Games and Economic Behavior 148 (2024) 68–81

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

# Failing to utilize potentially effective focal points: Prominence can stymie coordination on distinct actions

Uri Gneezy, Yuval Rottenstreich *

*Rady School of Management, University of California, San Diego, United States*

ARTICLE INFO

Original content: Experiment 3 data.xlsx (Reference data)

*Keywords:*
Coordination
Focal points
Level-k
Team reasoning

ABSTRACT

Are people skillful in utilizing potential focal points? We find a class of situations for which the answer is negative: the presence of prominent actions appears to stymie the use of distinct actions for coordination. Across several experimental games, we consistently observe that players readily coordinate on a categorically distinct action when all available actions are non-prominent but not when some actions are prominent. For instance, given the action set {Franklin Pierce, James Buchanan, Tianjin}, most players select the Chinese city Tianjin. Yet, given {Abraham Lincoln, George Washington, Tianjin}, they are roughly equally likely to select either American president and unlikely to select Tianjin, and given {Abraham Lincoln, George Washington, Shanghai}, their choices are distributed approximately uniformly. The observation that prominence stymies reliance on distinctiveness informs cognitive hierarchy and team reasoning theories of how people recognize focality.

## 1. Introduction

Coordination is essential to many economic interactions (Myerson, 2009). Schelling (1960) emphasized that it can often be achieved if the players involved recognize "some clue … some focal point for each person's expectations of what the other expects him to be expected to do (p. 57)." Building on Schelling's insight, a large experimental and theoretical literature has asked how skillful people are in utilizing potential focal points. This literature has established that people's coordination efforts can evince both impressive abilities and meaningful limitations (e.g., Bardsley, Mehta, Starmer, & Sugden, 2010; Crawford, Gneezy, & Rottenstreich, 2008; Faillo, Smerilli, & Sugden, 2013, 2017; Hargreaves Heap, Rojo Arjona, & Sugden, 2014, 2017). Here, we document a previously-unappreciated limitation. We show that when prominent actions are available, people frequently fail to leverage the potential focality of distinct actions. As we will detail, this finding informs the two leading theories of how people recognize focality.

To lay the groundwork for our argument, consider playing the following "American Presidents and Chinese Cities Game" with one other player and without communication. You and your counterpart will each select an action from the set {James Buchanan, Franklin Pierce, Tianjin} and receive $5 for successful coordination on the same action. None of these actions is especially prominent. Buchanan and Pierce are relatively little-known American presidents, and Tianjin is perhaps not an absolutely leading Chinese city. Thus, considering an action's level of prominence does not identify a potential focal point for coordination. But considering category membership does. Tianjin is clearly distinct from the two presidents, who are relatively similar to one another. Given all this, which action would you select? Not surprisingly, when we conducted this experiment, most participants selected Tianjin. The resulting

* Corresponding author.
*E-mail address:* yuval@rady.ucsd.edu (Y. Rottenstreich).

https://doi.org/10.1016/j.geb.2024.07.010
Received 29 January 2022;
Available online 30 August 2024

coordination rate was relatively substantial.

Now, consider a revamped game in which you and your counterpart each select an action from the set {Abraham Lincoln, George Washington, Shanghai}. These actions are all highly prominent. Abraham Lincoln and George Washington are well-known American presidents; they are eminent historical figures. Shanghai is among the world's leading metropolises. Again, an action's level of prominence does not identify a potential focal point. But, also as before, category membership does. Shanghai is clearly distinct from the two presidents, who are relatively similar to one another. In the revamped game, then, which action would you select? Participants in our experiment did *not* coordinate on Shanghai. Instead, their responses were roughly uniformly distributed. They therefore tended to miscoordinate.

In a third game, having the action set {Abraham Lincoln, George Washington, Tianjin}, our participants also did not coordinate effectively. They were roughly equally likely to select either prominent president and unlikely to select distinct but non-prominent Tianjin.

Our paper is built around results of this form. We study pure coordination games whose action sets include a categorically distinct action. When all available actions are non-prominent, we observe fairly good coordination on the distinct action. In contrast, when multiple actions are prominent, we observe much miscoordination.

To understand such behavior, it is useful to analyze our results via two types of theories of what makes an action focal. Cognitive hierarchy theories stress notions of salience that may often be orthogonal to distinctiveness (Bacharach and Stahl, 2000; Camerer, Ho, and Chong, 2004; Crawford, Costa-Gomes, and Iriberri, 2013; Crawford and Iriberri, 2007; Stahl and Wilson, 1995). In contrast, theories of team reasoning naturally implicate assessments of distinctiveness (Bacharach, 1999; Sugden, 1995; see also Crawford and Haller, 1990). We next detail each type of theory and discuss additional data we collect which speaks to them.

Cognitive hierarchy theories assume that there may be non-strategic players who simply pick an action that is for some reason salient to them. The salience may stem from an action's prominence. It could also reflect liking of an action or myriad other factors. An action that for whatever reason appeals to many non-strategic players is said to have primary salience (Mehta, Starmer, and Sugden, 1994). An action has secondary salience for strategic players who believe others may follow primary salience and accordingly best respond. Higher orders of salience are similarly defined. These forms of salience can drive focality. In the settings we examine, if enough strategic players think in line with secondary or higher order salience and believe some action has sufficient primary salience, coordination on that action ensues.

Team reasoning offers an alternative construction of focality. Players who engage in team reasoning try to generate a rule that, if followed by all of them, will be to the advantage of each of them. In the settings we examine, selecting the categorically distinct action is perhaps the most natural such rule.

To help assess the role of cognitive hierarchies in our experiments, we collect measures concerning primary and secondary salience. Using a method developed by Mehta, Starmer, and Sugden (1994; see also Bardsley, Mehta, Starmer, and Sugden, 2010), we have participants initially pick whichever element from a set they wish to pick; then, we offer these participants $5 for correctly guessing what a counterpart picked. The picking task provides a measure of primary salience, and the guessing task provides a measure of secondary salience.

To help assess thinking in line with team reasoning, we collect measures concerning the identification of distinct actions. In particular, we ask a single initial participant to select the element of the set he or she believes is different from the others; then, we offer additional participants $5 if both they and a counterpart select the action that has been deemed different. Responses to this "search-for-the-different" task provide a measure of participants' ability to identify a distinct action when explicitly directed to do so.

Our data on primary salience, secondary salience, and identification of distinctiveness jointly suggest that with sets composed entirely of non-prominent actions, people invoke team reasoning, but given prominent actions people attempt to coordinate on the basis of primary or higher orders of salience. With {James Buchanan, Franklin Pierce, Tianjin}, we observe that play of the coordination game is intermediate between guesses and an explicit search-for-the-different. It statistically departs from both. It appears that many players ably team reason about distinctiveness and thereby coordinate successfully. In contrast, with {Abraham Lincoln, George Washington, Shanghai}, play of the coordination game closely matches guesses, which are distributed relatively evenly over all available actions, and it is quite far from an explicit search-for-the-different. That is, coordination attempts are consistent with players following their appraisals of primary or higher orders of salience rather than team reasoning. Consequently, players tend to miscoordinate because while team reasoning about distinctiveness can isolate a unique focal point in the set {Abraham Lincoln, George Washington, Shanghai}, thinking in line with cognitive hierarchies cannot.

In sum, we suggest that absent prominent actions, people ably team reason about distinctiveness. However, given prominent actions, they largely think in line with cognitive hierarchies. They follow secondary and higher orders of salience even if doing so cannot facilitate coordination while team reasoning about distinctiveness can.

### *1.1. Relationship to Prior Research*

Our results build on important research by Bardsley, Mehta, Starmer, and Sugden (2010) and Hargeaves Heaps, Rojo Arjona, and Sugden (2014). These researchers also examined action sets having a distinct element. Bardsley at al.'s work is especially relevant to ours. Within broader studies, they examined action sets composed of a distinct, non-prominent element placed alongside similar, prominent elements, for example {Mannheim, Berlin, Brussels, Lisbon, Madrid} and {Calais, Berlin, Paris, Prague, Rome}. Furthermore, they implemented pick, guess, and coordination treatments though not a search-for-the-different treatment.

Interestingly, Bardsley et al.'s perspective differs from ours: they lauded people's skill in leveraging potential focal points. To appreciate Bardsley et al.'s perspective and why it departs from ours, we note that a cross-country comparison was integral to their

empirics. They studied some action sets in the Netherlands and others in England. In the Netherlands, players appeared to invoke team reasoning about distinctiveness, while in England players' behavior was consistent with cognitive hierarchies and salience. For instance, Dutch participants encountered {Mannheim, Berlin, Brussels, Lisbon, Madrid}, and relative to their guesses (17%), their coordination attempts were substantially more concentrated on the distinct Mannheim (44%). On the other hand, English participants encountered {Calais, Berlin, Paris, Prague, Rome}, and both their guesses (51%) and coordination attempts (61%) concentrated on Paris.

Based on evidence that each location showed relatively successful coordination – yet via differing strategies – Bardsley et al. suggested that players savvily appreciate whether circumstances call for thinking in line with cognitive hierarchies or team reasoning. They wrote:

> "Within each [location], a large majority … used a common mode … for identifying focal points … This behavior reveals an ability to solve coordination problems at a conceptual level above that of the theories of cognitive hierarchy and team reasoning (p. 77)."

Cross-country comparison is an intriguing methodology, and Bardsley et al.'s findings reflect an impressive ability to leverage focal points (for another cross-country comparison see Experiment 3 in Sontuoso & Bhatia, 2021). Why do we stress a less sanguine perspective of people's skill in leveraging potential focal points? Because our methodology generates two additional types of comparisons, and these comparisons cast participants' coordination efforts in a less positive light.

First, unlike Bardsley et al.'s studies, our experiments systematically manipulate prominence and distinctiveness. For example, Bardsley et al. consider play of actions sets such as {Mannheim, Berlin, Brussels, Lisbon, Madrid} and {Calais, Berlin, Paris, Prague, Rome}, which include a lone non-prominent town alongside several prominent locales. But they do not contrast such action sets with action sets composed of the same distinct, non-prominent town alongside non-prominent locales that differ from it and are similar to one another. To see the implications, suppose that with a revised action set such as {Mannheim, Nottingham, Newcastle, Hull, Derby}, a large proportion of participants attempted to coordinate on Mannheim, perhaps meaningfully more than the 44% that attempt to coordinate on it in the original set. Such a result would cast players' skill in leveraging distinctiveness given prominent actions in a less positive light. But that is precisely the sort of result we obtain in each of our experiments, for instance with {Tianjin, James Buchanan, Franklin Pierce} versus {Tianjin, Abraham Lincoln, George Washington}.

Second, as we have mentioned, unlike Bardsley et al.'s studies, our experiments include an explicit search-for-the-different treatment. Suppose that in this treatment a large proportion of participants would tap Mannheim from the original action set {Mannheim, Berlin, Brussels, Lisbon, Madrid}, perhaps meaningfully more than 44%. That result would also cast players' skill in using distinctiveness in a less positive light. And, again, that is precisely the sort of result we obtain. For instance, when alongside Abraham Lincoln and George Washington, Tianjin is selected much more often in our search-for-the-different treatment than coordination treatment.

In sum, our argument is not that players do not invoke team reasoning in the presence of prominent actions. Rather, it is that they invoke it relatively less than might be expected, in two specific ways: relative to the absence of non-prominent actions and relative to an explicit search-for-the-different. To corroborate this argument, we implement a methodology that complements cross-country comparison. Systematically varying prominence and distinctiveness allows us to thoroughly assess the influence of these factors. Examining an explicit search-for-the-different is likewise instructive.

Our experiments thereby yield data that are consistent with past work. But they include new comparisons that lead to a less sanguine perspective on people's skill in leveraging potential focal points. These comparisons indicate that prominence stymies the use of distinctiveness for coordination. In our concluding discussion, we consider how a full view of people's coordination abilities must integrate dual perspectives. As Bardsely et al. show, people can evince an impressive ability to leverage potential focal points. As we show, people can also evince meaningful limitations on this ability.

## 2. Experiments

Both of our first two experiments include coordination, explicit search-for-the-different, and pick-then-guess treatments. These treatments were conducted between-subjects. Coordination and search-for-the-different participants encountered just one game. Pick-then-guess participants encountered one game from Experiment 1 or Experiment 2 along with additional unrelated games.

Participants were UC San Diego undergraduates enrolled in business courses having an experimental requirement. They came to the behavioral lab at the business school and took part in sessions of between nine to sixteen participants.

Each session was devoted to one treatment. Every participant was anonymously matched to a counterpart from the same session and paid in cash on the basis of the choices made by the two of them. We offered participants $5 if they and their counterpart selected the same action in the coordination game, did so in the explicit search-for-the-different task, or provided a correct guess in the guessing task.

To eliminate serial position of actions as a coordination device (e.g., having players attempt to coordinate on the first action), we implemented a "card-based" method in all treatments. Each participant received face-down index cards and an instruction/response sheet (see Appendix for all instructions). An experimenter read the instructions aloud. These instructions indicated that every participant received cards listing the same entities. Participants were asked to shuffle their cards for fifteen seconds before turning them over and reviewing them. At that point, they selected one of the cards. This method made clear that the order in which the actions were laid out for a player need not match the order in which they were laid out for their counterpart.

### 2.1. Experiment 1: American Presidents and Chinese Cities

We created four "American Presidents and Chinese Cities" action sets, each of which included a pair of presidents alongside a single Chinese city. As Table 1 indicates, two sets were "unmixed": composed of only prominent actions {Abraham Lincoln, George Washington, Shanghai} or only non-prominent actions {James Buchanan, Franklin Pierce, Tianjin}. Two sets were "mixed": composed of prominent presidents and a non-prominent Chinese city {Abraham Lincoln, George Washington, Tianjin} or non-prominent presidents and a prominent Chinese city {James Buchanan, Franklin Pierce, Shanghai}. The cards participants received showed both the first and last name of each president, to make plain that the reference was to the person rather than a namesake town or city (such as Lincoln, Nebraska or Washington, D.C.).

Unmixed sets allow for a test of a weak version of our hypothesis: that encountering a set composed entirely of prominent actions stymies use of distinctiveness and engenders reliance on primary or higher orders of salience. The mixed set that includes prominent Abraham Lincoln and George Washington and non-prominent Tianjin allows for a test of a strong version of our hypothesis: that encountering even a subset of prominent actions stymies the use of distinctiveness and engenders reliance on primary or higher orders of salience. We include the mixed set containing Buchanan, Pierce, and Shanghai for completeness; with this set, thinking in line with cognitive hierarchies cannot be differentiated from team reasoning about distinctiveness (both lead to choices of Shanghai).

Following Bardsley, Mehta, Starmer, and Sugden (2010) for each treatment we define a "normalized coordination index," denoted $c^*$. Let $m_j$ be the number of participants who select action $j$, $n$ be the total number of participants, and $A = 3$ be the number of available actions. The probability that two participants chosen at random without replacement select the same action is given by $\sum[m_j(m_j - 1)]/[n(n-1)]$. The probability that two participants select the same action when one of them plays randomly is $1/A$. The normalized coordination index $c^*$ is the ratio of these terms, $c^* = A\sum[m_j(m_j - 1)]/[n(n-1))]$. It equals one when participants' selections are no more concentrated than chance. It rises as their selections become more concentrated.

Table 1 displays our results. For convenience, cells containing information about distinct actions are in grey. The unmixed sets, {Abraham Lincoln, George Washington, Shanghai} and {James Buchanan, Franklin Pierce, Tianjin} yield the following three-fold pattern. (1) Guesses are distributed roughly evenly in either set; $c^*$ equals 1.14 for the prominent actions and 1.07 for the non-prominent actions. (2) Players show marked success in an explicit search-for-the-different in either set. They readily identify both lone prominent Shanghai and lone non-prominent Tianjin as distinct; $c^*$ equals 2.60 and 2.27, respectively. (3) Nevertheless, coordination attempts differ vastly across the two sets. For the non-prominent actions, play of the coordination games is intermediate between guesses and an explicit search-for-the-different. Players thus coordinate decently well; $c^*$ equals 1.41. On the other hand, for the prominent actions, coordination attempts resemble guesses in being roughly uniform. Players tend to miscoordinate; $c^*$ equals only 0.98. This three-fold pattern is in line with a meaningful fraction of participants attempting to coordinate via team reasoning about distinctiveness given non-prominent actions but via cognitive hierarchy thinking and salience given prominent actions.

**Table 1**
Experiment 1 Results. Numbers in each president and city row are the percentage of players selecting that response. Distinctive actions within each action set appear in bolded italics. Unmixed action sets include only prominent or only non-prominent actions; mixed actions sets contain both. $c^*$ is a normalized concentration measure. $p$-values in rows labeled $c^*$ compare $c^*$ in the relevant treatment with the treatment to its immediate left. $p$-values in rows labeled "Comparing …" indicate whether the two action sets above yield statistically distinct values of $c^*$ in the relevant treatment.

| | | PICK | GUESS | COORDINATION | SEARCH-FOR-THE-DIFFERENT |
|---|---|---|---|---|---|
| **UNMIXED ACTION SETS** | ***Shanghai*** | .40 | .51 | .38 | .92 |
| | Washington | .29 | .21 | .29 | .04 |
| | Lincoln | .30 | .28 | .33 | .04 |
| | *n* | 109 | | 58 | 58 |
| | $c^*$ | 1.00 | 1.14 | 0.98, *ns* | 2.60, $p < .01$ |
| | Buchanan | .28 | .21 | .21 | .05 |
| | Pierce | .25 | .33 | .15 | .08 |
| | ***Tianjin*** | .48 | .46 | .64 | .87 |
| | *n* | 80 | | 75 | 38 |
| | $c^*$ | 1.07 | 1.07 | 1.41, $p < .01$ | 2.27, $p < .01$ |
| | Comparing unmixed action sets | ns | *ns* | $p < .01$ | *ns* |
| **MIXED ACTION SETS** | ***Shanghai*** | .71 | .76 | .74 | .94 |
| | Buchanan | .17 | .17 | .18 | .03 |
| | Pierce | .13 | .07 | .08 | .03 |
| | *n* | 72 | | 77 | 35 |
| | $c^*$ | 1.62 | 1.83 | 1.75, *ns* | 1.83, *ns* |
| | Washington | .31 | .35 | .40 | .08 |
| | Lincoln | .42 | .49 | .37 | .12 |
| | ***Tianjin*** | .27 | .16 | .22 | .80 |
| | *n* | 71 | | 89 | 59 |
| | $c^*$ | 1.01 | 1.15 | 1.03, *ns* | 1.95, $p < .01$ |
| | Comparing mixed action sets | $p < .01$ | $p < .01$ | $p < .01$ | $p < .05$ |

The mixed sets {James Buchanan, Franklin Pierce, Shanghai} and {George Washington, Abraham Lincoln, Tianjin}, yield a closely related three-fold pattern. (1) Guesses concentrate on lone prominent Shanghai, $c^*$ equals 1.83, but are distributed more evenly given the prominent pair Washington and Lincoln, $c^*$ equals 1.15. (2) In an explicit search-for-the-different, participants readily identify both lone prominent Shanghai and lone non-prominent Tianjin as distinct; $c^*$ equals 2.66 and 1.95, respectively. (3) Nevertheless, for {George Washington, Abraham Lincoln, Tianjin} play of the coordination games more closely resembles guesses than an explicit search-for-the-different; players do not coordinate well, $c^*$ equals 1.03. Meanwhile, players do coordinate fairly well given {James Buchanan, Franklin Pierce, Shanghai}, $c^*$ equals 1.75. This three-fold pattern is in line with the strong version of our hypothesis: Given one or more prominent actions people appear to be largely stymied from team reasoning about distinctiveness. They act in accord with cognitive hierarchy theories and salience.

A quartet of statistical analyses corroborate the foregoing patterns and the conclusions we draw from them. First, and most simply, we can consider the percentage of players who attempt to coordinate on the distinct action. The 64% who select less prominent Tianjin when it is alongside the less prominent presidents is statistically different from the 38% who select more prominent Shanghai when it is alongside the more prominent presidents ($p = .005$ by a two-sample test for equality of proportions with continuity correction). Furthermore, the 74% who select Shanghai when it is alongside the less prominent presidents is statistically distinct from the 22% who select Tianjin when it is alongside the more prominent presidents, ($p < .0001$).

Second, in each treatment, we can consider whether responses are more concentrated in one mixed action set than the other and one unmixed action set than the other. To do so, we use a bootstrap method to assess the null hypothesis that responses are equally concentrated. For a given pair of action sets, the method runs as follows. Take one of the sets and, using the sample size it has in our experiment and the observed relative frequencies it yields, draw a simulated sample of responses. Calculate the resulting $c^*$. Perform the same steps for the other action set. Then, calculate the difference between the two $c^*$ measures. Repeat this process 20,000 times. Finally, derive a $p$-value by comparing critical values in the resulting distribution of $c^*$ differences to zero. In Table 1, the row labeled "Comparing unmixed action sets" contains these $p$-values. The unmixed action sets yield statistically indistinguishable concentrations of responses in the pick, guess, and search-for-the-different treatments. Yet, in the coordination treatment, {James Buchanan, Franklin Pierce, Tianjin} shows statistically greater concentration than {Abraham Lincoln, George Washington, Shanghai}. Meanwhile, the mixed action sets yield statistically distinguishable concentrations in every treatment.

Third, we can consider whether for a given action set, responses in a target treatment are more concentrated than responses in a reference treatment. To do so, we use a bootstrap method to assess the null hypothesis that target treatment responses are drawn from a distribution reflecting the relative frequencies observed in the reference treatment. For a given target treatment, the method runs as follows. Construct 20,000 simulated samples using the participant count in the target treatment and the observed relative frequencies in the reference treatment. Calculate the $c^*$ measure associated with each simulated sample. Then, compare the observed $c^*$ from the target treatment to critical values from the distribution of 20,000 $c^*$ measures. In Table 1, $p$-values located within rows labeled $c^*$ indicate whether the $c^*$ in the relevant cell is statistically greater than the $c^*$ to the immediate left. Consistent with our arguments, the only action set for which coordination attempts are more concentrated than guesses is {James Buchanan, Franklin Pierce, Tianjin}. Furthermore, reflecting stymied team reasoning, coordination attempts are generally less concentrated than search-for-the-different responses; indeed, even with {James Buchanan, Franklin Pierce, Tianjin} there is less than full leveraging of the distinct action.

Fourth, we can consider whether for a given action set, the extent of the match between coordination attempts and search-for-the-different responses is greater than the extent of the match between coordination attempts and guesses. To do so, we again use a bootstrap method. It runs as follows. Using the observed relative frequencies, for each action set draw 20,000 coordination attempts, search-for-the-different responses, and guesses. For each draw, code a match between two treatments as a 1 and a mismatch as a zero. Then, compare the proportions of matches. Consistent with our arguments, coordination/search-for-the-different matches are reliably more common than coordination/guess matches for {James Buchanan, Franklin Pierce, Tianjin} but not {Abraham Lincoln, George Washington, Shanghai}, $p < .01$ and *ns* respectively. They are also reliably more common for {James Buchanan, Franklin Pierce, Shanghai}, $p < .01$, but are directionally less common for {Abraham Lincoln, George Washington, Tianjin}.

All four analyses lend credence to the hypothesis that absent prominent actions, team reasoning about distinctiveness readily emerges, but given prominent actions such thinking is largely stymied in favor of thinking in line with cognitive hierarchies and salience.

Finally, we wished to corroborate the intuition that Shanghai, Abraham Lincoln, and George Washington are relatively prominent, while Tianjin, James Buchanan, and Franklin Pierce are not. Accordingly, in later sessions with new participants from the same lab, we collected data which indicate that our players (a) mostly agree that, by pertinent definitions, Shanghai, Abraham Lincoln, and George Washington are prominent but Tianjin, James Buchanan, and Franklin Pierce are not, (b) by default classify the Chinse cities and U.S. Presidents as prominent or non-prominent in a manner that is roughly consistent with these definitions, and (c) to a meaningful extent expect that others will as well.

Participants were presented with every action from Experiment 1: both Chinese cities and all four U.S. Presidents. First, they classified each action as prominent or not prominent. Second, they were anonymously paired with another participant and predicted that person's classifications. The actions were presented three at a time, in either the two unmixed or two mixed action sets. The order in which the action sets were presented and the order in which actions were listed within a set were determined randomly for each participant.

We considered both a dictionary definition of prominence and a definition taken from research on coordination. In an initial run of sessions, roughly half of the participants were provided with the Merriam-Webster (2023) definition, which characterizes an entity as prominent if it is "widely and popularly known" in the relevant context. These participants were instructed to classify actions as prominent or not prominent on the basis of this definition. They were then asked to predict how another participant would classify

each action by this definition. The other half of participants from this initial run were not provided with any definition, so that we could assess their default notions of prominence. These participants classified each action with no explicit instruction as to what constitutes prominence and likewise predicted a counterpart's classifications. Participants in both the Merriam-Webster and default groups were paid $5 if all their predictions were correct.

In a subsequent run of sessions we turned to a complementary definition invoked by Hargreaves Heap, Rojo Arjona, and Sugden (2017). These authors characterized prominence as "surpassing most others in its category or especially distinguished." Roughly half of the participants in this run were instructed to rely on Hargreaves Heap's et al. (2017) definition, while the other half were not provided with any definition. Participants in both groups were paid $1 if all their predictions were correct.

Table 2A summarizes the results of the initial run, while Table 2B summarizes the results of the subsequent run. The same pattern emerges each time. In both tables, the left-most column of numbers displays the percentage of participants classifying an action as prominent. Per (b), responses with a definition (top) and without a definition (bottom) showed substantial agreement. Per (a), overwhelming majorities classified Abraham Lincoln, George Washington, and Shanghai as prominent, but substantially smaller percentages classified James Buchanan, Franklin Pierce, and Tianjin as prominent.

Per (c), participants' predictions regarding Shanghai, Abraham Lincoln, and George Washington overwhelmingly matched their own classifications. Across these three actions, given the Merriam-Webster definition, only 5% of predictions differed from classifications, and absent a definition only 9% differed. The corresponding figures based on the Hargreaves Heap et al. definition were similar, 9% and 12%, respectively. Participants' predictions regarding Tianjin, James Buchanan, and Franklin Pierce also tended to match their own classifications, though differences were more common. Given the Merriam-Webster definition, 26% of predictions differed from a participant's own selections, and absent a definition 23% differed. The corresponding figures based on the Hargreaves Heap et al. definition were again similar, 22% and 23%, respectively.

For statistical tests concerning observations (a), (b), and (c) we fit logistic regressions in which the dependent variable was the Phase 1 classification. We also fit logistic regressions and in which the dependent variable was whether the Phase 1 classification agreed with the Phase 2 prediction. We fit separate regressions for the initial run, involving the Merriam-Webster definition, and the subsequent run, involving the Hargreaves Heap et al. definition. In every regression, the independent variables included dummies for the triple from which the action came (Shanghai, Abraham Lincoln, George Washington versus Tianjin, James Buchanan, Franklin Pierce), whether a definition was provided, and whether mixed or unmixed action sets were presented. In every regression, the only significant variable ($p < .0001$) was the triple from which the action came.

In sum, whether or not they are provided with a definition of prominence, and irrespective of whether that definition characterizes prominence as "widely and popularly known" or as "surpassing most others in its category," almost all of our participants agreed that Shanghai, Abraham Lincoln, and George Washington are prominent. A substantial fraction also perceive consensus about the non-prominence of Tianjin, James Buchanan, and Franklin Pierce. We suggest that while neither definition is complete or authoritative, together they provide a viable characterization of what constitutes prominence.

**Table 2a**

Results for the follow-up to Experiment 1. Participants either were (top) or were not (bottom) provided with a definition of prominent as "widely and popularly known." In each row, the left-most number shows the percentage who classified an action as prominent rather than non-prominent. The following numbers summarize the relationship between participants' own classifications and their predictions of others' classifications. For instance, 5% of participants provided with a definition classified Shanghai as prominent and predicted that others would classify it as non-prominent.

| WITH A DICTIONARY DEFINITION (n = 87) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Own Classification | Own Classification → Predict Others' | | | | | |
| | % Prominent | NP → P | NP → NP | P → P | P → NP | % Same in Both | % Different |
| Shanghai | .99 | .00 | .01 | .94 | .05 | .95 | .05 |
| Lincoln | .98 | .01 | .01 | .91 | .07 | .92 | .08 |
| Washington | .97 | .00 | .03 | .94 | .02 | .98 | .02 |
| *average* | *.98* | *.00* | *.02* | *.93* | *.05* | *.95* | *.05* |
| Tianjin | .43 | .11 | .46 | .24 | .18 | .70 | .30 |
| Buchanan | .18 | .21 | .61 | .13 | .06 | .74 | .26 |
| Pierce | .32 | .11 | .56 | .22 | .10 | .78 | .22 |
| *average* | *.31* | *.15* | *.54* | *.20* | *.11* | *.74* | *.26* |

| WITHOUT A DEFINITION (n = 100) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Own Classification | Own Classification → Predict Others' | | | | | |
| | % Prominent | NP → P | NP → NP | P → P | P → NP | % Same in Both | % Different |
| Shanghai | .96 | .03 | .01 | .87 | .09 | .88 | .12 |
| Lincoln | .90 | .03 | .07 | .85 | .05 | .92 | .08 |
| Washington | .95 | .02 | .03 | .91 | .04 | .94 | .06 |
| *average* | *.94* | *.03* | *.04* | *.88* | *.06* | *.91* | *.09* |
| Tianjin | .49 | .11 | .40 | .24 | .25 | .64 | .36 |
| Buchanan | .26 | .10 | .64 | .16 | .10 | .80 | .20 |
| Pierce | .42 | .05 | .53 | .35 | .07 | .88 | .12 |
| *average* | *.39* | *.09* | *.52* | *.25* | *.14* | *.77* | *.23* |

**Table 2b**
Results for the follow-up to Experiment 1. Participants either were (top) or were not (bottom) provided with Hargreaves Heap et al.'s (2017) definition of prominent as "surpassing most others in its category or especially distinguished." In each row, the left-most number shows the percentage who classified an action as prominent rather than non-prominent. The following numbers summarize the relationship between participants' own classifications and their predictions of others' classifications. For instance, 4% of participants provided with a definition classified Shanghai as prominent and predicted that others would classify it as non-prominent.

| WITH THE HARGEAVES HEAP ET AL. DEFINITION (n = 135) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Own Classification | Own Classification → Predict Others' | | | | | |
| | % Prominent | NP → P | NP → NP | P → P | P → NP | % Same in Both | % Different |
| Shanghai | .89 | .04 | .07 | .84 | .04 | .91 | .09 |
| Lincoln | .96 | .04 | .01 | .90 | .06 | .91 | .10 |
| Washington | .95 | .03 | .02 | .90 | .05 | .92 | .08 |
| *average* | *.93* | *.04* | *.03* | *.88* | *.05* | *.91* | *.09* |
| Tianjin | .34 | .10 | .56 | .14 | .20 | .70 | .30 |
| Buchanan | .21 | .05 | .73 | .09 | .13 | .82 | .18 |
| Pierce | .19 | .09 | .73 | .10 | .09 | .83 | .17 |
| *average* | *.25* | *.08* | *.67* | *.11* | *.14* | *.78* | *.22* |

| WITHOUT A DEFINITION (n = 161) | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Own Classification | Own Classification → Predict Others' | | | | | |
| | % Prominent | NP → P | NP → NP | P → P | P → NP | % Same in Both | % Different |
| Shanghai | .92 | .04 | .04 | .80 | .12 | .84 | .16 |
| Lincoln | .92 | .05 | .03 | .86 | .06 | .89 | .11 |
| Washington | .96 | .02 | .02 | .89 | .07 | .91 | .09 |
| *average* | *.93* | *.04* | *.03* | *.85* | *.08* | *.88* | *.12* |
| Tianjin | .34 | .12 | .54 | .17 | .17 | .71 | .29 |
| Buchanan | .17 | .14 | .69 | .10 | .07 | .79 | .21 |
| Pierce | .22 | .10 | .68 | .14 | .08 | .82 | .18 |
| *average* | *.24* | *.12* | *.64* | *.14* | *.11* | *.77* | *.23* |

### 2.2. *Experiment 2: A Gentleman Among Ladies*

To examine the generality of our results, we studied a different family of games in which action sets were composed of profile pictures. Each set included profiles of three women and one man. The women were either all celebrities or all non-celebrities, and the man was either a celebrity or a non-celebrity. See Table 3; for privacy concerns, non-celebrities' pictures are blurred in the table but were of course shown in a clear format in the experiment.

By crossing the celebrity status of the women and man we obtained four action sets. In every set, gender distinguished one action from the rest. In some sets, the same action was also distinguished from the others by its celebrity/non-celebrity status. We thus had two unmixed sets: four celebrities, one of whom was a man and four non-celebrities, one of whom was a man. We also had two mixed sets: three female celebrities alongside a male non-celebrity and three female non-celebrities alongside a male celebrity. We considered each of the celebrities to be prominent, in the sense of "widely and popularly known."

Table 3 displays our results. The results are consistent with Experiment 1. It again appears that in the absence of prominent actions, people ably team reason about distinctiveness. Yet given prominent actions, they are stymied from doing so and engage in thinking that reflects cognitive hierarchies and appraisals of salience – even if such thinking cannot identify a unique focal but team reasoning about distinctiveness can.

As before, a quartet of statistical analyses support our argument. First, and most simply, we compare the percentage of participants who attempt to coordinate on the distinct action in one unmixed set versus the other and in one mixed set versus the other. The 57% who select the lone male among four non-celebrities is statistically different from the 35% who select the lone male among four celebrities, ($p < .001$ by a two-sample test for equality of proportions with continuity correction). The 90% who select the male celebrity placed alongside three female non-celebrities is also statistically different from the 25% who select the male non-celebrity alongside three female celebrities, ($p < .0001$).

Second, we use a bootstrap method to assess the null hypothesis that in a given treatment, responses in the two mixed actions sets are equally concentrated. We likewise assess the parallel null hypothesis for unmixed sets. In Table 3, the row labeled "Comparing unmixed action sets" contains the relevant *p*-values. The unmixed action sets yield statistically indistinguishable concentrations of responses in the pick, guess, and search-for-the-different treatments. Yet they yield statistically distinguishable concentrations of coordination attempts, with four non-celebrities showing greater concentration than four celebrities. Meanwhile, the mixed action sets yield statistically distinguishable concentrations of guesses and coordination attempts but not search-for-the-different responses.

Third, we use another bootstrap method to assess whether for a given action set, responses in a target treatment are more concentrated than responses in a reference treatment. In Table 3, *p*-values located within rows labeled $c^*$ indicate whether the $c^*$ in the relevant cell is statistically greater than the $c^*$ to the immediate left. Coordination attempts are more concentrated than guesses given four non-celebrities but not given four celebrities. But note that coordination attempts are more concentrated than guesses for both

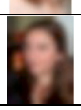
mixed action sets. That is, prominence stymies team reasoning – but not entirely.

Fourth, we use a last bootstrap method to assess whether for a given action set, the extent of the match between coordination attempts and search-for-the-different responses is greater than the extent of the match between coordination attempts and guesses. Coordination/search-for-the-different matches are reliably more common than coordination/guess matches for only two of the action sets: given four non-celebrities ($p < .01$) and given a male celebrity alongside three female non-celebrities ($p < .05$).

We should note that Bacharach's (2006) variable frame theory does not fully accommodate our results. By this theory, which blends elements of cognitive hierarchy logic and team reasoning, players perceive subsets of the action set they encounter as possible options, and then choose among options on the basis of payoff dominance. For instance, given a mixed set composed of a male celebrity and three female non-celebrities, players may perceive the category-based options "select the male celebrity" and "select a female non-celebrity." They would then select the man, because the equilibrium in which they do so payoff dominates the equilibrium in which they (randomly) select a woman. Bacharach assumes that in addition to category-based options, players often perceive the option "select a prominent" but not the option "select a non-prominent." Critically, this assumption implies that with a mixed set containing

**Table 3**

Experiment 2 Results. Numbers within borders are the percentage of players making each response. Grey marks distinctive actions. $c^*$ is a normalized concentration measure. $p$-values in rows labeled $c^*$ indicate whether $c^*$ in the relevant treatment is statistically greater than in the treatment to the immediate left. $p$-values in rows labeled "Comparing …" indicate whether the two action sets above yield statistically distinct values of $c^*$ in the relevant treatment. Non-celebrities are blurred here but were shown in a clear format in the experiment.

| | PICTURE DISPLAYED | TEXT PROVIDED | PICK | GUESS | COORDINATION | SEARCH-FOR-THE-DIFFERENT |
|---|---|---|---|---|---|---|
| **Unmixed Sets: All Celebrities or All Non-Celebrities** |  | Zach Galifianakis | .08 | .06 | .35 | .82 |
| |  | Jennifer Lawrence | .38 | .51 | .27 | .04 |
| |  | Scarlett Johansson | .28 | .29 | .16 | .04 |
| |  | Mila Kunis | .26 | .14 | .22 | .10 |
| | | *n* | 114 | | 86 | 50 |
| | | $c^*$ | 1.16 | 1.44 | 1.04, *ns* | 2.72, $p < .01$ |
| |  | | .19 | .33 | .57 | .87 |
| |  | | .13 | .07 | .12 | .06 |
| |  | | .47 | .41 | .23 | .02 |
| |  | | .20 | .19 | .08 | .04 |
| | | *n* | 98 | | 75 | 47 |
| | | $c^*$ | 1.24 | 1.24 | 1.57, $p < .01$ | 3.05, $p < .01$ |
| | Comparing unmixed action sets | | *ns* | *ns* | $p < .01$ | *ns* |

(*continued on next page*)

**Table 3** (*continued*)

| | PICTURE DISPLAYED | TEXT PROVIDED | PICK | GUESS | COORDINATION | SEARCH-FOR-THE-DIFFERENT |
|---|---|---|---|---|---|---|
| **Mixed Sets: Both Celebrities and Non-Celebrities** |  | Zach Galifianakis | .57 | .71 | .90 | .86 |
| |  | Non-Celebrity | .03 | .03 | .02 | .02 |
| |  | Non-Celebrity | .19 | .17 | .08 | .05 |
| |  | Non-Celebrity | .21 | .09 | .00 | .07 |
| | | *n* | 91 | | 51 | 44 |
| | | *c** | 1.60 | 2.16 | 3.27, $p < .01$ | 2.99, *ns* |
| |  | Non-Celebrity | .10 | .08 | .25 | .75 |
| |  | Jennifer Lawrence | .37 | .39 | .25 | .04 |
| |  | Scarlett Johansson | .25 | .32 | .23 | .13 |
| |  | Mila Kunis | .29 | .22 | .28 | .08 |
| | | *n* | 104 | | 69 | 53 |
| | | *c** | 1.13 | 1.19 | .96, *ns* | 2.35, $p < .01$ |
| | Comparing mixed action sets | | $p < .01$ | $p < .01$ | $p < .01$ | *ns* |

three female celebrities and a male non-celebrity, participants would coordinate on the male non-celebrity, which is contrary to the data.

### *2.3. Experiment 3: The Effect of Feedback*

Can people learn to team reason about distinct actions? To examine this question, we conducted a seven-round experiment. On each round, participants played a coordination game with a different, anonymous counterpart and were paid $1 for successful coordination. After each round, they received feedback about the distribution of choices that had just been made by all the players in their session. They did not, however, receive feedback about what action had been selected by their specific counterpart. Sessions included between 8 and 17 players.

We implemented two treatments. In the "uniform" treatment, participants played the same game on every round. This game had the action set {Jennifer Lawrence, Scarlett Johansson, Adam Driver}.

In the "variable" treatment, participants played a different game on each round. They encountered {Jennifer Lawrence, Scarlett Johansson, Adam Driver} on the last round. Before that, they played six other games in an order determined randomly for each session. These games had the action sets {Instagram, TikTok, Southwest Airlines}, {Honda, Toyota, Pacific Ocean}, {Beijing, Shanghai, London}, {Microsoft, Amazon, Germany}, {Thomas Jefferson, Franklin Roosevelt, Isaac Newton}, and {Apple, Google, Tesla}.

Players sat at individual stations and were given a face down packet. The first sheet in the packet included introductory instructions and was read aloud by an experimenter. The rest of each packet consisted of "selection sheets." At the beginning of a round, each participant turned over the top selection sheet in their packet. This sheet listed the available actions, and participants selected an action by circling it. In place of the card-based method, participants were informed that the order in which actions were listed on each

selection sheet had been randomly determined, so that their counterparts could be facing the same or a different ordering.

We linked participants to specific counterparts and paid them using the following procedure. Consider a session with $n$ participants. Randomly number them from 1 to $n$. On round $j$, player $k$ was paid based on their own action and the action selected by player $(k+j)$ when $(k+j) \leq n$ and was paid based on their own action and the action selected by player $[(k+j) \bmod n]$ when $(k+j) > n$. Here, "mod" denotes the modulo operator. This amounts to a "wrap-around" linking procedure. For example, with 8 players, in consecutive rounds player 1 was linked to players 2, 3, 4, 5, 6, 7, and 8; player 2 was linked to players 3, 4, 5, 6, 7, 8, and 1; player 8 was linked to players 1, 2, 3, 4, 5, 6, and 7.

Before conducting the uniform and variable treatments, we gathered preliminary data which confirmed that most players perceived Adam Driver as distinct from Jennifer Lawrence and Scarlett Johannsson. In particular, we implemented a "search-for-the-different" set-up akin to that of our earlier experiments: We began by asking a single initial participant to select the element from {Jennifer Lawrence, Scarlett Johansson, Adam Driver} that seemed different from the others. Then, we offered a sample of 100 participants (who were separate from the participants in the uniform and variable treatments) $1 if both they and a counterpart selected the action that had thus been deemed different. Eighty-four of them selected Adam Driver.

We considered the possibility of both narrow and general forms of learning. Narrow learning concerns insight about a specific action set gleaned from repeated experience with it. Tests for narrow learning focus on the uniform treatment. They assess how play changes over rounds. Does coordination improve? Does it improve because of increasing play of the distinct action, Adam Driver?

General learning cuts across action sets. It concerns portable insights. Tests for it consider the variable treatment. Perhaps the cleanest test of general learning compares last round variable treatment play with first round uniform treatment play. In each case, the action set is {Jennifer Lawrence, Scarlett Johansson, Adam Driver}, but in the variable treatment, players have previously encountered six isomorphic action sets. Additional tests of general learning assess how variable treatment play changes over rounds. Again, does coordination improve? Does it improve because of increasing play of the distinct action?

The data reveal three takeaways. First, there is substantial narrow learning but little general learning. Over the seven rounds, uniform participants advantageously adjust their coordination efforts much more than variable participants. Second, the narrow learning that occurs is, however, not about leveraging distinct actions. Coordination improvements achieved by uniform participants tend to arise from participants whose initial choice is relatively uncommon switching to the most common choice. For example, participants in a group in which more players chose Jennifer Lawrence than either Adam Driver or Scarlett Johansson in round 1, typically coordinate on Jennifer Lawrence by round 7. Third, though there is little general learning, the adjustments to coordination efforts that do arise in the variable treatment are in the direction of leveraging distinct actions.

Table 4 displays observed coordination rates by session. Recall that in each round of each session, every participant was

**Table 4**
Coordination rates in each round of each session of Experiment 3.

| **UNIFORM TREATMENT** | | | | | | | |
|---|---|---|---|---|---|---|---|
| # of Players in the Session | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 |
| 17 | .53 | .41 | .47 | .76 | .88 | .76 | .88 |
| 16 | .25 | .31 | .50 | .31 | .50 | .75 | .63 |
| 16 | .31 | .19 | .25 | .31 | .63 | .75 | 1.00 |
| 14 | .36 | .64 | .14 | .50 | .43 | .86 | .71 |
| 14 | .43 | .50 | .29 | .57 | .86 | .86 | .86 |
| 13 | .31 | .62 | .85 | .85 | .85 | .85 | .85 |
| 13 | .38 | .23 | .46 | .69 | .85 | 1.00 | 1.00 |
| 11 | .18 | .55 | .55 | .27 | .55 | .64 | .55 |
| 11 | .36 | .82 | .82 | 1.00 | 1.00 | 1.00 | 1.00 |
| 10 | .90 | .50 | .40 | .60 | .60 | .60 | .90 |
| 9 | .44 | .44 | .33 | .78 | .56 | .78 | .56 |
| 8 | .75 | .38 | .38 | .00 | .75 | .50 | .75 |
| median | .37 | .47 | .43 | .59 | .69 | .77 | .85 |
| **VARIABLE TREATMENT** | | | | | | | |
| # of Players in the Session | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 |
| 17 | .47 | .41 | .29 | .53 | .53 | .29 | .18 |
| 17 | .47 | .41 | .35 | .41 | .29 | .71 | .47 |
| 16 | .63 | .38 | .38 | .19 | .50 | .56 | .44 |
| 10 | .40 | .20 | .50 | .20 | .40 | .60 | .40 |
| 10 | .40 | .40 | .40 | .20 | .20 | .40 | .20 |
| 10 | .10 | .20 | .30 | .40 | .20 | .10 | .30 |
| 9 | .44 | .33 | .33 | .44 | .11 | .44 | .22 |
| 9 | .44 | .33 | .11 | .44 | .33 | .22 | .44 |
| 8 | .38 | .25 | .13 | .00 | .50 | .75 | .25 |
| 8 | .25 | .00 | .38 | .25 | .63 | .38 | .38 |
| median | .42 | .33 | .34 | .33 | .37 | .42 | .34 |

anonymously linked to another participant. The table lists the percentage of these links that yielded coordination rather than miscoordination. There are two reasons we present data in terms of this observed coordination rate rather than the normalized coordination rate, $c^*$, as in Experiments 1 and 2. First, $c^*$ can be unstable with small samples, and several of the sessions in the present experiment included only eight or nine participants. Second, observed coordination rates are more amenable to round-by-round statistical analyses.

Consistent with substantial narrow learning, Table 4 shows that coordination in the uniform treatment improves markedly over rounds. The median observed coordination rate across sessions is only .37 in the initial round, but it climbs monotonically and reaches .85 in the last round. Tests of proportions comparing uniform coordination rates in round 7 versus round 5 or any earlier round are statistically significant ($p < .015$ versus round 5 and $p < .0001$ for each earlier round). So are parallel tests comparing round 6 with round 4 or earlier ($p < .0001$ in each case) and round 5 with round 3 or earlier ($p < .0001$ in each case). Furthermore, while the observed coordination rate is not statistically distinguishable from chance in round 1, it statistically departs from chance beginning in round 2 and in every round thereafter ($p < .01$ for rounds 2 and 3 and $p < .0001$ from there).

On the other hand, consistent with scant general learning, observed coordination rates in the variable treatment do not change much across rounds. The median rate across sessions begins at .42 and never rises above that figure. On the last round – in which, to reiterate, variable and uniform participants face the same actor-and-actresses action set – it is only .34. Notably, the coordination rate is slightly lower on round 7 of the variable treatment than on round 1 of the uniform treatment (.34 versus .37). In addition, tests of proportions comparing variable coordination rates between any two rounds are never significant (the lowest p-value is .0417, which is insufficient given Bonferroni corrections), and the variable coordination rate is statistically distinguishable from chance only in round 6 ($p < .01$).

Table 5 displays the proportion of players selecting the distinct action on each round. Its data are consistent with the notion that the narrow learning in the uniform treatment is not about leveraging distinct actions. As we have mentioned, the coordination improvements achieved by uniform participants tend to arise from participants whose initial choice are relatively uncommon switching to the most common choice. Indeed, the percentage of uniform participants who select the distinct action falls over rounds. It is 22% on the first round and only 13% on the last round.

Though the variable treatment results suggest little general learning, there is some evidence that what learning does occur is in the direction of leveraging distinct actions. The 30% who attempt to coordinate on the distinct Adam Driver in the last round of the variable treatment is marginally statistically larger than the 22% who attempt to coordinate on him in the initial round of the uniform treatment, ($p = .08$ by a one-way test for equality of proportions). Moreover, as Table 5 shows, rounds 5 and 6 of the variable treatment show intriguingly frequent selections of the distinct action. Accordingly, to look for further evidence of whether both uniform and variable participants learn to leverage distinctiveness, we fit linear probability models for each treatment. In these models, the dependent variable indicated whether a player selected the distinct action. The independent variables included dummies for rounds two through seven and, in the variable treatment, dummies for each game other than the actor-and-actresses game. We fit versions of these models with session fixed effects but not subject fixed effects and vice versa. The results of these two approaches, displayed in Table 5, were nearly identical. They include an interesting juxtaposition: Relative to round 1, play of the distinct action is less common on rounds 5, 6, and 7 of the uniform treatment. In contrast, it is more common on rounds 5 and 6, though not round 7, of the variable treatment (Table 6).

In sum, the uniform treatment shows substantial narrow learning. However, the learning is not to team reason about distinct actions. Meanwhile, the variable treatment shows limited evidence of general learning in the direction of team reasoning about distinctiveness.

The relative lack of general learning in our experiment contrasts with an instance of general learning reported by Alberti, Sugden, and Tsutsui (2012). On each of many rounds, these authors presented a new quartet of images, and players attempted to coordinate on one of the images with a partner. Importantly, partnerships were maintained across rounds. After each round, each player was told what image their partner had chosen. Furthermore, there were similarities in images across rounds, so that a player might try to select an image akin to the image their partner selected on the previous round. In early rounds, participants' selections were highly correlated with their self-reports of which images they liked. However, as the experiment progressed, this correlation diminished, and within-pair conventions emerged. In later rounds, many pairs' matching frequency exceeded chance though the specific pictures matched on differed markedly across pairs – and these pictures were not necessarily liked by either member of the pair. Thus, while our variable treatment participants could not exploit the consistent presence of a distinct, prominent action alongside two other prominent actions, Alberti et al.'s participants were able to define and home in on distinct pictures alongside more-liked pictures.

Some basic structural features likely contribute to why Alberti et al.'s setting engenders general learning whereas our setting does not. In Alberti et al.'s experiment, the information players receive after each round concerns just a single individual, namely the partner they maintain for the entire, repeated interaction. In contrast, in our experiment, players receive information only about the distribution of actions in their entire session, they never receive information about a specific counterpart, and, indeed, they do not maintain partnerships across rounds. It is possible that because of these structural features, Albert et al.'s participants are able to

**Table 5**
Percent of players picking the distinct action in each round of the two treatments of Experiment 3.

| | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 |
|---|---|---|---|---|---|---|---|
| **UNIFORM** | .22 | .16 | .16 | .16 | .14 | .11 | .13 |
| **VARIABLE** | .32 | .32 | .39 | .35 | .48 | .52 | .30 |

**Table 6**
Fits of linear probability models predicting play of the distinct action in each of round of Experiment 3. t-statistics in parentheses; * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

| | UNIFORM TREATMENT | | VARIABLE TREATMENT | |
|---|---|---|---|---|
| | Model U1 | Model U2 | Model V1 | Model V2 |
| Round 2 | -.059 | -.059 | -.020 | -.020 |
| | (-1.57) | (-1.67) | (-.31) | (-.34) |
| Round 3 | -.066 | -.066 | .080 | .080 |
| | (-1.75) | (-1.85) | (1.25) | (1.36) |
| Round 4 | -.059 | -.059 | .092 | .092 |
| | (-1.57) | (-1.67) | (1.47) | (1.60) |
| Round 5 | -.079* | -.079* | .136* | .136* |
| | (-2.09) | (-2.22) | (2.15) | (2.34) |
| Round 6 | -.118** | -.118** | .146* | .146* |
| | (-3.14) | (-3.34) | (2.28) | (2.49) |
| Round 7 | -.092* | -.092* | -.001 | -.001 |
| | (-2.09) | (-2.59) | (-.01) | (-.01) |
| Constant | .094*** | .211 | .106 | -.084 |
| | (2.09) | (1.77) | (1.33) | (-.50) |
| Session FE | Yes | No | Yes | No |
| Player FE | No | Yes | No | Yes |
| Game FE | N/A | N/A | Yes | Yes |
| Adjusted $R^2$ | .181 | .273 | .099 | .240 |
| Observations | 1064 | | 798 | |

develop a productive dyadic dynamic, while a productive dynamic is out of reach of our participants.

To pursue this possibility and investigate what features are necessary for general learning, it would be useful to conduct additional experiments in both settings. These experiments could contrast maintained partnerships with round-by-round random pairing. They could also contrast feedback that is about only one's partner with feedback about an entire group.

Structural features notwithstanding, note that the two settings require different kinds of team reasoning and raise different obstacles to such thinking. In our experiment, variable treatment players do not learn to leverage categorically distinct actions, perhaps in part because prominence stymies such learning and induces salience-based play. In Alberti, Sugden, and Tsutsui's experiment, players do learn to leverage visual cues to distinctiveness. They are not stymied by their liking of certain images which could in principle also induce salience-based play. Below, we further juxtapose how prominence-based and liking-based salience may interact with team reasoning about distinctiveness.

## 3. Conclusion

We observed fairly good coordination on distinct elements of sets that had no prominent actions but not of sets containing prominent actions. This finding, along with supplementary findings from pick-then-guess and explicit search-for-the-different tasks, suggests that given prominent actions, people follow primary and higher orders of salience even if doing so cannot facilitate coordination, while team reasoning based on distinctiveness can. Additional data indicate that, at least when given group-level information, players do not readily learn to invoke team reasoning about distinctiveness.

Hargreaves Heap, Rojo Arjona, and Sugden (2017) also investigated reliance on prominence and distinctiveness. They did so as part of a larger study that examined other coordination methods, including selecting the best-liked action and the most typical action. Hargreaves Heap et al defined prominence differently than we have, as being at or near the "top of the most natural ranking" for a category, citing the example of Everest for the category mountains. Among several intriguing findings, Hargreaves Heap et al. reported that play was more consistent with reliance on prominence and distinctiveness given restricted sets, as when attempting to coordinate on an element of {Ferrari, Ford, Mercedes, BMW, and Honda}, than unrestricted sets, as when attempting to coordinate on any car manufacturer. For instance, Ferrari was the modal restricted choice, perhaps because its racecar heritage grants it prominence in the sense of "top of ranking" or renders it distinct from the other manufacturers listed. On the other hand, Ford, which is a more typical consumer vehicle, was the modal unrestricted choice.

It is worth emphasizing the differences – and consequent complementarity – that mark our work and that of Hargreaves Heap et al. Like Bardsley, Mehta, Starmer, and Sugden (2010), Hargreaves Heap et al. do not set prominence and distinctiveness against each other, as we have. They highlight ways in which people savvily grasp what circumstances mesh best with these coordination methods.

To explicate their argument, again consider Ferrari, prominence, and distinctiveness. As we have mentioned, in the restricted set, Ferrari's racing heritage may make it stand out. However, a racing heritage will not make Ferrari stand out in the unrestricted set, which contains other manufacturers with such histories as well as manufacturers that are prominent or distinct for other reasons. This pattern is likely general: consensus about prominence and distinctiveness may be more forthcoming given restricted than unrestricted sets, and players may appreciate that. They may understand that reliance on prominence and distinctiveness is more advantageous given restricted than unrestricted sets and select their coordination method accordingly. In contrast to Hargreaves Heap et al., we are expressly concerned with exploring limits on people's savvy use of different coordination methods. We thus purposely set prominence and distinctiveness against one another and collect evidence suggesting that prominence stymies reliance on distinctiveness even when only the latter can yield substantial coordination.

Of course, a full view of people's coordination abilities must integrate both perspectives. As Bardsely et al. and Hargreaves Heap et al show, people can evince an impressive ability to leverage potential focal points. As we show, people can also evince meaningful limitations in their ability to leverage potential focal points.

Future research might explore limits on people's coordination prowess by setting factors other than prominence against distinctiveness. Put differently, while we have detected what we characterized as over-reliance on the salience of prominent actions, additional work could consider over-reliance on the salience of other types of actions. Consider the liking or attractiveness of actions (which is implicated in the learning study by Alberti, Sugden, & Tsutsui, 2012). This factor sometimes co-occurs with prominence – Ferrari, for instance, is surely well-liked by many players – and our studies notably did not disentangle the two. Yet, prominence and liking are certainly distinct, and it would be useful to determine if both prominent and well-liked actions can stymie distinctiveness, or only actions that are both prominent and well-liked do so, or precisely what the case may be. We suspect that both sources of salience, and indeed others as well, often stymie distinctiveness.

We have conducted a pilot study that aims to disentangle prominence and attractiveness in settings akin to our main experiments. We had UCSD participants consider the action sets {Abraham Lincoln, George Washington, Tianjin} and {American Airlines, United Airlines, jetBlue}. We asked participants to rate each member of each set on the scale {-2 = Dislike a lot, -1 = Dislike a little, 0 = Neither like nor dislike, 1 = Like a little, 2 = Like a lot}. The mode and median rating for both presidents were around 1, while those of Tianjin were around zero; in other words, our participants tended to like these prominent presidents but neither liked nor disliked Tianjin. On the other hand, the mode and median rating of all three airlines was zero; our participants did not tend to like or dislike any of them. Meanwhile, when asked to indicate which airline was the "most prominent," 50% selected American, 38% selected United, and only 18% selected jetBlue. Unlike the U.S. Presidents and Chinese cities action set, the airline action set may help in disentangling prominence and liking. Nevertheless, it yielded our typical result: In a search-for-the-different condition, 82% selected jetBlue, whereas in a standard coordination game only 20% selected jetBlue, 40% selected American, and 40% selected United.

Though our main experiments did not disentangle prominence and attractiveness, and indeed might have confounded them, our results do clearly indicate that pronounced salience of action labels, whatever its basis, may stymie team reasoning about label distinctiveness. Positioning of our findings as about over-reliance on label salience in comparison to label distinctiveness, highlights parallels with an important stream of research that has investigated over-reliance on payoff salience in comparison to payoff distinctiveness. For instance, Faillo, Smerilli, and Sugden (2013, 2017; see also Isoni, Poulsen, Sugden, & Tsutsui, 2014, and Bardsley et al's "numbers" experiments) had participants attempt to coordinate on either (a) 10 points for Player 1 and 9 points for Player 2, (b) 9 points for Player 1 and 10 points for Player 2, or (c) 8 points for Player 1 and 7 points for Player 2. Participants rarely selected (c). If non-strategic players are attracted to high individual payoffs or repelled by Pareto inferior payoffs, and strategic players accordingly best-respond, then such results indicate that team reasoning about distinctiveness may be stymied by payoff-based salience.

Relatedly, Crawford, Gneezy, and Rottenstreich (2008) set the potential salience of labels and payoffs against each other. They observed that powerful label-based focal points in pure coordination games can lose their efficacy given even slightly mixed motives (see Isoni, Poulsen, Sugden, and Tsutsui, 2013 for a class of games which nuances this result). For instance, when asked to select either the letter *X* or *Y* in a two-player game in which each player would receive $5 for matching, a large majority selected *X*. In contrast, when one player would receive $5 for coordination on *X* but $6 for coordination on *Y*, while the other player faced the reverse payoffs, majorities in each role selected the letter that would yield them the greater payoff. When the payoffs were $5 and $5.10, majorities in each role selected the letter that could yield them the lesser payoff. Given either asymmetry, people evidently based their play on how they felt about the different payoffs and how their counterpart might feel, so that *X* no longer stood out.

In conclusion, our findings about how pronounced salience of action labels may stymie team reasoning about label distinctiveness reinforces a recurring research theme: While people are in many ways savvy coordinators, there are large classes of situations in which they frequently fail to leverage potential focal points.

## Declaration of competing interest

None.

## Data availability

Data for Experiments 1 and 2 are binary choices that are fully described in the tables. Data for Experiment 3 are available via the link provided. Experiment 3 data.xlsx (Reference data) (OSF).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.geb.2024.07.010.

## References


Bacharach, Michael., 1999. Interactive Team Reasoning: A Contribution to the Theory of Cooperation. Research in Economics 53 (20), 117–147.

Bacharach, Michael, Stahl, Dale O., 2000. Variable-Frame Level-n Theory. Games. Econ. Behav. 32 (2), 220–246.

Bardsley, Nicholas, Mehta, Judith, Starmer, Chris, Sugden, Robert, 2010. Explaining Focal Points: Cognitive Hierarchy Theory versus Team Reasoning. The Economic Journal 120 (543), 40–79.

Camerer, Colin F., Ho, Teck-Hua, Chong, Juin-Kuan, 2004. A Cognitive Hierarchy Model of Games. Quarterly Journal of Economics 119 (3), 861–898.

Crawford, V.P., Costa-Gomes, M.A., Iriberri, N., 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. J. Econ. Lit. 51 (1), 5–62.

Crawford, Vincent.P., Gneezy, Uri, Rottenstreich, Yuval, 2008. The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. Am. Econ. Rev. 98 (4), 1443–1458.

Crawford, Vincent P., Haller, Hans, 1990. Learning How to Cooperate: Optimal Play in Repeated Coordination Games. Econometrica 58 (3), 571–595.

Crawford, Vincent P., Iriberri, Nagore, 2007. Fatal Attraction: Salience, Naivete, and Sophistication in Experimental 'Hide-and-Seek' Games. American Economic Review 97 (5), 1731–1750.

Faillo, M., Smerilli, A., Sugden, R., 2013. The roles of level-k and team reasoning in solving coordination games. University of Trento Cognitive and Experimental Economics Laboratory Working Paper, pp. 6–13.

Faillo, M., Smerilli, A., Sugden, R., 2017. Bounded best-response and collective-optimality reasoning in coordination games. J. Econ. Behav. Organ.

Hargreaves Heap, Shaun, Arjona, David Rojo, Sugden, Robert, 2014. How Portable Is Level-0 Behavior? A Test of Level-k Theory in Games With Non-Neutral Frames. Econometrica 82 (3), 1133–1151.

Hargreaves Heap, S.P., Rojo Arjona, D., Sugden, R., 2017. Coordination when there are restricted and unrestricted options. Theory and Decision 83 (1), 107–129.

Isoni, A., Poulsen, A., Sugden, R., Tsutsui, K., 2013. Focal points in tacit bargaining problems: Experimental evidence. Eur. Econ. Rev. 59, 167–188.

Isoni, A., Poulsen, A., Sugden, R., Tsutsui, K., 2014. Efficiency, equality, and labeling: An experimental investigation of focal points in explicit bargaining. American Economic Review 104 (10), 3256–3287.

Mehta, Judith, Starmer, Chris, Sugden, Robert, 1994. The Nature of Salience: An Experimental Investigation of Pure Coordination Games. American Economic Review 84 (3), 658–673.

Merriam-Webster, 2023. Prominent. Merriam-Webster.com dictionary. Retrieved from. https://www.merriam-webster.com/dictionary/prominent.

Myerson, R.B., 2009. Learning from Schelling's strategy of conflict. J. Econ. Lit. 47 (4), 1109–1125.

Schelling, Thomas., 1960. The Strategy of Conflict. Harvard University Press, Cambridge, MA.

Stahl, Dale O., Wilson, Paul, 1995. On Players' Models of Other Players: Theory and Experimental Evidence. Games. Econ. Behav. 10 (1), 218–254.

Sugden, Robert., 1995. A Theory of Focal Points. Economic Journal 105 (430), 533–550.