE overlap = 0.0%; 2nd only vs Serial search: PD = 0.99, ROPE overlap = 0.0%). Additionally, this strategy has also a tendency to outperform more compensatory strategies (Serial search vs Tallying PD = 0.986, ROPE overlap = 2.7%; Serial search vs Partial Tallying PD = 0.98, ROPE overlap = 9.8%) however there is a lot of uncertainty in our estimates as both 89% HDI for both comparisons still have some of their mass inside the designated ROPE. All paired comparisons are reported in detail in table 3

## Fixation patterns in each strategy

Although the decision models we created considered a wide array of information usage, there are some underlying assumptions that can only be explored by actually looking at the sampling patterns from the participants. For this reason, we used a series of eye tracking benchmarks to evaluate the starting point, spread of search and stopping point of the different models during T1.

To assess the starting point of the sampling process, we looked at the location of the first fixation using a hierarchical multinomial model, with a variable intercept for every participant, nested within decision groups. Each trial had four different outcomes, one for each of the domains presented. Figure 10 shows the proportion of first fixations allocated to each of the 4 domains by order of importance.

Next, to benchmark the spread of the sampling, we examined the allocation of fixations across the different domains. For this we used a hierarchical multinomial model with 4 outcomes similar to the one described above, but considering all the fixations in T1. In figure (figure 11) we can see the overall allocation of fixations across domains for each decision group.

Finally, to review the stopping point of the sampling process, we looked at the location of the last fixation in each trial. Again, using a hierarchical multinomial model with 4 outcomes, but using the only the last fixation recorded in each trial. In figure (figure 12) we can see the distribution of last fixations across domains.

Now, even though these metrics allow us to present a more detailed picture of the sampling process, they are insufficient to check its consistent application during the task. Moreover, some of these strategies, specifically, the ones using serial sampling, have a dynamic nature, depending on the particular states in each domain, and their ranking in terms of relative importance to other pieces of information. Specifically, these assume that the participant would start searching in descending order of importance, and stop sampling as soon as she determines there is a difference between cues in the highest ranking domain. This means that we would expect more sampling for information as the best discriminating domain is located further down in the scale of importance.

For these reasons, to analyze if this behavior was present in any of the decision strategies, we grouped trials according to which domain was discriminating between cues, in descending order of importance (6). This creates four decision scenarios:

* When the cue states in the 1st domain (ranked by importance) are different (A).
* When cue states are the same in 1st domain, but different in the 2nd (B).
* When cue states in both 1st and 2nd domain are the same, but the 3rd shows different states (C).
* All the domains but the least important have the same state (D).

We added these as a predictor in our statistical models, and allowed its slope to vary across participants, nested within decision groups. We applied this to assess if this had an effect to both the start (figure 14), spread (figure 14), and end (figure 15) of the sampling process.

### First only.

This model assumes that participants would only rely on the most informative domain to make their choices. This means that any instance where the most important cue is not different between cues, they should simply choose randomly between them. Therefore, we would expect a consistent preference towards the most informative domain in all the eye tracking measures.

We found that participants in this group tend to direct their first fixations towards the most important domain more often (figure 19, panel A), in comparison to the other domains (PD > 0.999, ROPE overlap = 0.00%). Using the decision scenario as a predictor in this metric (figure @ref(fig:first-scenario-comparisons-A, panel A)) did not reveal any change in this pattern (scenario A: PD > 0.999, ROPE overlap = 0.00%; scenario B: PD > 0.999, ROPE overlap = 0.00%; scenario C: PD > 0.999, ROPE overlap = 0.00%; scenario D: PD > 0.999, ROPE overlap = 0.00%). This pattern repeats when we take a look to the overall allocation of fixations (figure 20, panel A), showing a higher proportion towards the first domain vs the rest of the pieces of information (PD > 0.999, ROPE overlap = 0.0%). When we consider the different decision scenarios (figure ?? panel A), this result remains consistent, still favoring the most important domain (scenario A: PD > 0.999, ROPE overlap = 0.00%; scenario B: PD > 0.999, ROPE overlap = 0.00%; scenario C: PD > 0.999, ROPE overlap = 0.00%; scenario D: PD > 0.999, ROPE overlap = 0.00%). Moreover, when we look at the stopping point for this strategy (figure 21, panel A), we also see a preference for the most important domain (PD > 0.999, ROPE overlap = 0.00%). Finally, this remains consistent when we introduce the decision scenario as a predictor (figure ??, panel A), still displaying their preference towards the most informative domain (scenario A: PD > 0.999, ROPE overlap = 0.000%; scenario B: PD > 0.999, ROPE overlap = 0.000%; scenario C: PD > 0.999, ROPE overlap = 0.000%; scenario D: PD > 0.999, ROPE overlap = 0.000%).

### Second only.

This strategy is similar to the previously described, differing only in its reliance instead on the second most informative domain. This can happen as a overestimation of its importance by the participants using this strategy, leading them to think this domain is the most predictive of success in the task. Again, this should be reflected by a consistent preference for the second most important domain, which should not be affected by decision scenarios.

As expected, we see how these subjects preferred the 2nd most important domain as their starting point for their sampling (figure 19, panel B; PD = 0.999, ROPE overlap = 0.00%). However, we see that this effect is not as big as the observed in the 1st only strategy, which manifests further down when we take a look at this metric across decision scenarios (figure ??, panel B). Here, although the probability of direction favors the 2nd most important domain (figure ??, panel B), our estimation carries a lot of uncertainty, and therefore we cannot conclude this effect is meaningful in any of the decision scenarios (scenario A: PD = 0.895, ROPE overlap = 18.10%; scenario B: PD = 0.883, ROPE overlap = 19.90%; scenario C: PD = 0.891, ROPE overlap = 17.90%; scenario D: PD = 0.794, ROPE overlap = 23.50%). Looking at the sampling spread for this strategy, it shows us an overall preference for the second most important domain (figure 20, panel B; PD > 0.999, ROPE overlap = 0.0%). This pattern holds for scenario A and B, but the scenario C and D, while they show a probability of direction favoring the 2nd domain, they still show high uncertainty in its magnitude (figure ??; scenario A: PD = 0.998, ROPE overlap = 0.000%; scenario B: PD > 0.999, ROPE overlap = 0.000%; scenario C: PD = 0.991, ROPE overlap = 4.800%; scenario D: PD = 0.949, ROPE overlap = 21.400%). Finally, the stopping point for this strategy shows a tendency to favor the expected domain, however, this difference overlaps with the defined ROPE, which indicates this may not be a meaningful effect (figure 21, panel B; PD = 0.999, ROPE overlap = 0.00%). When we look more closely at each decision scenario, we can see that the last fixation only shows a meaningful effect favoring the 2nd domain in scenario B (figure ??, panel B; scenario A: PD = 0.973, ROPE overlap = 7.600%; scenario B: PD = 0.999, ROPE overlap = 0.000%; scenario C: PD = 0.802, ROPE overlap = 39.800%; scenario D: PD = 0.549, ROPE overlap = 44.800%).

### Serial search.

So far, the strategies analyzed predict static patterns, preferring always a particular domain over the others. For users of a strategies that rely in conditional sampling, this changes depending on the state of the information presented in a given trial. In particular, for the participants grouped in our Serial search group, we predict they should start by looking at the most important domain, comparing it between cues, and only continue sampling if this does not discriminate between cues. Therefore, we expect an overall preference for the most important domain in all the metrics we used, however, as we divide the trials by decision scenarios, we would expect that participants using this strategy should allocate more fixations to the other domains, and they should finish their sampling further down in the domain rankings.

Our results show that in effect, these participants start on the most important domain (figure 19, panel C; PD > 0.999, ROPE overlap = 0.00%), remaining constant across decision scenarios (figure ??, panel C; scenario A: PD > 0.999, ROPE overlap = 0.00%; scenario B: PD > 0.999, ROPE overlap = 0.00%; scenario C: PD = 0.999, ROPE overlap = 0.00%; scenario D: PD > 0.999, ROPE overlap = 0.00%). For the proportion of fixations, we were interested in testing whether participants allocated their fixations following a linear trend, for which purpose we performed a series of paired comparisons (figure 20). We found meaningful differences between the most important levels (1st vs 2nd: PD = 0.991, ROPE overlap = 0.0%; 2nd vs 3rd: PD = 0.992, ROPE overlap = 0.0%), but not for the 3rd vs 4th (PD = 0.613, ROPE overlap = 55.1%). When comparing the magnitude of the meaningful effects we found, we did not find any differences between them, indicating a linear relationship between fixation proportions and domain importance from the 1st to the 3rd most important domain (PD = 0.66, ROPE overlap = 24.0%). Looking at the sampling across scenarios (??), we can see that the fixations in scenario A, favor the most important domain (PD > 0.999, ROPE overlap = 0.000%). In scenario B, the difference between the fixations to domain 1 and 2 narrows, making it not meaningful anymore (PD = 0.91, ROPE overlap = 14.000%), while still being larger than the proportion of fixations directed to the least informative domains (PD > 0.999, ROPE overlap = 0.000%). This pattern stops when we reach scenario C, where the proportion of fixations towards domain 1 still is larger than to domain 3 (PD = 0.989, ROPE overlap = 0.000%), which repeats again in scenario D (PD = 0.993, ROPE overlap = 0.000%). This suggest that the amount of sampling only equalizes for the two most important domains, and does not occur as consistently for the scenarios where the best discriminating domain is located further down in the scale, more consistent with an incomplete use of the domains. Finally, the analysis of last fixations only the comparison between the 1st and 2nd most important domain shows a meaningful difference, favoring the latter (PD = 0.983, ROPE overlap = 0.00%). When looking across scenarios, this benchmark shows a preference for the most important domain in scenario A (PD > 0.999, ROPE overlap = 0.000%), while in scenaro B we can see a tendency favoring the 2nd most important domain, but still overlapping slightly with the ROPE (PD = 0.984, ROPE overlap = 1.600%). Both scenario C and D do not show a meaningful difference towards the domain 3 and 4, respectively.

### Tallying.

This strategy integrates the sum of all the attributes as a single predictor. This makes all the domain equal in terms of the importance of its information, selecting the cue with more favorable features. Considering this, we should see no difference between the allocation of fixations to the different domains, reflecting an equal consideration of all of them by the participants. Moreover, the sampling should remain pretty much the same across decision scenarios, since these conditions should bear no effect in how much information they need to make their choices.

We can see that the starting point (figure 19, panel D) does not seem to be allocated to any of the domains in particular. However, there is a trend favoring the most informative domain compared to the 2nd (PD = 0.896, ROPE overlap = 8.20%), 3rd (PD = 0.902, ROPE overlap = 7.90%), and 4th (PD = 0.922, ROPE overlap = 6.40%), but the uncertainty associated to its estimate does not allow us to mark it as meaningful. Looking at this metric across decision scenarios does not show any disruption in the general pattern observed previously (figure ??). Additionally, the overall allocation of fixations does not seem to lean towards any of the domains in particular (figure 20). However, when we analyze these comparisons across decision scenarios, we can see a meaningful difference between the allocation of fixations to domain 1 and 4 in scenario A, favoring the latter (PD = 0.995, ROPE overlap = 0.000%). This does indicate at least a distinction between the most informative and least informative domains from the participants classified in this group by the choice model. This observation is reinforced by the results on their stopping point, showing a meaningful difference which favors domain 1 when compared to domain 4 (PD = 0.989, ROPE overlap = 0.00%). Furthermore, the analysis of the location of last fixation in the decision scenarios shows a meaningful difference that favors the most informative domain over the 2nd (PD = 0.982, ROPE overlap = 0.000%), 3rd (PD = 0.993, ROPE overlap = 0.000%), and 4th (PD = 0.998, ROPE overlap = 0.000%) domains in scenario A.

### Partial Tallying.

The last strategy we analyze in detail, only differs from its more compensatory parent in the number of domains added up to determine the cue with more positive attributes. In this particular case, the participants classified in this group only integrate the 1st and 2nd most important domains, therefore identifying explicitly these two as superioir sources of information compared to the rest of the domains. Still, these two domains share the same level in the hierarchy, making both equally important to determine their choice. This small differentiation makes a some different predictions for some of the fixation metrics that we review. Overall, we should see a preference for these two domains in all the metrics when compared to the least informative domains, but we should see no difference between the 1st and 2nd domain, given its equal subjective importance.

For the starting point, we can see that this strategy does not show the expected pattern of favoring the most important domains over the least important ones (figure 19, panel E; PD = 0.677, ROPE overlap = 41.50%). The different scenarios do not make these comparisons differ in a meaningful way (figure ??). The overall proportion of fixations for this strategy (figure 20, panel E) does show the predicted preference towards the most important domains (PD = 0.998, ROPE overlap = 0.0%), while there is no meaningful difference between domain 1 and 2 (PD = 0.517, ROPE overlap = 74.0%), and domains 3 and 4 (PD = 0.673, ROPE overlap = 67.6%). This pattern only holds for scenario A (PD > 0.999, ROPE overlap = 0.000%) and B (PD = 0.999, ROPE overlap = 0.000%). Finally, the stopping point shows no preference towards the most important domains (figure 21, panel E; PD = 0.974, ROPE overlap = 17.60%), and only shows a tendency favoring them in scenario A (PD = 0.995, ROPE overlap = 1.200%).

# Discussion

# References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. <https://doi.org/10.1163/156856897X00357>

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>

Drugowitsch, J. (2019). Variational Bayesian inference for linear and logistic regression. *arXiv:1310.5438 [Stat]*. Retrieved from <http://arxiv.org/abs/1310.5438>

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823. <https://doi.org/10.3758/s13428-016-0822-1>

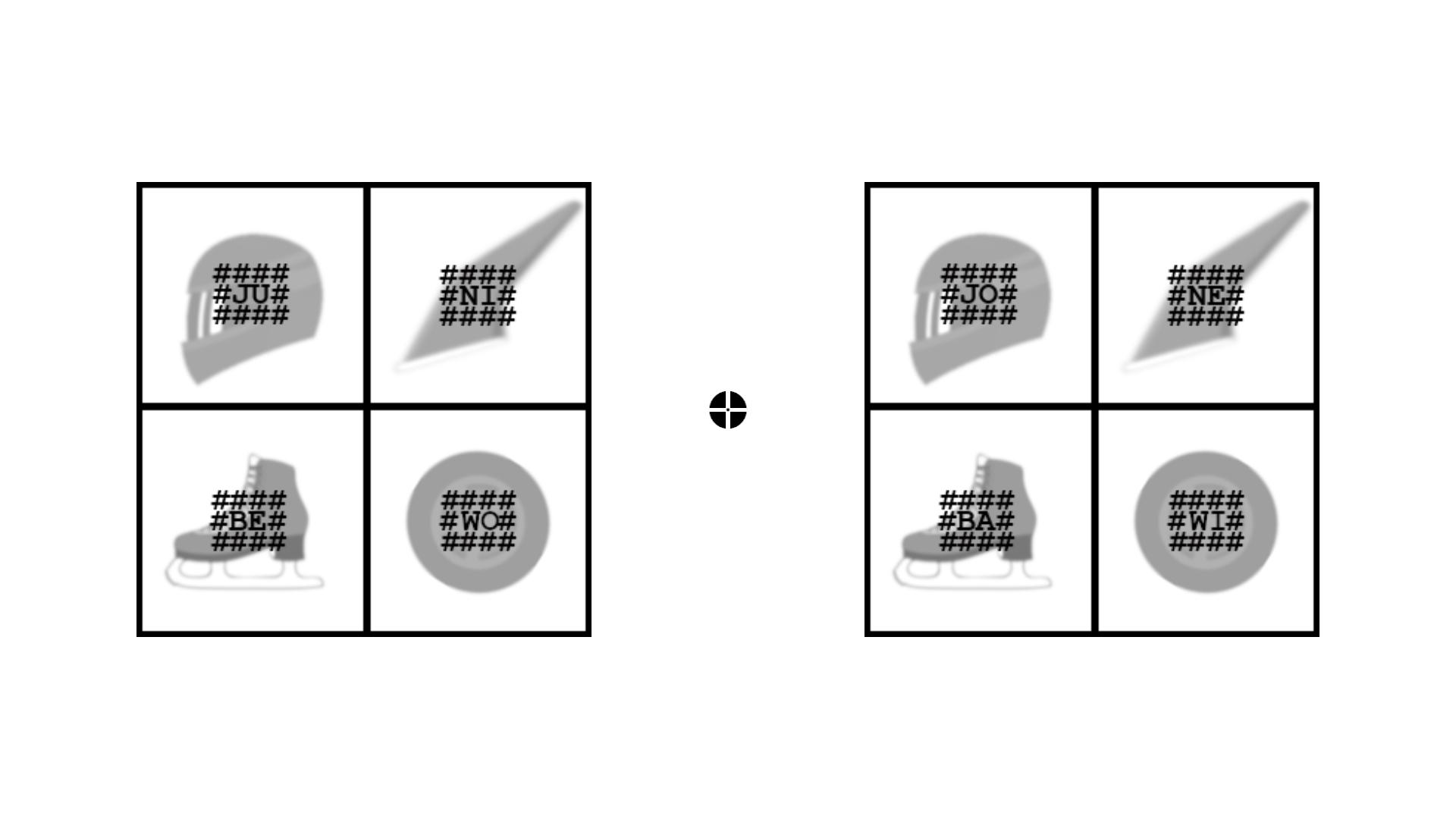
Niehorster, D. C., & Nyström, M. (2020). SMITE: A toolbox for creating Psychophysics Toolbox and PsychoPy experiments with SMI eye trackers. *Behavior Research Methods*, *52*(1), 295–304. <https://doi.org/10.3758/s13428-019-01226-0>

Oh, H., Beck, J. M., Zhu, P., Sommer, M. A., Ferrari, S., & Egner, T. (2016). Satisficing in split-second decision making is characterized by strategic cue discounting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1937–1956. <https://doi.org/10.1037/xlm0000284>

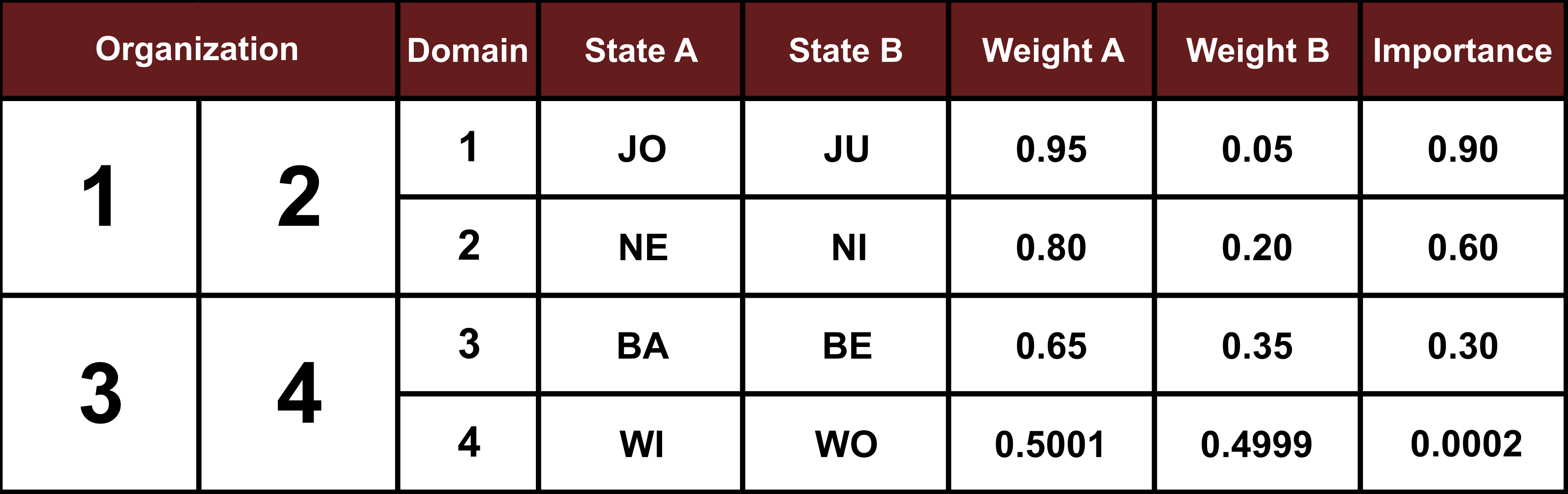
R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Stan Development Team. (2020). RStan: The R interface to Stan. Retrieved from <http://mc-stan.org/>

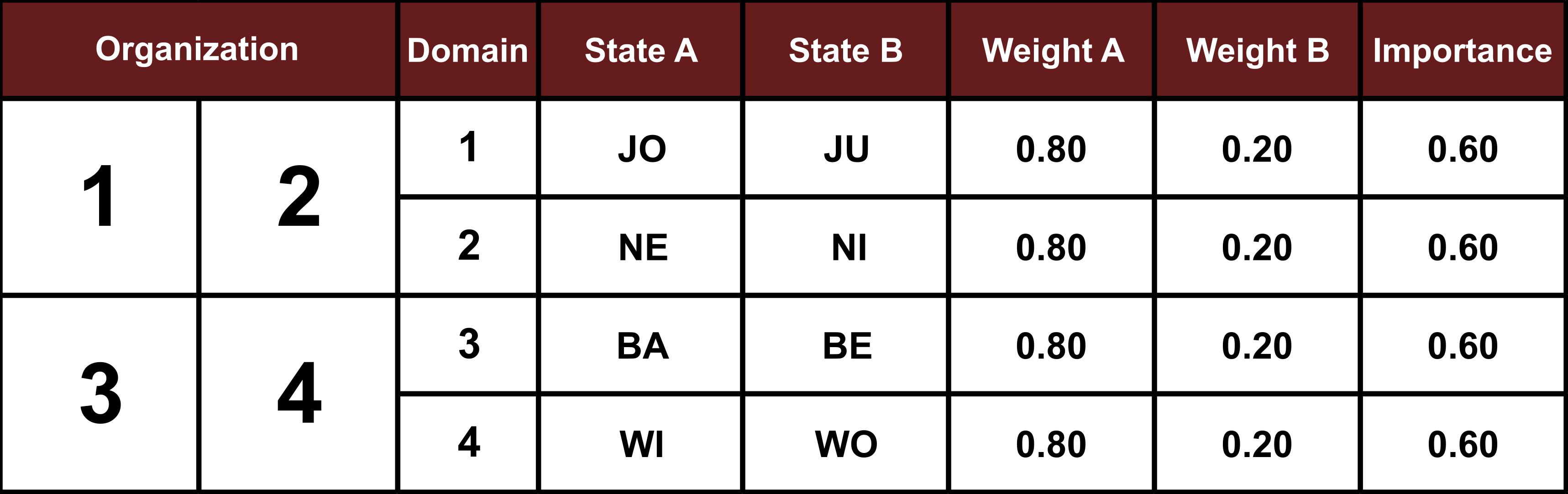
# Figures and Tables



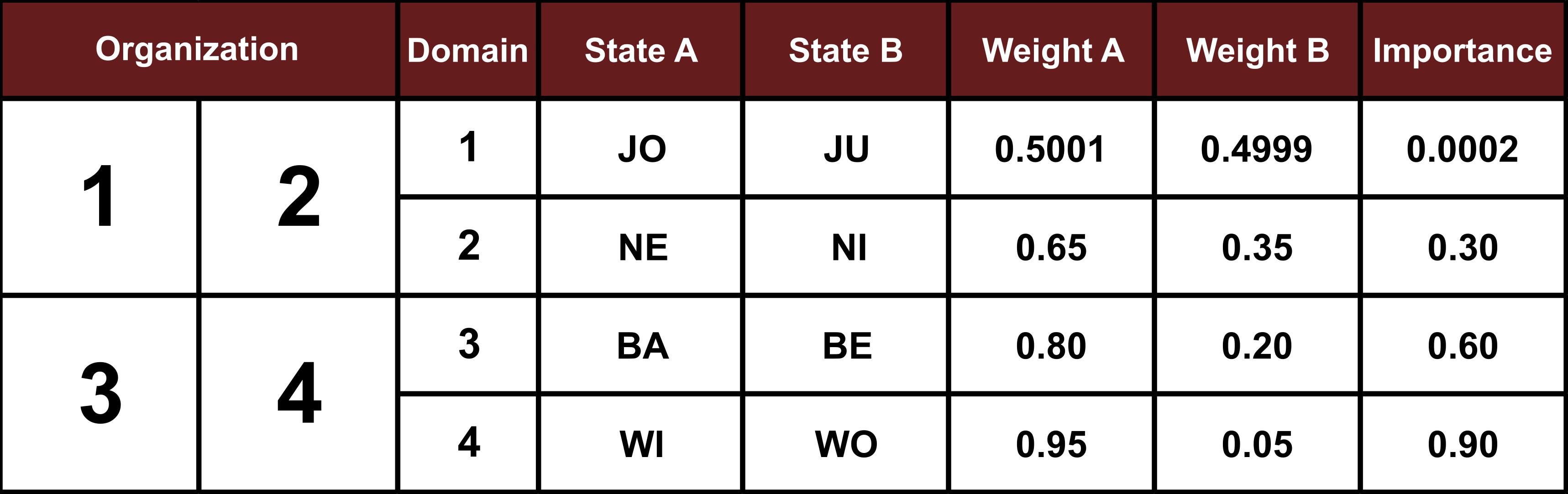
*Figure* *1.*  Sample stimulus presented to participants.



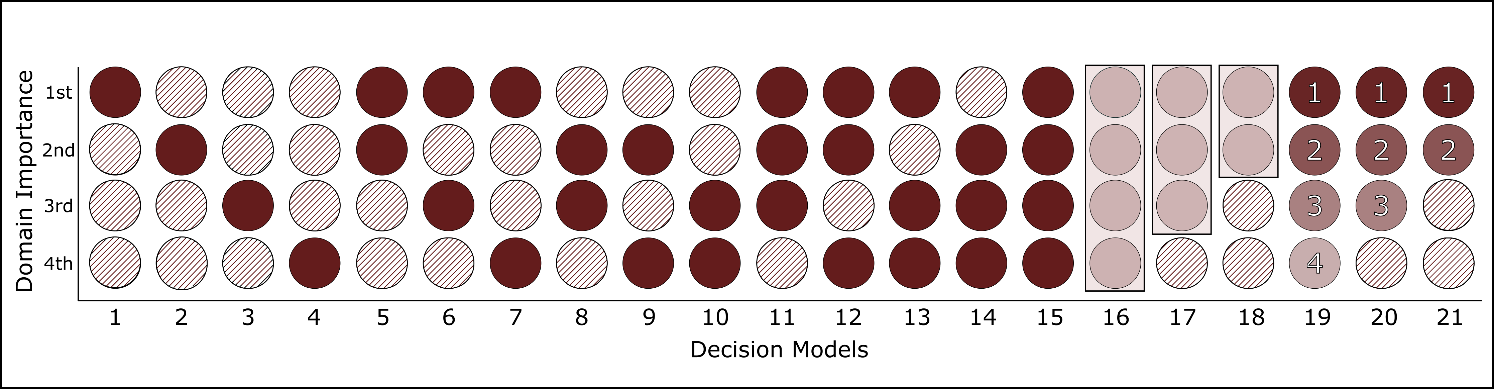
*Figure* *2.*  Example of the organization of cue weights across domains during training and T1. All the equipment (domains) and brands (states) kept their position constant across subjects, but the allocation of weights was assigned semirandomly across them.



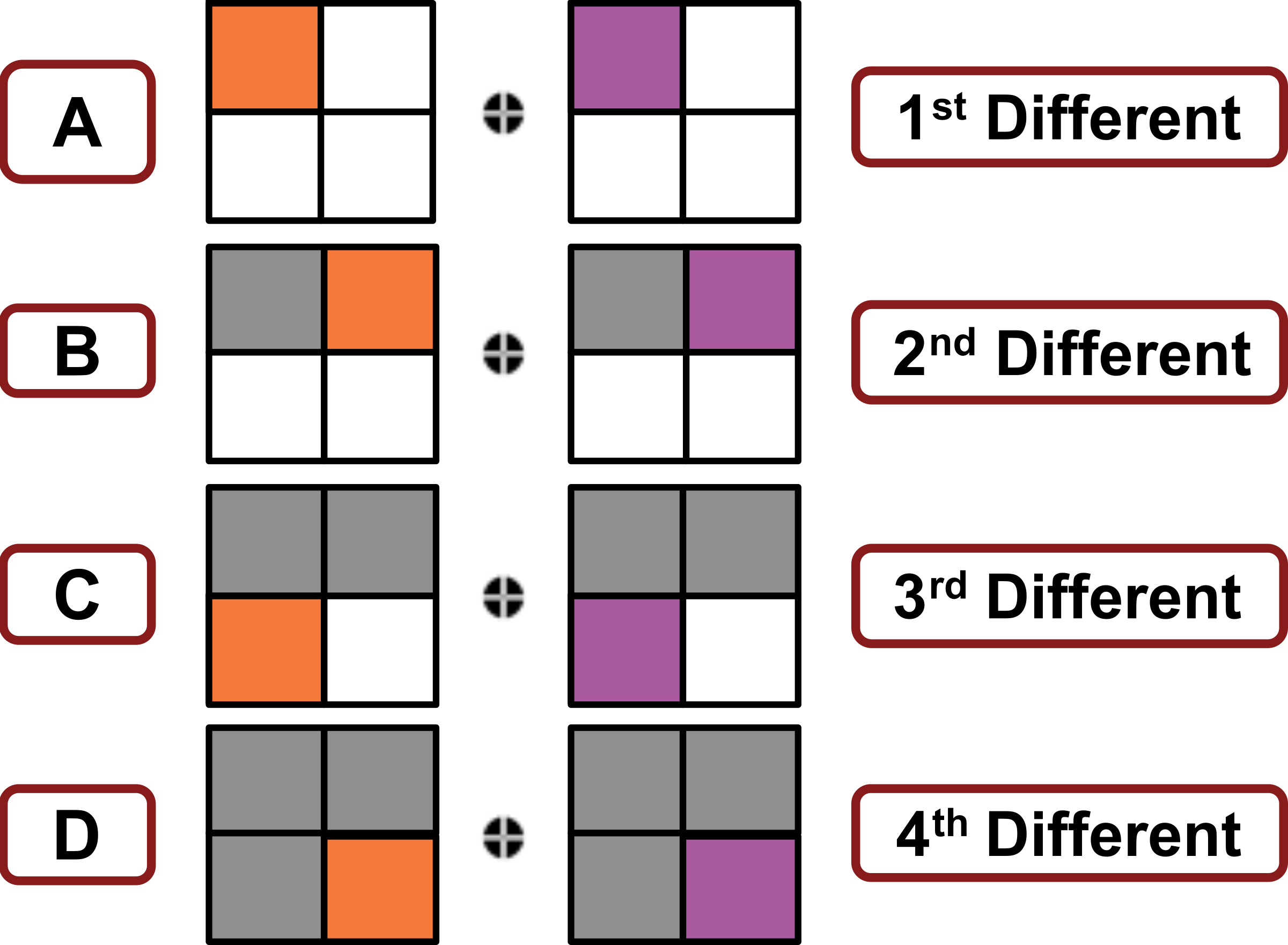
*Figure* *3.*  Example of the organization of cue weights across domains during T2 in E1. All the different pairs had the same contingencies, preserving winning states.



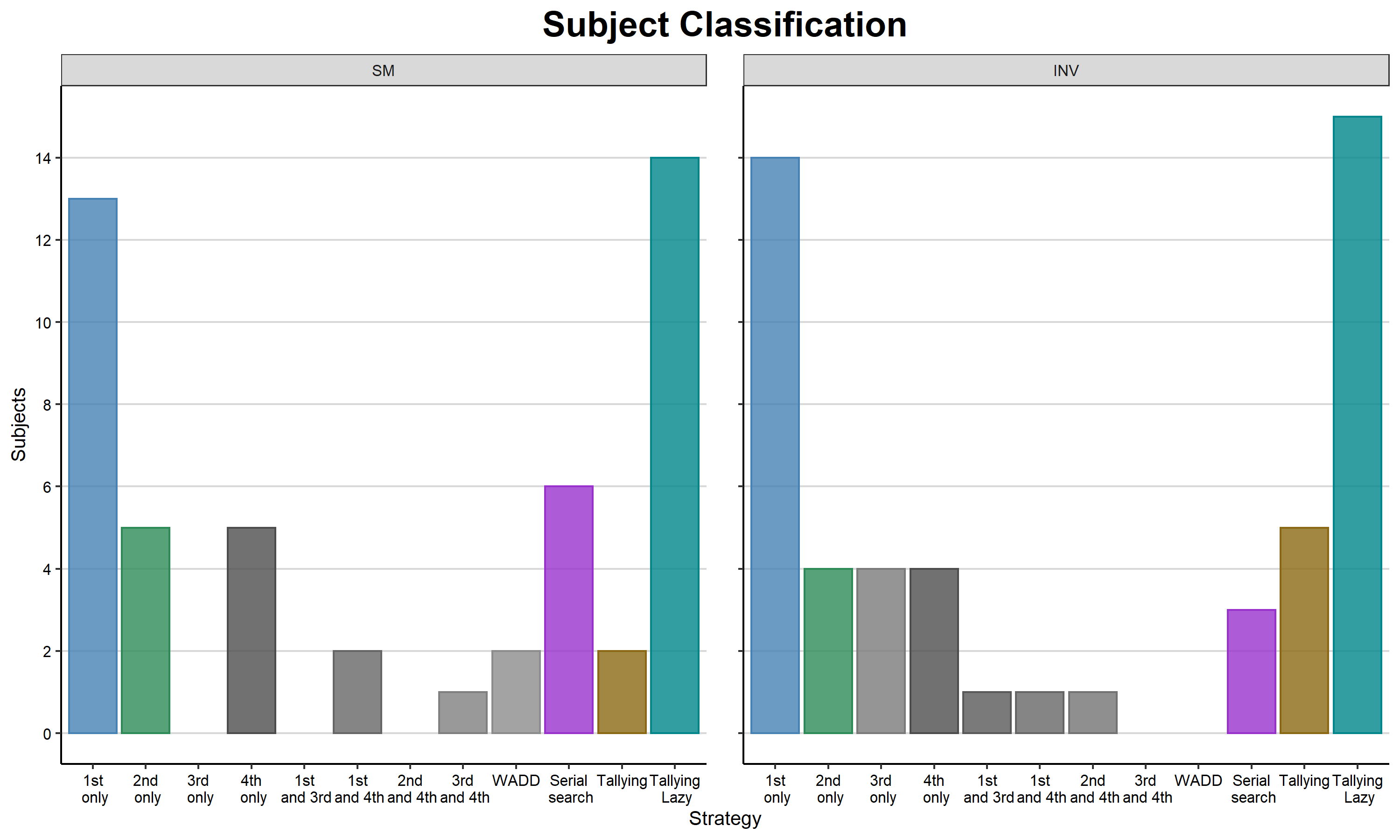
*Figure* *4.*  Example of the organization of cue weights across domains during T2 in E2. The different domains show now inverted contingencies, preserving winning states.



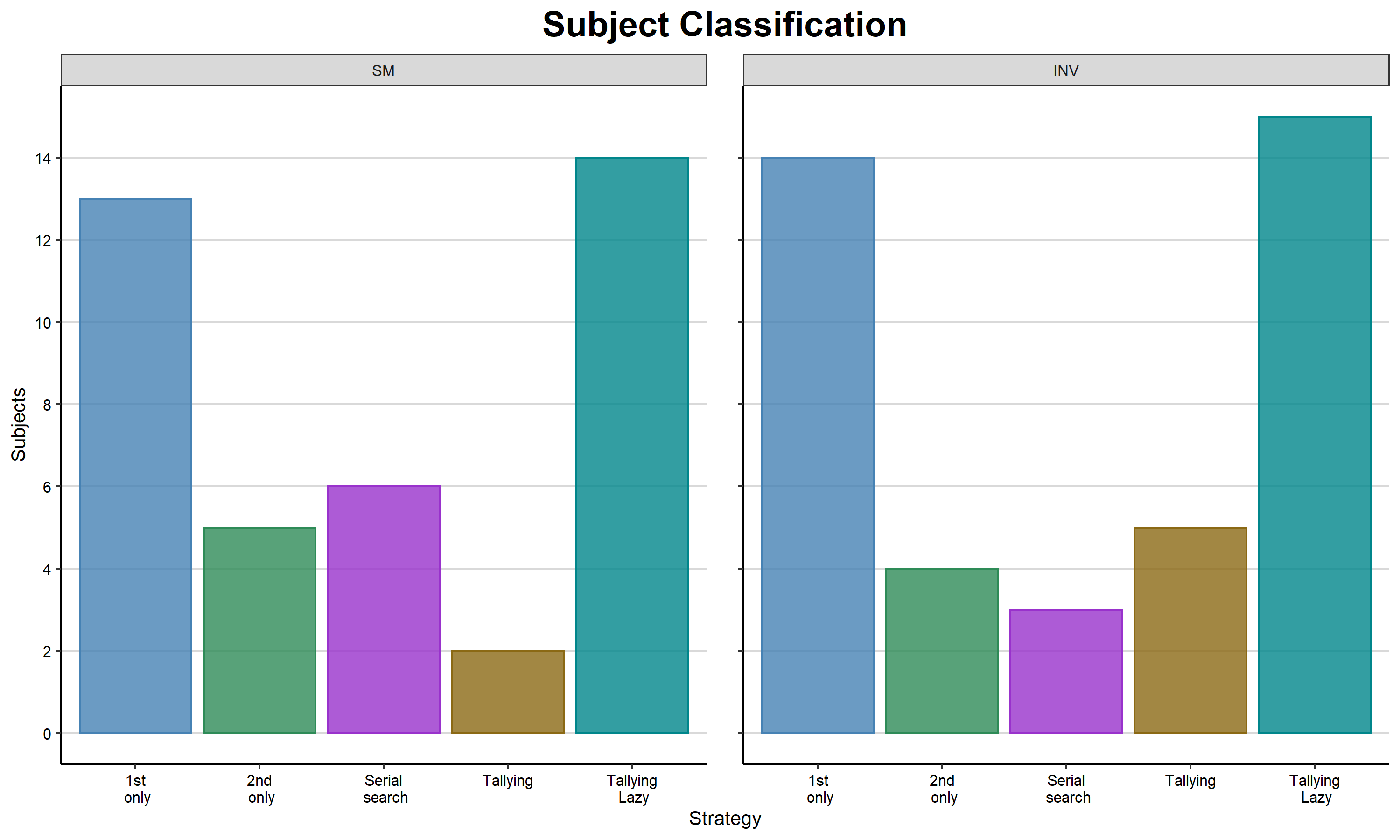
*Figure* *5.*  Information usage in the different decision models. In models 1-15, black circles indicate the domain(s) used by each strategy. For models 16-18 we created a model that integrated the domains in a single piece of information (i.e. a sum of positive traits). Models 19-21 consider a serial search for information, starting in the most informative domain, continuing in descending order of importance, and stoping as soon as any of these domains discriminates between cues.



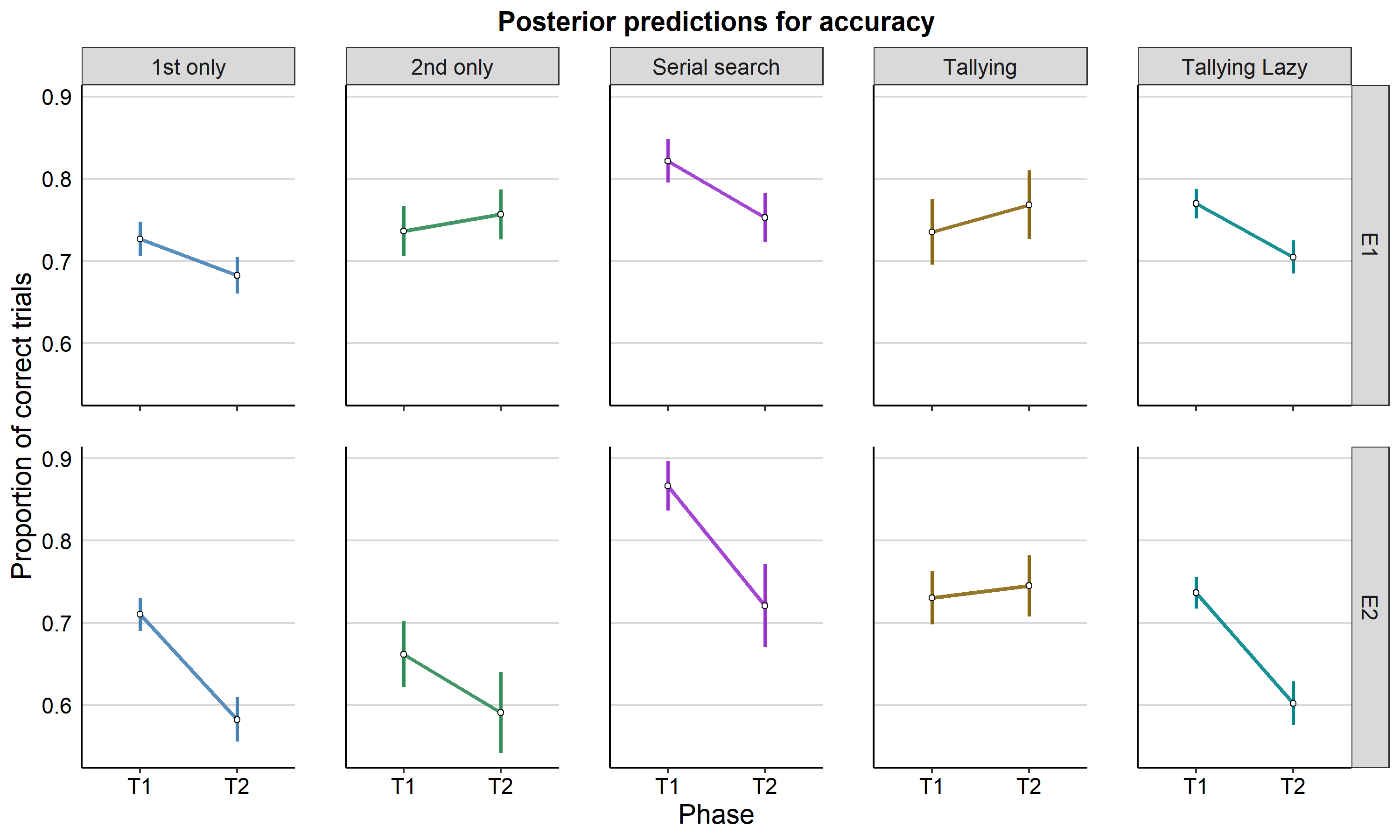
*Figure* *6.*  Decision scenarios according to the best discriminating domain between cues.



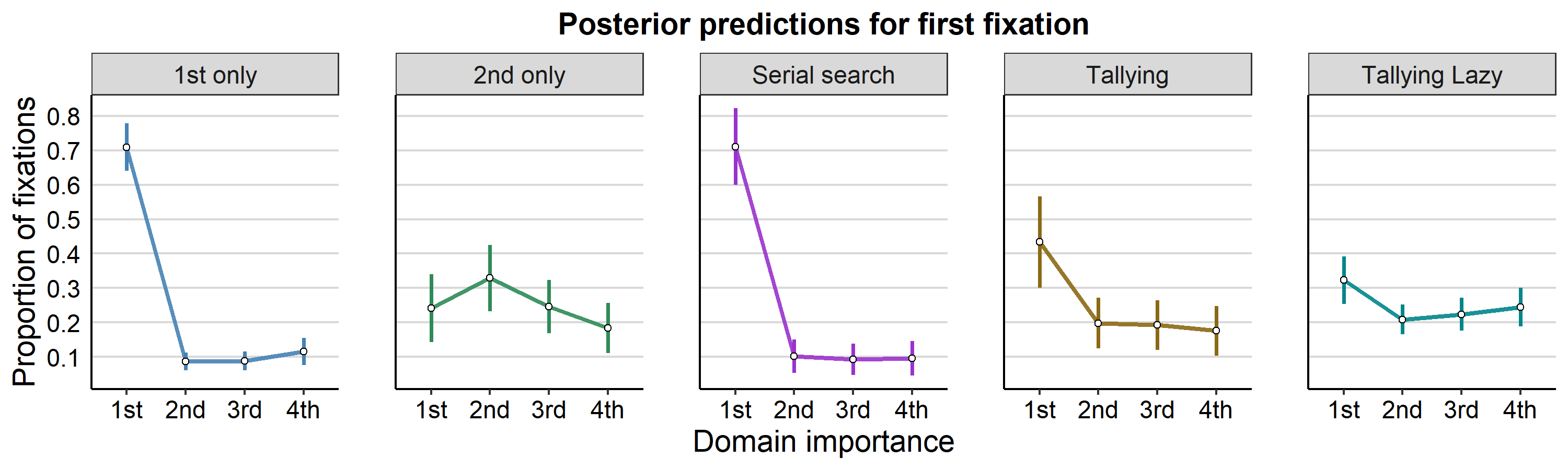
*Figure* *7.*  Strategy group classification based on the variational Bayesian inference.



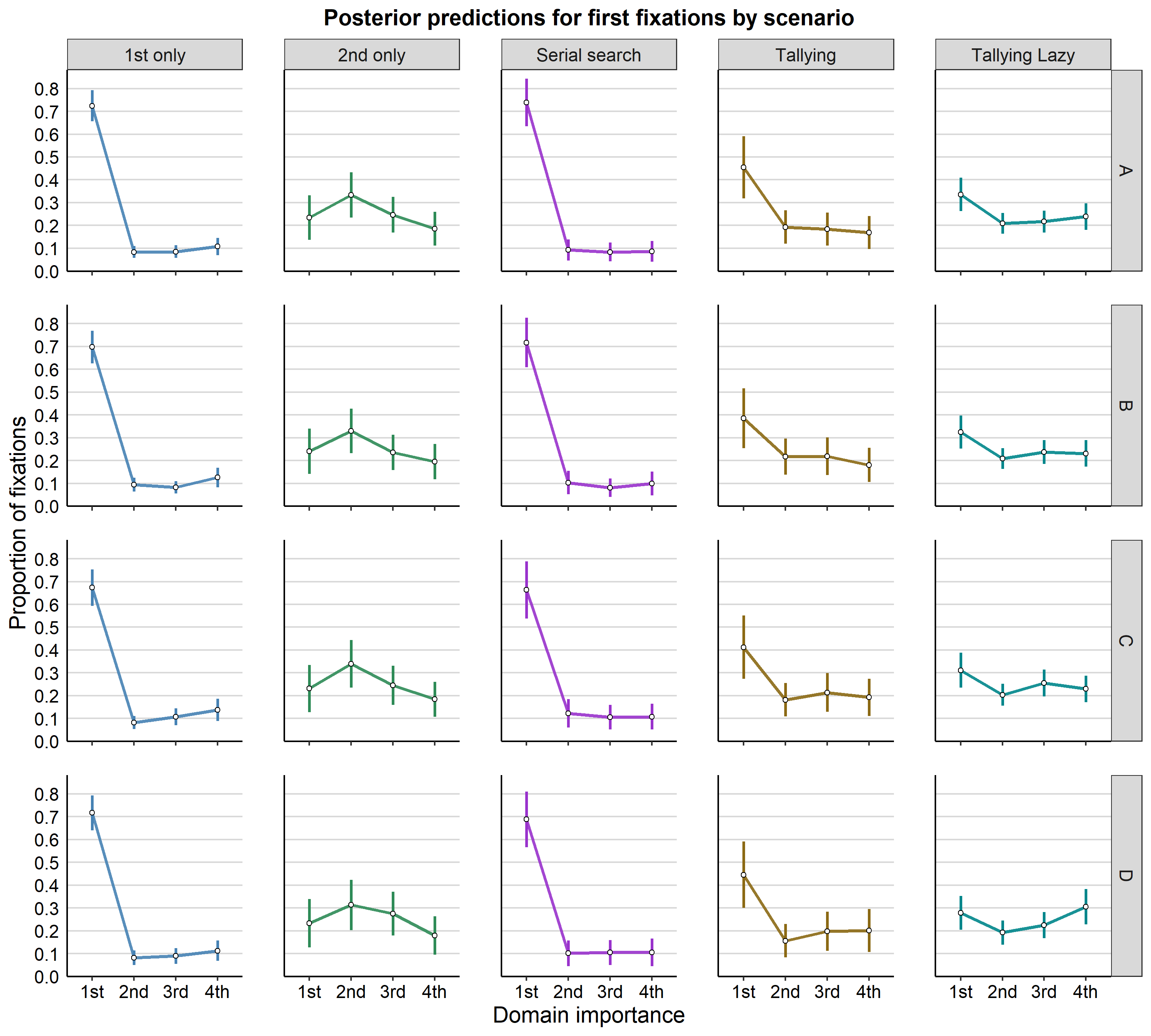
*Figure* *8.*  Decision groups studied in detail for this work.



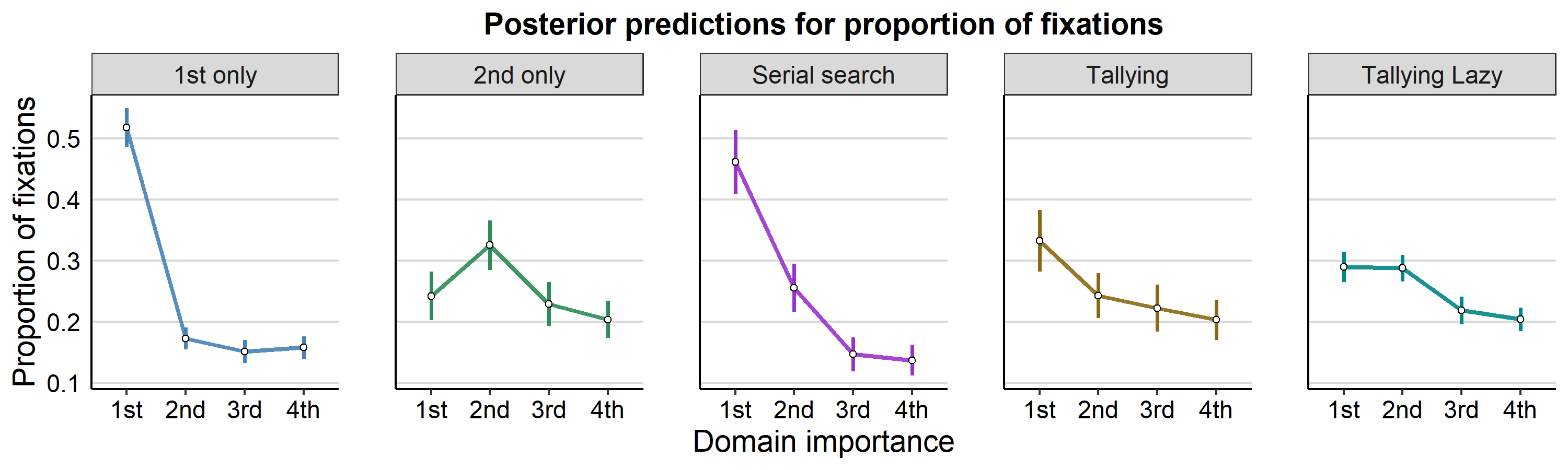
*Figure* *9.*  Performance (as percent of accurate trials) across testing phases. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



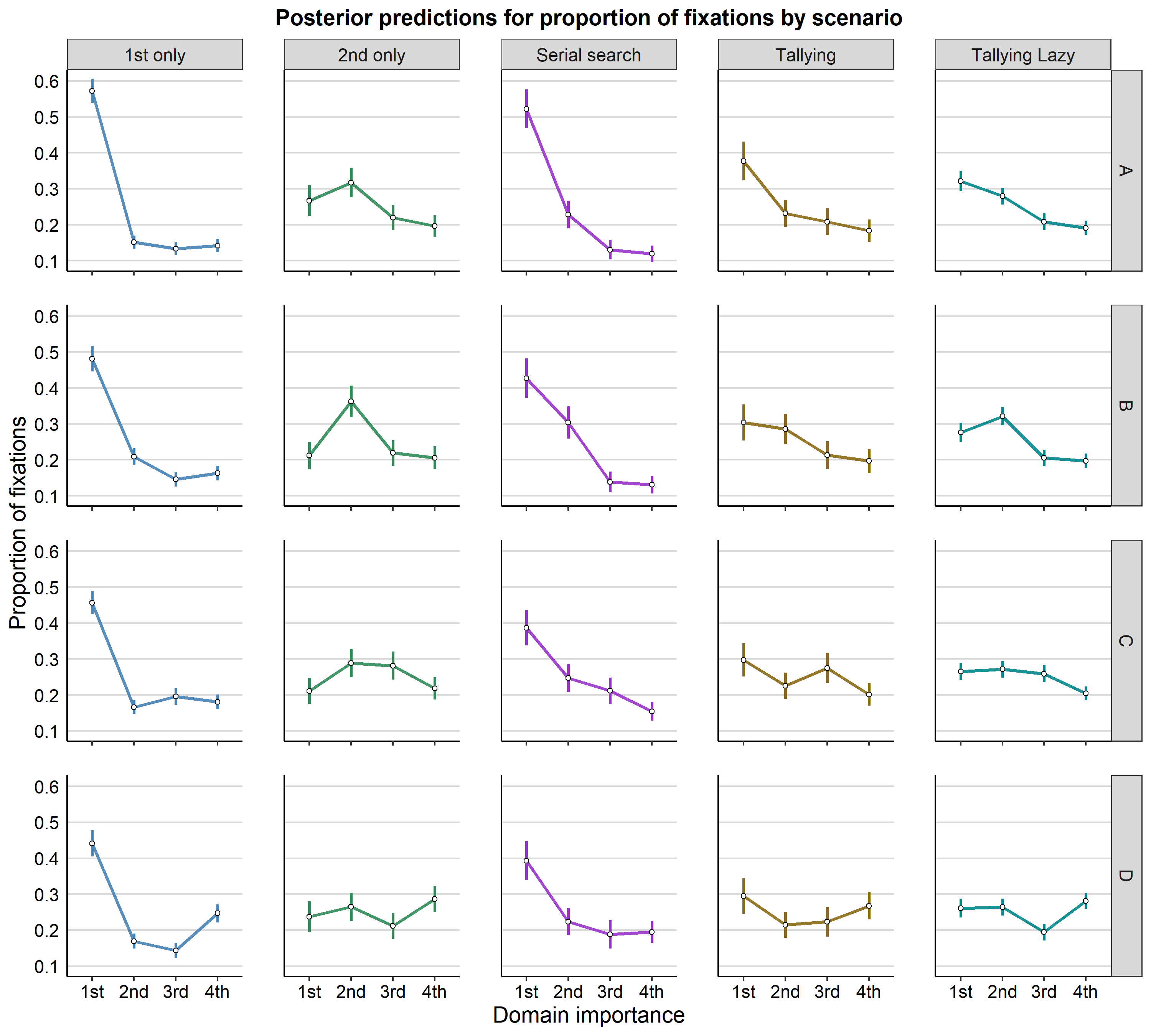
*Figure* *10.*  Allocation of first fixations across domains in T1. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



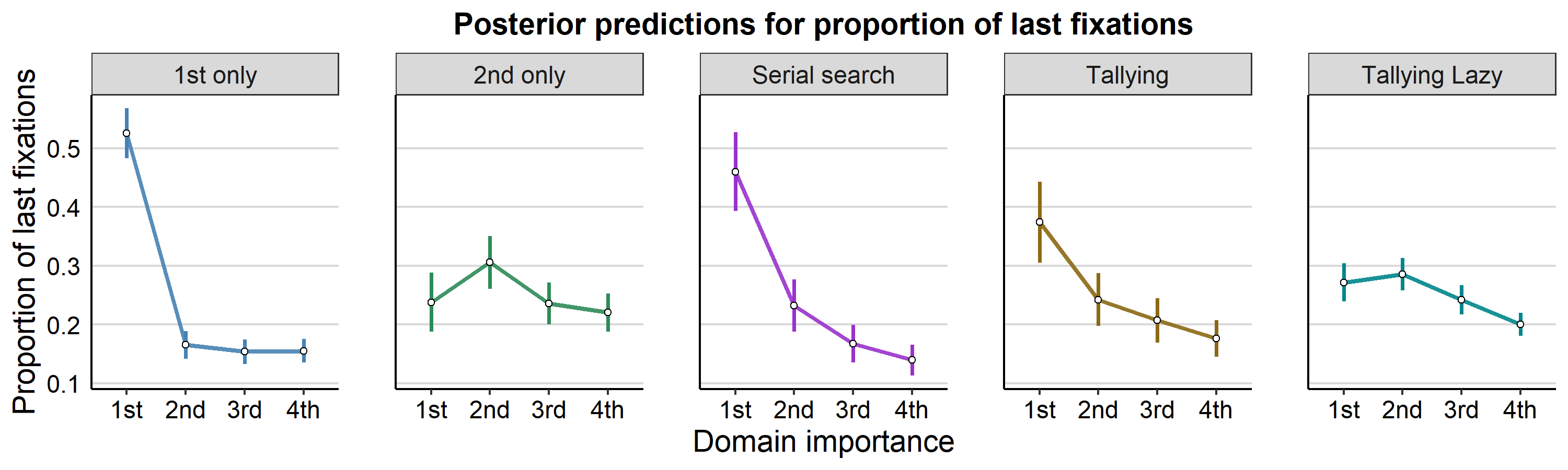
*Figure* *13.*  Allocation of first fixations across domains in the different decision scenarios in T1. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



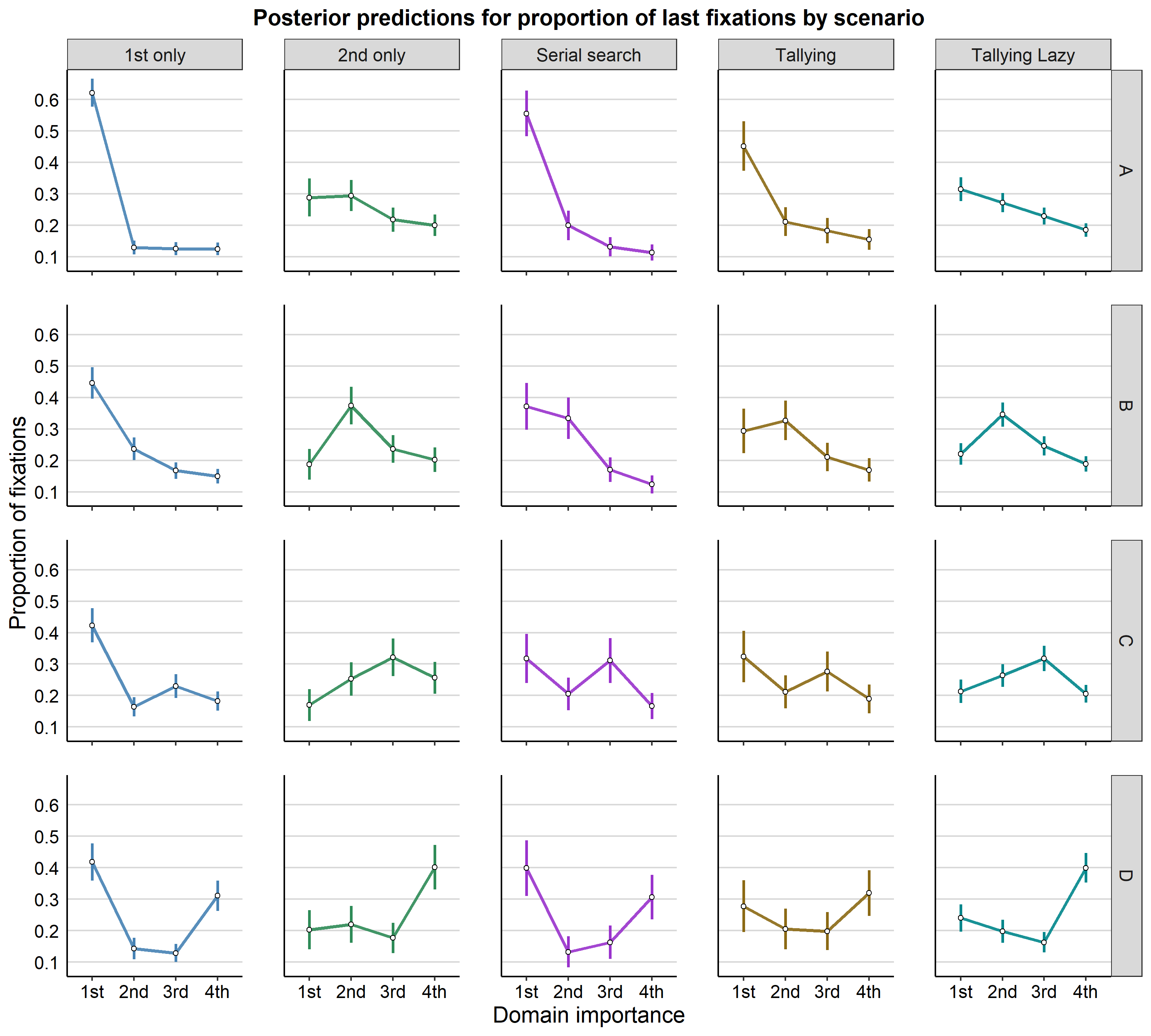
*Figure* *11.*  Allocation of fixations across domains in T1 for the different decision groups. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



*Figure* *14.*  Allocation of fixations across domains in T1 for the different decision groups. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



*Figure* *12.*  Allocation of the last fixations across domains in T1. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.



*Figure* *15.*  Allocation of last fixations across domains in the different decision scenarios in T1. Individual dots correspond to the median from the posterior draws from each participant. Error bars represent standard deviations.

Table 1:

Accuracy between phases across groups

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| experiment | strategy | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| E1 | 1st only | 0.35 | 0.98 | 35.40% | = 0.981 |
| E1 | 2nd only | 0.68 | 0.74 | 68.40% | = 0.735 |
| E1 | Serial search | 0.03 | 1.00 | 2.60% | = 0.995 |
| E1 | Tallying | 0.48 | 0.77 | 47.80% | = 0.77 |
| E1 | Tallying Lazy | 0.00 | 1.00 | 0.50% | > 0.999 |
| E2 | 1st only | 0.00 | 1.00 | 0.00% | > 0.999 |
| E2 | 2nd only | 0.24 | 0.94 | 24.50% | = 0.941 |
| E2 | Serial search | 0.00 | 1.00 | 0.00% | > 0.999 |
| E2 | Tallying | 0.65 | 0.63 | 65.30% | = 0.633 |
| E2 | Tallying Lazy | 0.00 | 1.00 | 0.00% | > 0.999 |

Table 2:

Paired comparisons between groups on the difference in performance across phases

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| experiment | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| E1 | first\_lazy | 0.72 | 0.82 | 71.50% | = 0.82 |
| E1 | first\_second | 0.21 | 0.96 | 21.10% | = 0.965 |
| E1 | first\_serial | 0.44 | 0.86 | 44.30% | = 0.859 |
| E1 | first\_tally | 0.17 | 0.94 | 17.10% | = 0.945 |
| E1 | second\_lazy | 0.04 | 0.99 | 3.50% | = 0.989 |
| E1 | second\_serial | 0.03 | 0.99 | 3.40% | = 0.986 |
| E1 | second\_tally | 0.58 | 0.62 | 58.00% | = 0.617 |
| E1 | serial\_lazy | 0.71 | 0.67 | 71.10% | = 0.671 |
| E1 | serial\_tally | 0.03 | 0.98 | 3.40% | = 0.977 |
| E1 | tally\_lazy | 0.04 | 0.98 | 3.90% | = 0.98 |
| E2 | first\_lazy | 0.81 | 0.64 | 81.30% | = 0.635 |
| E2 | first\_second | 0.33 | 0.88 | 33.40% | = 0.877 |
| E2 | first\_serial | 0.24 | 0.92 | 24.50% | = 0.915 |
| E2 | first\_tally | 0.00 | 1.00 | 0.00% | = 0.996 |
| E2 | second\_lazy | 0.24 | 0.90 | 24.40% | = 0.904 |
| E2 | second\_serial | 0.02 | 0.98 | 1.70% | = 0.976 |
| E2 | second\_tally | 0.23 | 0.91 | 23.50% | = 0.909 |
| E2 | serial\_lazy | 0.33 | 0.88 | 32.80% | = 0.875 |
| E2 | serial\_tally | 0.00 | 1.00 | 0.00% | = 0.999 |
| E2 | tally\_lazy | 0.00 | 1.00 | 0.00% | = 0.999 |

Table 3:

Paired comparisons between groups on the performance in T1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| first\_lazy | 0.51 | 0.89 | 50.7% | = 0.892 |
| first\_second | 0.62 | 0.62 | 61.5% | = 0.616 |
| first\_serial | 0.00 | 1.00 | 0.0% | = 0.997 |
| first\_tally | 0.68 | 0.66 | 68.0% | = 0.66 |
| second\_lazy | 0.35 | 0.90 | 34.7% | = 0.905 |
| second\_serial | 0.00 | 0.99 | 0.0% | = 0.99 |
| second\_tally | 0.48 | 0.71 | 47.9% | = 0.712 |
| serial\_lazy | 0.10 | 0.98 | 9.8% | = 0.98 |
| serial\_tally | 0.03 | 0.99 | 2.7% | = 0.986 |
| tally\_lazy | 0.65 | 0.68 | 65.1% | = 0.677 |

Table 4:

Follow up comparisons for the proportion of first fixations across decision scenarios.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| strategy | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | d1\_all | 0.00 | 1.00 | 0.00% | > 0.999 |
| 2nd only | d2\_all | 0.00 | 1.00 | 0.00% | = 0.999 |
| Serial search | d1\_all | 0.00 | 1.00 | 0.00% | > 0.999 |
| Tallying | d1\_2 | 0.08 | 0.90 | 8.20% | = 0.896 |
| Tallying | d1\_3 | 0.08 | 0.90 | 7.90% | = 0.902 |
| Tallying | d1\_4 | 0.06 | 0.92 | 6.40% | = 0.922 |
| Tallying | d2\_3 | 0.29 | 0.53 | 29.40% | = 0.533 |
| Tallying | d2\_4 | 0.25 | 0.59 | 25.00% | = 0.593 |
| Tallying | d3\_4 | 0.26 | 0.57 | 26.50% | = 0.569 |
| Tallying Lazy | d1\_2 | 0.17 | 0.88 | 16.60% | = 0.885 |
| Tallying Lazy | d12\_34 | 0.42 | 0.68 | 41.50% | = 0.677 |
| Tallying Lazy | d3\_4 | 0.35 | 0.60 | 34.90% | = 0.601 |

Table 5:

Follow up comparisons for the proportion of fixations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| strategy | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | d1\_all | 0.00 | 1.00 | 0.0% | > 0.999 |
| 2nd only | d2\_all | 0.00 | 1.00 | 0.0% | > 0.999 |
| Serial search | d1\_2 | 0.00 | 0.99 | 0.0% | = 0.991 |
| Serial search | d2\_3 | 0.00 | 0.99 | 0.0% | = 0.992 |
| Serial search | d3\_4 | 0.55 | 0.61 | 55.1% | = 0.613 |
| Serial search | lin1 | 0.24 | 0.66 | 24.0% | = 0.66 |
| Serial search | lin2 | 0.16 | 0.89 | 16.3% | = 0.889 |
| Serial search | lin3 | 0.15 | 0.63 | 15.2% | = 0.629 |
| Tallying | d1\_2 | 0.20 | 0.89 | 20.1% | = 0.887 |
| Tallying | d1\_3 | 0.11 | 0.92 | 10.9% | = 0.924 |
| Tallying | d1\_4 | 0.04 | 0.97 | 3.9% | = 0.967 |
| Tallying | d2\_3 | 0.42 | 0.67 | 42.2% | = 0.667 |
| Tallying | d2\_4 | 0.36 | 0.78 | 36.4% | = 0.78 |
| Tallying | d3\_4 | 0.51 | 0.65 | 50.7% | = 0.647 |
| Tallying Lazy | d1\_2 | 0.74 | 0.52 | 74.0% | = 0.517 |
| Tallying Lazy | d12\_34 | 0.00 | 1.00 | 0.0% | = 0.998 |
| Tallying Lazy | d3\_4 | 0.68 | 0.67 | 67.6% | = 0.673 |

Table 6:

Follow up comparisons for the last fixations.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| strategy | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | d1\_all | 0.00 | 1.00 | 0.00% | > 0.999 |
| 2nd only | d2\_all | 0.04 | 0.98 | 4.20% | = 0.984 |
| Serial search | d1\_2 | 0.00 | 0.98 | 0.00% | = 0.983 |
| Serial search | d2\_3 | 0.22 | 0.91 | 21.90% | = 0.913 |
| Serial search | d3\_4 | 0.45 | 0.80 | 44.70% | = 0.795 |
| Serial search | lin1 | 0.13 | 0.83 | 13.40% | = 0.83 |
| Serial search | lin2 | 0.30 | 0.64 | 30.40% | = 0.641 |
| Serial search | lin3 | 0.13 | 0.69 | 12.90% | = 0.685 |
| Tallying | d1\_2 | 0.14 | 0.90 | 14.40% | = 0.9 |
| Tallying | d1\_3 | 0.04 | 0.96 | 4.00% | = 0.962 |
| Tallying | d1\_4 | 0.00 | 0.99 | 0.00% | = 0.989 |
| Tallying | d2\_3 | 0.39 | 0.75 | 39.20% | = 0.749 |
| Tallying | d2\_4 | 0.21 | 0.90 | 21.10% | = 0.898 |
| Tallying | d3\_4 | 0.46 | 0.77 | 45.50% | = 0.768 |
| Tallying Lazy | d1\_2 | 0.56 | 0.61 | 55.90% | = 0.608 |
| Tallying Lazy | d12\_34 | 0.18 | 0.97 | 17.60% | = 0.974 |
| Tallying Lazy | d3\_4 | 0.38 | 0.90 | 38.40% | = 0.897 |

Table 7:

Follow up comparisons for the proportion of first fixations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| strategy | comparison | scenario | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | d1\_all | A | 0.00 | 1.00 | 0.00% | > 0.999 |
| 1st only | d1\_all | B | 0.00 | 1.00 | 0.00% | > 0.999 |
| 1st only | d1\_all | C | 0.00 | 1.00 | 0.00% | > 0.999 |
| 1st only | d1\_all | D | 0.00 | 1.00 | 0.00% | > 0.999 |
| 2nd only | d2\_all | A | 0.18 | 0.90 | 18.10% | = 0.895 |
| 2nd only | d2\_all | B | 0.20 | 0.88 | 19.90% | = 0.883 |
| 2nd only | d2\_all | C | 0.18 | 0.89 | 17.90% | = 0.891 |
| 2nd only | d2\_all | D | 0.23 | 0.79 | 23.50% | = 0.794 |
| Serial search | d1\_all | A | 0.00 | 1.00 | 0.00% | > 0.999 |
| Serial search | d1\_all | B | 0.00 | 1.00 | 0.00% | > 0.999 |
| Serial search | d1\_all | C | 0.00 | 1.00 | 0.00% | = 0.999 |
| Serial search | d1\_all | D | 0.00 | 1.00 | 0.00% | > 0.999 |
| Tallying | d1\_2 | A | 0.07 | 0.92 | 6.80% | = 0.916 |
| Tallying | d1\_2 | B | 0.12 | 0.81 | 11.80% | = 0.815 |
| Tallying | d1\_2 | C | 0.08 | 0.89 | 8.10% | = 0.891 |
| Tallying | d1\_2 | D | 0.04 | 0.94 | 3.90% | = 0.936 |
| Tallying | d1\_3 | A | 0.06 | 0.92 | 6.00% | = 0.923 |
| Tallying | d1\_3 | B | 0.12 | 0.80 | 12.30% | = 0.804 |
| Tallying | d1\_3 | C | 0.11 | 0.83 | 10.70% | = 0.83 |
| Tallying | d1\_3 | D | 0.08 | 0.88 | 7.80% | = 0.883 |
| Tallying | d1\_4 | A | 0.05 | 0.93 | 5.10% | = 0.934 |
| Tallying | d1\_4 | B | 0.09 | 0.87 | 9.20% | = 0.869 |
| Tallying | d1\_4 | C | 0.09 | 0.86 | 9.30% | = 0.865 |
| Tallying | d1\_4 | D | 0.08 | 0.87 | 8.20% | = 0.869 |
| Tallying | d2\_3 | A | 0.28 | 0.54 | 28.40% | = 0.544 |
| Tallying | d2\_3 | B | 0.26 | 0.50 | 26.20% | = 0.502 |
| Tallying | d2\_3 | C | 0.24 | 0.63 | 23.80% | = 0.63 |
| Tallying | d2\_3 | D | 0.20 | 0.66 | 20.30% | = 0.661 |
| Tallying | d2\_4 | A | 0.26 | 0.61 | 26.20% | = 0.607 |
| Tallying | d2\_4 | B | 0.23 | 0.64 | 23.20% | = 0.643 |
| Tallying | d2\_4 | C | 0.24 | 0.54 | 23.80% | = 0.541 |
| Tallying | d2\_4 | D | 0.18 | 0.65 | 18.30% | = 0.646 |
| Tallying | d3\_4 | A | 0.26 | 0.57 | 26.40% | = 0.567 |
| Tallying | d3\_4 | B | 0.23 | 0.63 | 23.00% | = 0.634 |
| Tallying | d3\_4 | C | 0.23 | 0.57 | 22.60% | = 0.572 |
| Tallying | d3\_4 | D | 0.21 | 0.51 | 21.40% | = 0.506 |
| Tallying Lazy | d1\_2 | A | 0.14 | 0.90 | 14.10% | = 0.898 |
| Tallying Lazy | d1\_2 | B | 0.17 | 0.88 | 16.60% | = 0.88 |
| Tallying Lazy | d1\_2 | C | 0.17 | 0.85 | 17.00% | = 0.851 |
| Tallying Lazy | d1\_2 | D | 0.20 | 0.80 | 20.10% | = 0.804 |
| Tallying Lazy | d12\_34 | A | 0.38 | 0.73 | 37.80% | = 0.729 |
| Tallying Lazy | d12\_34 | B | 0.40 | 0.66 | 40.30% | = 0.663 |
| Tallying Lazy | d12\_34 | C | 0.40 | 0.57 | 39.60% | = 0.566 |
| Tallying Lazy | d12\_34 | D | 0.35 | 0.66 | 34.90% | = 0.656 |
| Tallying Lazy | d3\_4 | A | 0.35 | 0.60 | 34.60% | = 0.6 |
| Tallying Lazy | d3\_4 | B | 0.33 | 0.54 | 32.60% | = 0.536 |
| Tallying Lazy | d3\_4 | C | 0.31 | 0.62 | 30.60% | = 0.623 |
| Tallying Lazy | d3\_4 | D | 0.22 | 0.76 | 21.60% | = 0.762 |

Table 8:

Follow up comparisons for the proportion of fixations across decision scenarios.

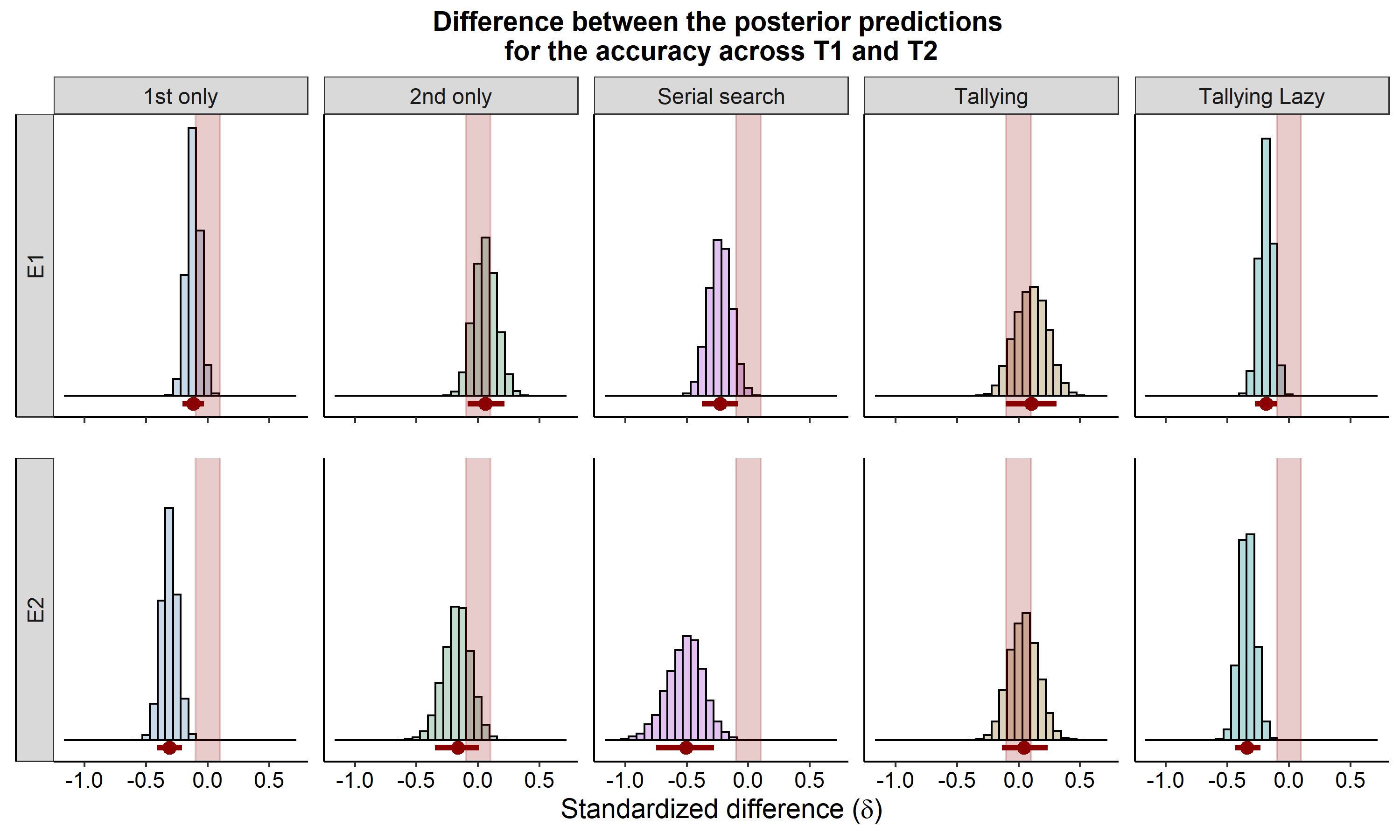
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| strategy | scenario | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | A | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | B | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | C | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | D | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 2nd only | A | d2\_all | 0.00 | 1.00 | 0.000% | = 0.998 |
| 2nd only | B | d2\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 2nd only | C | d2\_all | 0.05 | 0.99 | 4.800% | = 0.991 |
| 2nd only | D | d2\_all | 0.21 | 0.95 | 21.400% | = 0.949 |
| Serial search | A | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| Serial search | B | d1\_2 | 0.14 | 0.91 | 14.000% | = 0.91 |
| Serial search | B | d12\_34 | 0.00 | 1.00 | 0.000% | > 0.999 |
| Serial search | C | d1\_2 | 0.04 | 0.96 | 4.500% | = 0.963 |
| Serial search | C | d1\_3 | 0.00 | 0.99 | 0.000% | = 0.989 |
| Serial search | C | d2\_3 | 0.41 | 0.72 | 41.100% | = 0.717 |
| Serial search | C | d3\_4 | 0.20 | 0.90 | 20.400% | = 0.902 |
| Serial search | D | d1\_2 | 0.00 | 0.98 | 0.000% | = 0.981 |
| Serial search | D | d1\_3 | 0.00 | 0.99 | 0.000% | = 0.993 |
| Serial search | D | d2\_3 | 0.36 | 0.72 | 36.000% | = 0.72 |
| Serial search | D | d3\_4 | 0.46 | 0.57 | 46.400% | = 0.572 |
| Tallying | A | d1\_2 | 0.04 | 0.97 | 4.500% | = 0.969 |
| Tallying | A | d1\_3 | 0.00 | 0.98 | 0.000% | = 0.982 |
| Tallying | A | d1\_4 | 0.00 | 1.00 | 0.000% | = 0.995 |
| Tallying | A | d2\_3 | 0.40 | 0.68 | 40.400% | = 0.684 |
| Tallying | A | d2\_4 | 0.31 | 0.83 | 30.900% | = 0.834 |
| Tallying | A | d3\_4 | 0.48 | 0.69 | 48.300% | = 0.69 |
| Tallying | B | d1\_2 | 0.41 | 0.59 | 40.700% | = 0.587 |
| Tallying | B | d1\_3 | 0.19 | 0.89 | 18.700% | = 0.889 |
| Tallying | B | d1\_4 | 0.10 | 0.94 | 10.200% | = 0.94 |
| Tallying | B | d2\_3 | 0.22 | 0.87 | 21.600% | = 0.873 |
| Tallying | B | d2\_4 | 0.12 | 0.93 | 12.000% | = 0.932 |
| Tallying | B | d3\_4 | 0.50 | 0.61 | 49.700% | = 0.609 |
| Tallying | C | d1\_2 | 0.27 | 0.86 | 26.900% | = 0.858 |
| Tallying | C | d1\_3 | 0.40 | 0.62 | 39.500% | = 0.621 |
| Tallying | C | d1\_4 | 0.13 | 0.94 | 12.800% | = 0.936 |
| Tallying | C | d2\_3 | 0.37 | 0.78 | 37.400% | = 0.777 |
| Tallying | C | d2\_4 | 0.44 | 0.69 | 44.500% | = 0.694 |
| Tallying | C | d3\_4 | 0.21 | 0.91 | 20.700% | = 0.911 |
| Tallying | D | d1\_2 | 0.23 | 0.87 | 22.800% | = 0.87 |
| Tallying | D | d1\_3 | 0.25 | 0.83 | 25.000% | = 0.828 |
| Tallying | D | d1\_4 | 0.41 | 0.64 | 40.600% | = 0.644 |
| Tallying | D | d2\_3 | 0.44 | 0.54 | 44.500% | = 0.537 |
| Tallying | D | d2\_4 | 0.34 | 0.83 | 34.200% | = 0.831 |
| Tallying | D | d3\_4 | 0.34 | 0.78 | 33.900% | = 0.778 |
| Tallying Lazy | A | d1\_2 | 0.46 | 0.84 | 45.900% | = 0.837 |
| Tallying Lazy | A | d12\_34 | 0.00 | 1.00 | 0.000% | > 0.999 |
| Tallying Lazy | A | d3\_4 | 0.64 | 0.70 | 64.500% | = 0.703 |
| Tallying Lazy | B | d1\_2 | 0.42 | 0.85 | 42.500% | = 0.847 |
| Tallying Lazy | B | d12\_34 | 0.00 | 1.00 | 0.000% | = 0.999 |
| Tallying Lazy | B | d3\_4 | 0.70 | 0.59 | 69.800% | = 0.593 |
| Tallying Lazy | C | d1\_2 | 0.72 | 0.57 | 71.500% | = 0.574 |
| Tallying Lazy | C | d12\_34 | 0.43 | 0.93 | 43.100% | = 0.926 |
| Tallying Lazy | C | d3\_4 | 0.23 | 0.95 | 23.100% | = 0.946 |
| Tallying Lazy | D | d1\_2 | 0.68 | 0.54 | 67.900% | = 0.535 |
| Tallying Lazy | D | d12\_34 | 0.60 | 0.84 | 59.500% | = 0.842 |
| Tallying Lazy | D | d3\_4 | 0.02 | 0.99 | 2.000% | = 0.989 |

Table 9:

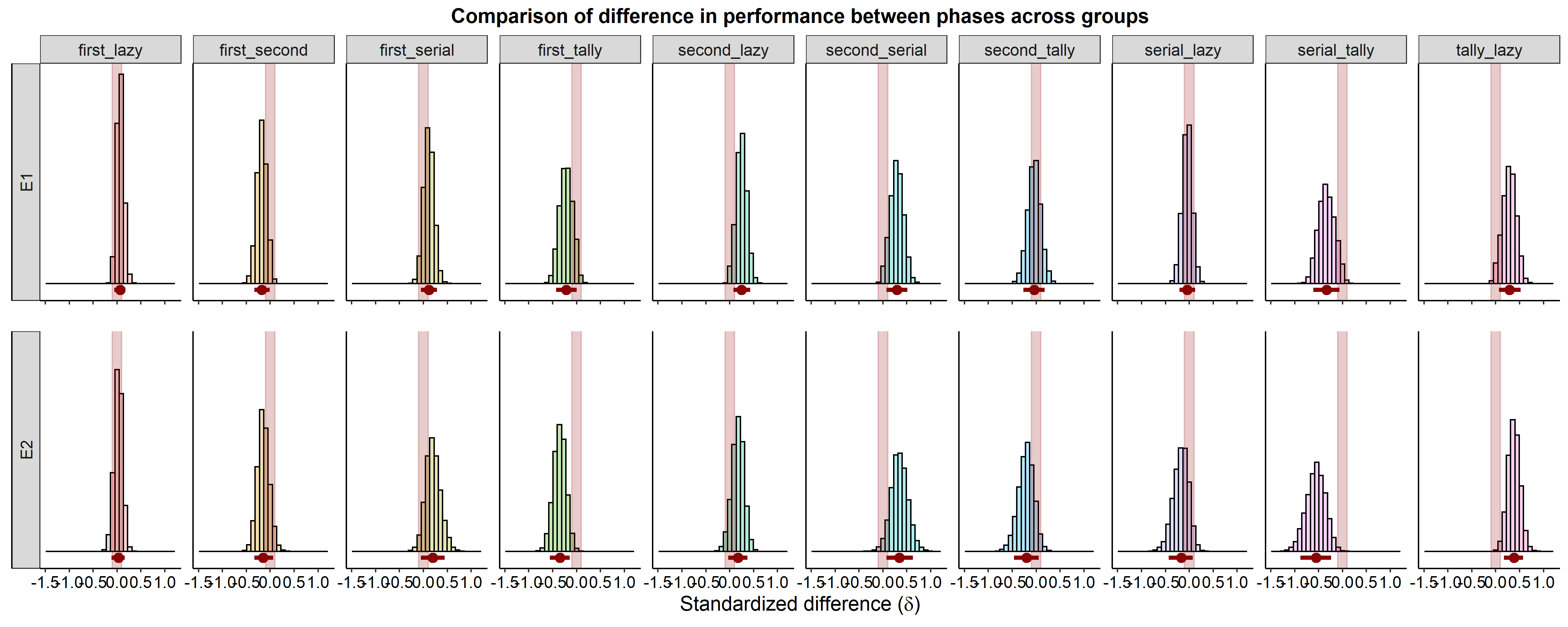
Follow up comparisons for the proportion of last fixations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| strategy | scenario | comparison | ROPE\_Percentage | rawdirection | p\_HDI | p\_direction |
| 1st only | A | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | B | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | C | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 1st only | D | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| 2nd only | A | d2\_all | 0.08 | 0.97 | 7.600% | = 0.973 |
| 2nd only | B | d2\_all | 0.00 | 1.00 | 0.000% | = 0.999 |
| 2nd only | C | d2\_all | 0.40 | 0.80 | 39.800% | = 0.802 |
| 2nd only | D | d2\_all | 0.45 | 0.55 | 44.800% | = 0.549 |
| Serial search | A | d1\_all | 0.00 | 1.00 | 0.000% | > 0.999 |
| Serial search | B | d2\_all | 0.02 | 0.98 | 1.600% | = 0.984 |
| Serial search | C | d3\_all | 0.20 | 0.90 | 20.300% | = 0.896 |
| Serial search | D | d4\_all | 0.20 | 0.87 | 19.800% | = 0.874 |
| Tallying | A | d1\_2 | 0.00 | 0.98 | 0.000% | = 0.982 |
| Tallying | A | d1\_3 | 0.00 | 0.99 | 0.000% | = 0.993 |
| Tallying | A | d1\_4 | 0.00 | 1.00 | 0.000% | = 0.998 |
| Tallying | A | d2\_3 | 0.40 | 0.72 | 39.700% | = 0.719 |
| Tallying | A | d2\_4 | 0.23 | 0.88 | 22.700% | = 0.875 |
| Tallying | A | d3\_4 | 0.42 | 0.75 | 42.300% | = 0.746 |
| Tallying | B | d1\_2 | 0.26 | 0.62 | 26.100% | = 0.62 |
| Tallying | B | d1\_3 | 0.22 | 0.80 | 22.200% | = 0.8 |
| Tallying | B | d1\_4 | 0.10 | 0.92 | 10.100% | = 0.92 |
| Tallying | B | d2\_3 | 0.11 | 0.92 | 11.200% | = 0.915 |
| Tallying | B | d2\_4 | 0.00 | 0.98 | 0.000% | = 0.978 |
| Tallying | B | d3\_4 | 0.34 | 0.77 | 33.600% | = 0.77 |
| Tallying | C | d1\_2 | 0.18 | 0.84 | 17.600% | = 0.839 |
| Tallying | C | d1\_3 | 0.25 | 0.64 | 25.100% | = 0.636 |
| Tallying | C | d1\_4 | 0.14 | 0.90 | 13.900% | = 0.901 |
| Tallying | C | d2\_3 | 0.25 | 0.77 | 25.000% | = 0.774 |
| Tallying | C | d2\_4 | 0.35 | 0.63 | 35.400% | = 0.626 |
| Tallying | C | d3\_4 | 0.19 | 0.86 | 19.400% | = 0.861 |
| Tallying | D | d1\_2 | 0.20 | 0.73 | 19.500% | = 0.732 |
| Tallying | D | d1\_3 | 0.19 | 0.76 | 19.400% | = 0.755 |
| Tallying | D | d1\_4 | 0.24 | 0.64 | 23.600% | = 0.639 |
| Tallying | D | d2\_3 | 0.27 | 0.53 | 26.900% | = 0.532 |
| Tallying | D | d2\_4 | 0.14 | 0.85 | 14.200% | = 0.847 |
| Tallying | D | d3\_4 | 0.13 | 0.87 | 13.200% | = 0.872 |
| Tallying Lazy | A | d1\_2 | 0.42 | 0.76 | 41.500% | = 0.758 |
| Tallying Lazy | A | d12\_34 | 0.01 | 1.00 | 1.200% | = 0.995 |
| Tallying Lazy | A | d3\_4 | 0.33 | 0.90 | 33.100% | = 0.899 |
| Tallying Lazy | B | d1\_2 | 0.03 | 0.98 | 3.400% | = 0.977 |
| Tallying Lazy | B | d12\_34 | 0.17 | 0.96 | 17.300% | = 0.958 |
| Tallying Lazy | B | d3\_4 | 0.23 | 0.92 | 23.000% | = 0.923 |
| Tallying Lazy | C | d1\_2 | 0.33 | 0.80 | 32.900% | = 0.805 |
| Tallying Lazy | C | d12\_34 | 0.59 | 0.69 | 58.900% | = 0.694 |
| Tallying Lazy | C | d3\_4 | 0.04 | 0.98 | 3.500% | = 0.979 |
| Tallying Lazy | D | d1\_2 | 0.34 | 0.74 | 33.700% | = 0.74 |
| Tallying Lazy | D | d12\_34 | 0.37 | 0.84 | 37.100% | = 0.84 |
| Tallying Lazy | D | d3\_4 | 0.00 | 1.00 | 0.000% | > 0.999 |

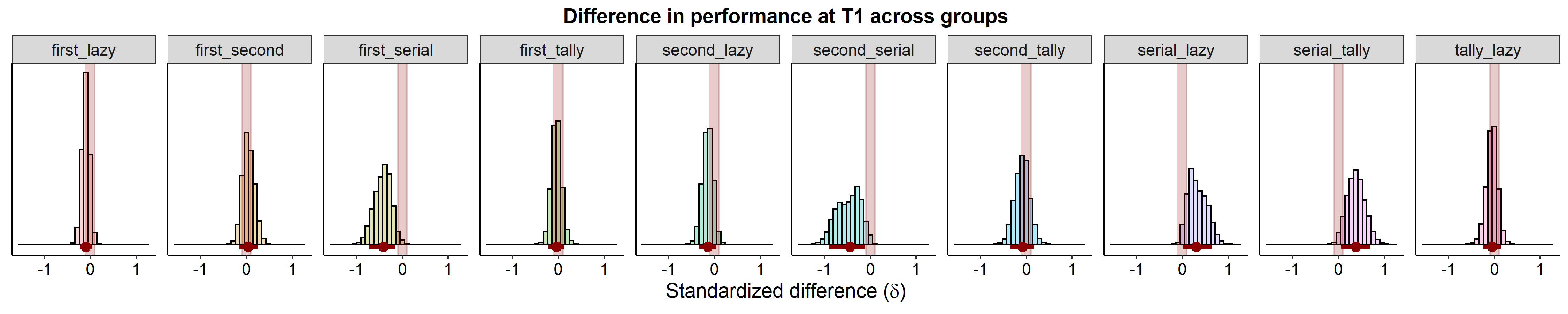
# Appendix



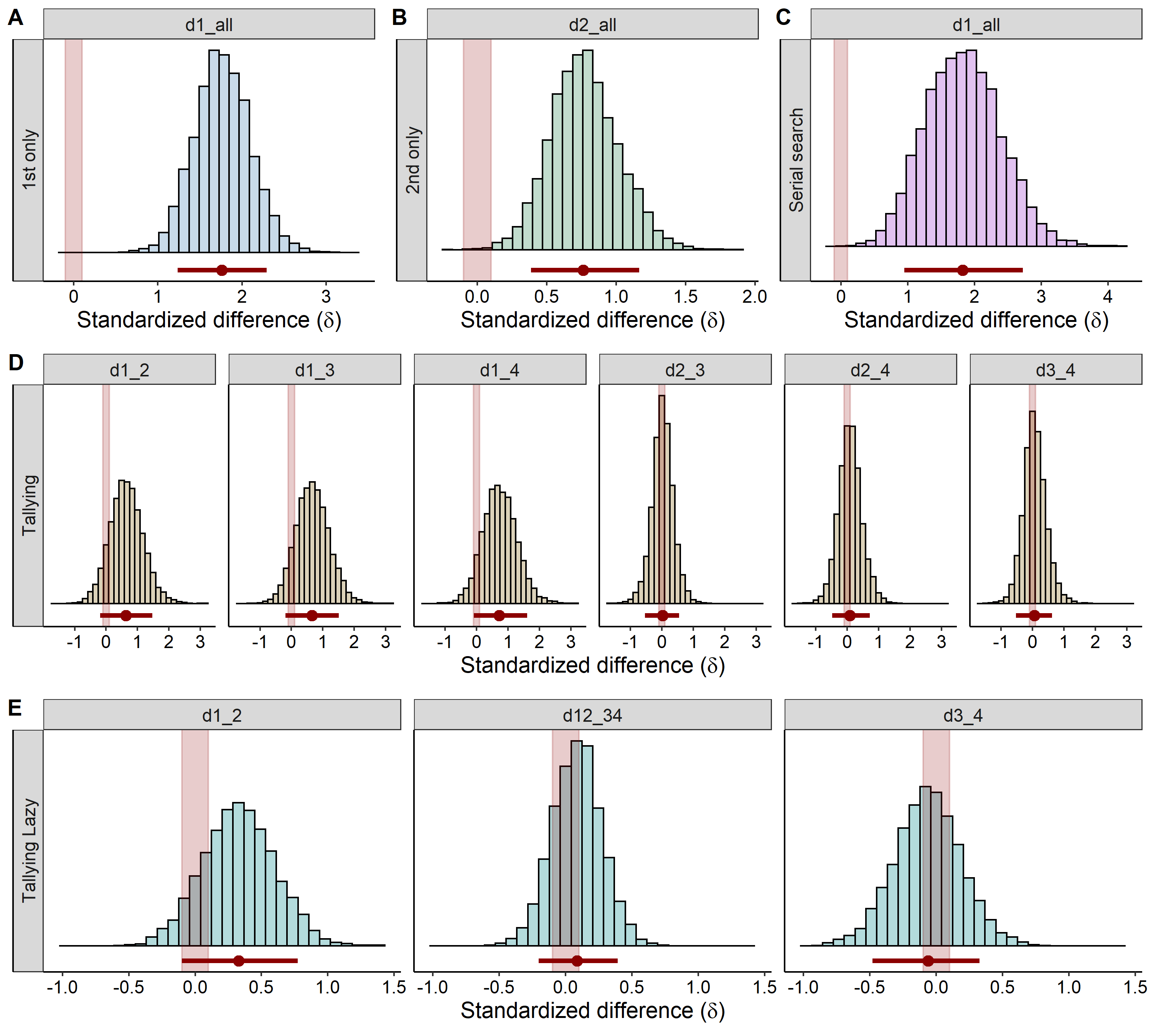
*Figure* *16.*  Difference in performance (as standardized difference) across testing phases for the different decision groups. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons



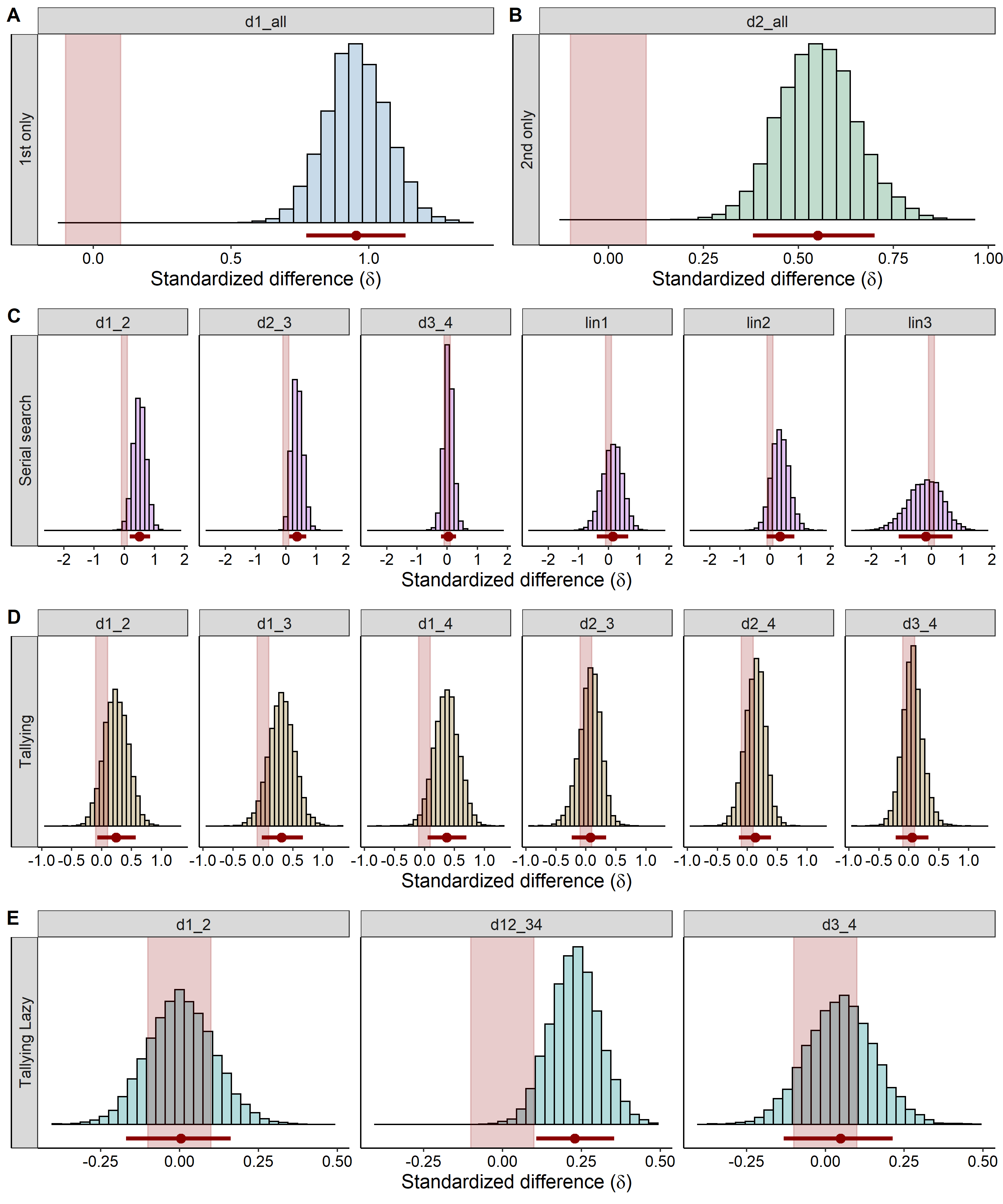
*Figure* *17.*  Paired comparisons between groupds on their performance change across phases. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons



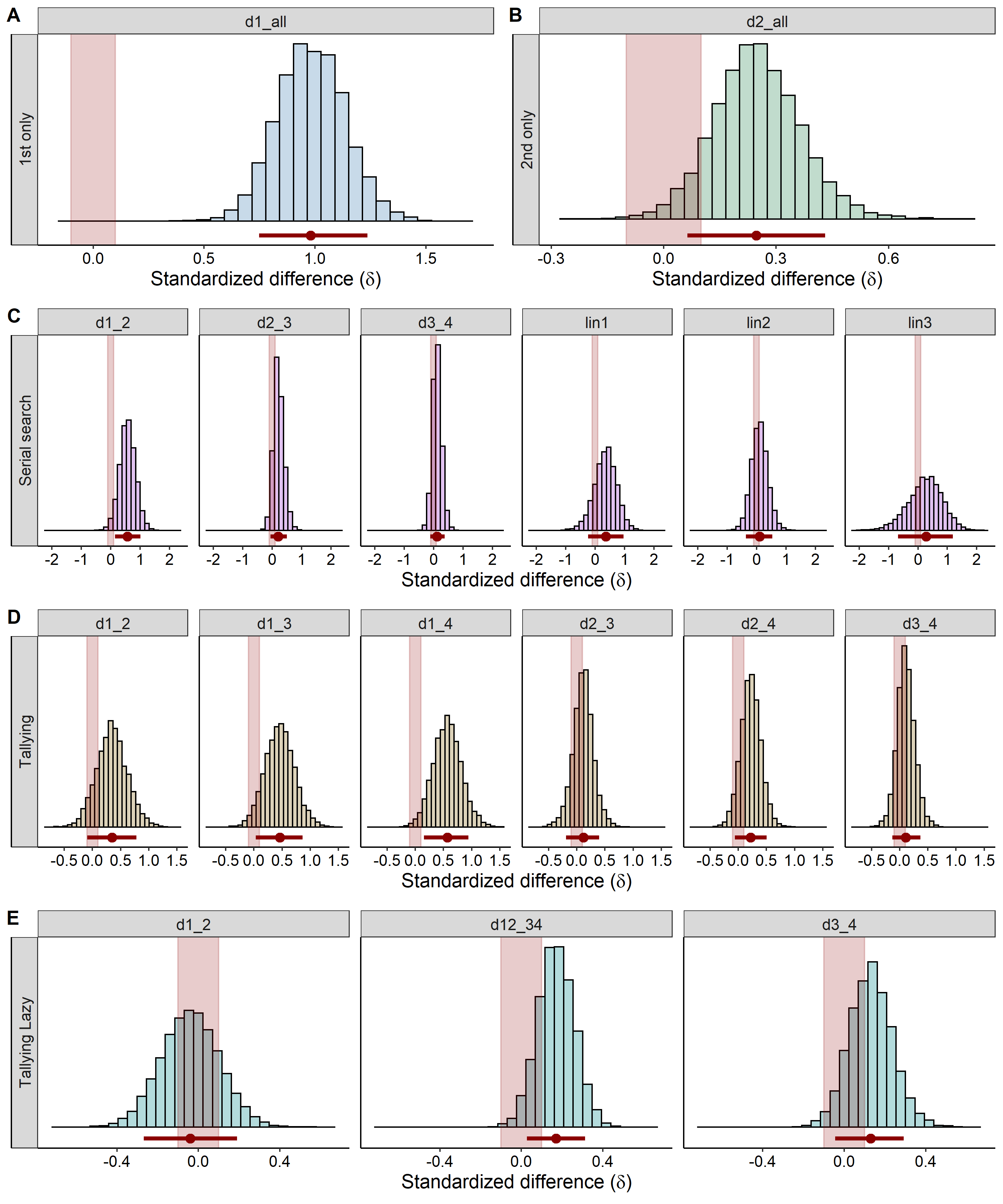
*Figure* *18.*  Difference in performance (as standardized difference) at T1 across decision groups. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons



*Figure* *19.*  Follow up comparisons for the allocation of first fixations, across decision groups. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons



*Figure* *20.*  Follow up comparisons for the allocation of fixations. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons



*Figure* *21.*  Follow up comparisons for the allocation of last fixations. The horizontal red bar represents the 89% HDI. The shaded area highlights the ROPE used in these comparisons