Quantitative Economics 12 (2021), 283–312 1759-7331/20210283

A notion of prominence for games with natural-language labels

Alessandro Sontuoso
Smith Institute for Political Economy and Philosophy, Chapman University and Philosophy, Politics and Economics, University of Pennsylvania

Sudeep Bhatia
Department of Psychology, University of Pennsylvania

We study games with natural-language labels (i.e., strategic problems where options are denoted by words), for which we propose and test a measurable characterization of prominence. We assume that—ceteris paribus—players find particularly prominent those strategies that are denoted by words more frequently used in their everyday language. To operationalize this assumption, we suggest that the prominence of a strategy-label is correlated with its frequency of occurrence in large text corpora, such as the Google Books corpus ("n-gram" frequency). In testing for the strategic use of word frequency, we consider experimental games with different incentive structures (such as incentives *to* and *not to* coordinate), as well as subjects from different cultural/linguistic backgrounds. Our data show that frequently-mentioned labels are more (less) likely to be selected when there are incentives to match (mismatch) others. Furthermore, varying one's knowledge of the others' country of residence significantly affects one's reliance on word frequency. Overall, the data show that individuals play strategies that fulfill our characterization of prominence in a (boundedly) rational manner.

Keywords. Focal points, salience, coordination, hide-and-seek, culture, language.

JEL classification. C72, C91.

1. Introduction

Coordination problems can affect a broad array of economic interactions, yet they are often solved by shared *cultural* understandings. Culture is in fact believed to facilitate coordination by informally codifying the common beliefs and practices that are shared by a society's members (Alesina and Giuliano (2015)). For example, global companies that wish to match potential demand in developing markets usually adjust their strategies in such a way to connect to (and hence coordinate with) culturally diverse consumers. In that case, cultural awareness affects an organization's marketing activities—from labeling to customer services—thereby making *some* products more salient for

Alessandro Sontuoso: sontuoso@sas.upenn.edu
Sudeep Bhatia: bhatiasu@sas.upenn.edu
We are grateful to Cristina Bicchieri, Gary Charness, Alexander Funcke, Einav Hart, Rosemarie Nagel, and David Rojo-Arjona for their helpful comments.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1212

some consumer segments (Kapferer (2012)). More generally, it has been suggested that cultural and sociolinguistic competence favor the coordination of economic interactions by promoting trade, development, and growth (Easterly and Levine (1997); Lazear (1999); Melitz (2008)). Natural languages are indeed an instance of a cultural code that eases the exchange of information, and hence facilitates economic activities (Ginsburgh and Weber (2020)).[1]

Thomas Schelling (1960) was the first to note that a wide range of economic interactions can be formally represented as a "coordination problem," that is, a symmetric simultaneous-move game with multiple pure-strategy Nash equilibria.[2] Schelling informally observed that coordination puzzles are often solved by exploiting *contextual* or *cultural* cues that drive expectations in such a way to make a specific course of action salient in specific circumstances. As we will elaborate below, more recent explanations of prominence (i.e., salience) have centered around features of the game that players perceive as distinctive, where such distinctive features do not necessarily vary across cultures.[3] Here, on the other hand, we propose and test a characterization of prominence that directly rests on players' culture.

To that end, we note that advances in online data collection and processing have made it possible for researchers to access (culture-dependent) digitized information at little cost. These advances have boosted the use of "big data" to investigate economic, psychological, and medical outcomes (e.g., Varian (2014); Einav and Levin (2014); Matz and Netzer (2017); Yin, Sulieman, and Malin (2019)). In particular, recent developments have facilitated the extraction of information from online text documents, which has made it possible to track language use across different cultures and over long periods of time (Michel et al. (2011)). Such developments have led us to propose the following notion of prominence for "games with natural-language labels" (i.e., strategic problems where options are denoted by words).

In short, we assume that *players find particularly prominent those strategies that are denoted by words more frequently used in their everyday language, ceteris paribus*. Specif-

[1]Research in political economy has shown that countries' per-capita GDP growth is inversely related to their ethno-linguistic fractionalization (Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg (2003)). Also, it has been suggested that production line workers coordinate tasks more efficiently if they speak the same language, although this might result in a form of labor market discrimination against cultural minorities (Lang (1986); Kossoudji (1988)).

[2]The class of coordination problems contains any situation in which there are multiple ways agents may "match" their behavior for mutual benefit. This class contains a vast and diverse array of interactions, including games with (slightly) conflicting interests, with and without Pareto-rankable equilibria. Specifically, interactions where players wish to coordinate—but have conflicting interests—are sometimes referred to as *impure* coordination games (e.g., Luce and Raiffa's (1957), battle of the sexes is a classic example). Bacharach's (2006) Hi-Lo game is an instance of a conflict-free problem with two equilibria, one of which is Pareto-dominated by the other. On the other hand, Schelling's (1960) driving game is a case where each player is completely indifferent between the equilibria. See Lewis (1969) for an early book-length account of coordination games.

[3]For example, Bacharach and Bernasconi (1997) noted that players solve coordination games by exploiting distinctive attributes of the strategy space. The authors went on to note that—in the class of games where strategy options are represented by visual objects—there «are attributes, such as colors, comparative sizes, and simple geometric shapes, whose saliencies are *universal* constants [...]. But for others, saliencies are *culture-dependent*» (Bacharach and Bernasconi (1997), p. 7, our italics).

ically, to operationalize this assumption we suggest that the prominence of a strategy-label is correlated with its frequency of occurrence in large text corpora, like the Google Books corpus ("n-gram" frequency).[4] In testing for the strategic use of word frequency, we consider experimental games with different incentive structures (such as incentives *to* and *not to* coordinate), participants from different cultural backgrounds, as well as participants with asymmetric knowledge about the counterpart's cultural background.

Before fleshing out our hypotheses and methods, we note that formal (game-theoretic) characterizations of prominence have typically revolved around features of the strategic problem, that would drive one's perception as to the uniqueness of a solution independently of one's culture. For example, Harsanyi and Selten's (1988) payoff-dominance criterion assumes that *any* rational player (upon facing a one-shot game with no pre-play communication) would discard solutions that are collectively suboptimal: in that case, the payoff structure of the game serves as a "cue," thereby directing players' expectations toward the collectively optimal solution. Related characterizations of prominence have integrated the payoff-dominance criterion into a theory of framing, whereby players select the collectively optimal solution to some perceptual description of the game.[5] In this respect, lab experiments have provided evidence confirming that *payoff* and *frame* asymmetries do—each to a different extent—affect behavior. That is, experiments have shown that if one of the game solutions is an "oddity" (in the sense that its label or payoff profile differ from the others), then the distinctiveness of that solution serves as a cue to facilitate coordination (e.g., Bacharach and Bernasconi (1997); Rubinstein, Tversky, and Heller (1997); Crawford, Gneezy, and Rottenstreich (2008)). In summary, per the above explanations of prominence, a solution is generally viewed as a "focal point" by virtue of culture-invariant cues. In what follows, instead, we explicitly study culture's role in affecting behavior by introducing an a-priori measurable proxy for prominence (i.e., word frequency).

Our approach rests on a psychologically grounded characterization of focality that has important implications for our understanding of strategic reasoning. Research in cognitive psychology has shown that the *frequency of exposure* to words is closely related to word *fluency*, that is, the ease with which an individual is able to recognize, retrieve, and process a word. Word frequency—through its effect on fluency—has been shown to have a role in a wide range of memory and language tasks (e.g., Anderson and Schooler (1991); Balota and Spieler (1999); Jescheniak and Levelt (1994); Morrison and

[4]This is a standard metric for word frequency in languages, and has so far been used for psychological, sociological, and historical research (e.g., Hills, Proto, Sgroi, and Seresinhe (2019); Garg, Schiebinger, Jurafsky, and Zou (2018)).

[5]Frame-based theories of coordination are sometimes divided into two broad classes, namely, team reasoning and level-k models. Theories of team reasoning assume that a group member follows the decision rule/frame that, if followed by other members, would be optimal for each of them (e.g., Crawford and Haller (1990); Bacharach (1993); Sugden (1995); Casajus (2000); Blume (2000); Janssen (2001); Alós-Ferrer and Kuzmics (2013)). By contrast, level-k theories assume a hierarchy of cognitive levels, whereby higher types best respond to lower-level players, anchoring their beliefs in the behavior of strategically naïve individuals (see Bacharach and Stahl (2000) for a frame-based model of level-k reasoning). Relatedly, Charness and Sontuoso (2019) took a hybrid approach such that team reasoning is reduced to the case where one best responds to other types, given one's partial awareness of frames.

Ellis (1995); Seidenberg and McClelland (1989)). Notably, there is evidence that individuals use word frequency as a cue in several *non*-strategic domains, like probability judgment (Dougherty, Franco-Watkins, and Thomas (2008); Tversky and Kahneman (1974)), risk perception (Hertwig, Pachur, and Kurzenhäuser (2005); Lichtenstein, Slovic, Fischhoff, Layman, and Combs (1978)), as well as factual judgment (Gigerenzer and Goldstein (1996); Hertwig, Herzog, Schooler, and Reimer (2008)). In all these domains, the frequency of occurrence of a word in everyday language is positively correlated with the tendency to select that word as a response, and to evaluate the object denoted by that word as being large, important, truthful, or desirable. Indeed, fluency is believed to be one of the mechanisms through which the *availability heuristic* operates (Tversky and Kahneman (1973); Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, and Simons (1991)).

The nonstrategic literature above caused us to speculate that word frequency could be used as a proxy for prominence in strategic domains. For example, consider a pure coordination game where the strategy set is given by {paprika, curry, chili}.[6] In this case, the strategy-label that is most frequently mentioned in everyday language may be most "fluent" (easy to process). Hence subjects could be drawn to that option, in the same way as they are drawn to fluent options in the non-strategic literature above. Yet—unlike the psychological literature—this paper aims to verify whether subjects *strategically* exploit word frequency. Here, one might argue that if there is common reason to believe that an option easily comes to mind to people with the same cultural background, then it may be optimal to select that option in pure coordination games.[7] But what about games with unaligned incentives? And what about players with different cultural/linguistic backgrounds? To address these questions and test for the strategic use of word frequency, we propose the following three studies.

We designed Study 1 as a preliminary test for predicting behavior in 2-player (one-shot) pure coordination games with a finite set of strategies (played by culturally alike participants). A few points are worth noting. First, our strategy options present no *payoff asymmetries*; in fact, note that we only consider games without Pareto-rankable equilibria, which allows us to rule out a common driver of prominence as an explanation for our data patterns.[8] Second, we designed each game by randomly drawing labels from lists of words in the same semantic domain (e.g., names of food ingredients): this means that our games have no obvious *frame asymmetries*, and hence we can rule out another

[6] *Pure* coordination games are characterized by the following payoff structure: if players select the same strategy, they each receive an identical positive payoff (say, 1 currency unit); otherwise, they each receive nothing.

[7] One mechanism supporting such a hypothesis (about the strategic use of word frequency) is that subjects may realize that their counterpart might be culturally alike, and hence view the problem in a similar way. For evidence on "projection," see Hedden and Zhang (2002), Sebanz, Knoblich, and Prinz (2003), and Rubinstein and Salant (2016).

[8] For early evidence on coordination games with Pareto-rankable equilibria, see Cooper, DeJong, Forsythe, and Ross (1990); for the case of repeated games, see also Van Huyck, Battalio, and Beil (1990). For more recent experiments with Pareto-rankable equilibria, see Bardsley and Ule (2017) and Faillo, Smerilli, and Sugden (2017).

common driver of prominence.[9] Having found a strong correlation between word frequency (i.e., *n-gram* frequency computed in the general English Google Books corpus) and behavior in coordination games, we designed two more studies to put to test our notion of prominence under different conditions.

Study 2 contrasts choice behavior in (i) a pure coordination game, against the behavior of participants in three alternative roles/conditions (with each condition featuring exactly the same list of labels). More specifically, we consider: (ii) the case in which a subject is prompted to *pick* an option, without any explicit objective; (iii) the case where a subject is prompted to avoid matching her counterpart's choice, under the assumption that her counterpart instead wants to match (i.e., the role of *Hider* in a Hide-and-Seek game); (*iv*) the case where a subject is prompted to match her counterpart's choice, under the assumption that her counterpart instead wants to avoid any such match (i.e., the role of *Seeker* in a Hide-and-Seek game). Since all our conditions involve the exact same options, note that if the effect of word frequency were merely due to an automatic (or naïve) response, then we should observe similar choice distributions across the four conditions. However, the data paint a different picture. Participants in problem *ii* (i.e., "Pickers") were about as likely to select the most frequently-mentioned label as were participants in problem *i* (i.e., "Coordinators"). On the other hand, Hiders were less likely to select the most frequently-mentioned label than Seekers and, in turn, Seekers were less likely than Coordinators. As shall be discussed, this pattern indicates a boundedly rational, strategic use of word frequency that is consistent with a particular specification of level-k reasoning (Nagel (1995); Stahl and Wilson (1995); Costa-Gomes, Crawford, and Broseta (2001)).

We designed Study 3 to delve further into the strategic use of word frequency in pure coordination games. To that end, each of the games of Study 3 involves labels that we purposely selected so that the option with the highest word frequency differs between the American- and British-English vocabularies (as measured by the *n-gram* frequency in the American- and British-English Google Books corpora, resp.).[10] We then varied the cultural/linguistic makeup of the subject pool by recruiting individuals residing in either the US or the UK; additionally, we manipulated our subjects' knowledge of the counterpart's country of residence. Consistent with our predictions, the data show that choice behavior differs between US and UK subjects and, in each case, it is positively related to the word frequency of the strategy-labels in the relevant vocabulary. Moreover, the data show that subjects are less likely to rely on word frequency as a means to guiding their behavior *if* they know that their counterpart resides in a different country. Put differently, if subjects are aware that their assigned partner is alike (in terms of cultural

[9]Recent experiments with *more or less* obvious frame asymmetries include, among others: Blume and Gneezy (2010); Bardsley, Mehta, Starmer, and Sugden (2010); Hargreaves Heap, Rojo Arjona, and Sugden (2014, 2017). Of particular interest is Hargreaves Heap et al.'s (2017) design, which elicits subjects' beliefs about alternative heuristics that may drive behavior in coordination games. (Note that, unlike our studies, their design does not involve an a-priori measurable proxy for prominence, nor does it vary subjects' incentives or cultural background.)

[10]For example, consider the game with strategy set {paprika, curry, chili}. There, "curry" has the highest n-gram frequency in *British English* and the lowest one in American English; conversely, "chili" has the highest n-gram frequency in *American English* and the lowest one in British English.

background), then they are more likely to select the label most frequently mentioned in their vocabulary.

We finally compared coordination rates that would be obtained if different subsamples were paired with each other, using Monte Carlo methods. In brief, successful coordination is more likely when subjects were *knowingly* paired with partners from their own country, as opposed to when they were knowingly or unknowingly paired with partners from a different country. Notably, subjects who were knowingly paired with partners from the same country exhibit expected coordination rates between 10 and 20 percentage points higher than chance.

To conclude, for the first time we propose and test an a-priori measurable proxy for prominence that explicitly rests on players' culture. Our experiments provide very robust evidence indicating that individuals play strategies fulfilling our notion of prominence in a (boundedly) rational manner. Remarkably, reliance on word frequency leads to higher rates of coordination than chance, and more so when individuals knowingly share a cultural background. The remainder of the article is organized in this manner: Section 2 lays out the experimental procedures, Sections 3–5 present our studies, and Section 6 concludes.

2. General procedures

Our studies were conducted online between September and November 2016. A primary motivation for running online experiments is the ease with which the experimenter can control the cultural makeup of the subject pool. Another advantage is the ease with which the experimenter can vary subjects' knowledge of the fellow participants' cultural characteristics: this makes online experimentation optimal for testing culture-related hypotheses. (For a methodological discussion of extra-laboratory experiments, we refer the reader to Charness, Gneezy, and Kuhn (2013).)

Participants were recruited through *Prolific Academic*, a crowdsourcing platform backed by Oxford University Innovation (https://www.prolific.co/). Participation in our study was limited to individuals with a Prolific Academic approval rate greater than 95%. At the beginning of each study, subjects were informed that they would be paired with a fellow participant at random, and that they would not know the identity of their counterpart or be able to communicate with them. Participants' responses were incentivized, as specified in the following sections. No subject was allowed to participate in more than one study.

As a proxy for the prominence of strategy-labels, we used the *case-insensitive average yearly n-gram frequency* (henceforth simply NGRAM) of the corresponding words in the Google Books corpus, for books published after 2000. NGRAM values represent the fraction of times phrases (in our case, words) occur in the corpus of interest; for Studies 1, 2 we used the general English corpus, whereas for Study 3 we used the American-English and British-English corpora. The values were obtained through the Google n-gram tool in August 2016, shortly before running the studies (https://books.google.com/ngrams/info). Note that n-gram frequencies are a reliable, standard metric for word popularity in corpus linguistics (Michel et al. (2011); Garg, Schiebinger, Jurafsky, and Zou (2018); Hills, Proto, Sgroi, and Seresinhe (2019)).

3. Study 1

Demographics

The subject pool for Study 1 consisted of 91 US resident individuals. The average participant was 33 years old, and 57% of the subjects were male. Participants took less than 10 minutes to review the instructions and complete all the tasks; they received a 0.5 GBP participation fee (in addition to the payoffs earned in each game), which is on par with typical wages on Prolific Academic or other Internet marketplaces such as MTurk.

Design

We designed Study 1 as a preliminary test for predicting behavior in 2-player pure coordination games. This study involves a series of (one-shot) games, with each game featuring a 3-element strategy set, such that: each member of a pair receives GBP 0.10 if both players choose the same option; each member of a pair receives nothing otherwise. Figure 1 represents the game structure in bimatrix form (there, for expositional purposes the set of strategies is denoted by $\{X, Y, Z\}$; note that subjects were not provided any such figure). Subjects played 10 instances of the game, with each instance differing from the others only in the names of the three options. Each subject was assigned the same partner for all the (10) games. *No feedback* was provided between games.

We ran two versions of the study: Version A's options consist of names of countries, whereas Version B's options consist of names of food ingredients; both versions are shown in Table 1 below. The reason we designed two versions is to verify that the (presumed) prominence of frequently-mentioned labels does not depend on the characteristics of a specific collection. Note that the option sets for Version A were obtained by drawing member states of the United Nations at random. The option sets for Version

		Player 2		
		X	**Y**	**Z**
	X	0.10 / 0.10	0 / 0	0 / 0
Player 1	**Y**	0 / 0	0.10 / 0.10	0 / 0
	Z	0 / 0	0 / 0	0.10 / 0.10

Figure 1. The coordination game. The bottom-left and top-right numbers in each cell represent the monetary payoffs to Player 1 and Player 2, respectively. (For expositional purposes, the set of strategies is denoted by $\{X, Y, Z\}$; subjects were not provided any such figure.)

Table 1. The option sets for Study 1. The left and right panels refer to Version A and B, respectively. Below each strategy-label is the relative n-gram frequency of that label, computed from the general English corpus including books published after the year 2000. Note: for each option, the reported number is obtained by dividing the NGRAM value of its label by the mean of the values of the three labels in the game (to normalize the data, simply divide each value by 3). For visual clarity, the option with the relatively highest NGRAM value is marked with an "h."

	Version A				Version B		
	[option X]	[option Y]	[option Z]		[option X]	[option Y]	[option Z]
1	Kyrgyzstan 0.4435	Tuvalu 0.1139	Morocco 2.4424 (***h***)	1	mandarin 0.7727	soybean 0.8652	cinnamon 1.3619 (***h***)
2	Turkey 2.1062 (***h***)	Jamaica 0.6537	Senegal 0.2399	2	raspberry 2.0997 (***h***)	sauerkraut 0.6453	scallion 0.2549
3	Yemen 0.6588	Benin 0.4876	Jamaica 1.8535 (***h***)	3	cashew 0.2170	vanilla 1.5978 (***h***)	yogurt 1.1850
4	Spain 2.2128 (***h***)	Angola 0.1943	Norway 0.5927	4	meat 2.5905 (***h***)	cocoa 0.3052	raspberry 0.1041
5	Bolivia 2.1420 (***h***)	Kyrgyzstan 0.5533	Palau 0.3046	5	horseradish 1.0992	rhubarb 1.1067 (***h***)	tarragon 0.7940
6	Afghanistan 0.3663	Germany 2.4533 (***h***)	Ghana 0.1802	6	butter 2.3233 (***h***)	peppermint 0.1339	cocoa 0.5426
7	Nepal 0.9101	Uzbekistan 0.3106	Thailand 1.7791 (***h***)	7	pineapple 1.6064 (***h***)	nutmeg 0.9394	buckwheat 0.4540
8	Bahamas 1.1349	Botswana 1.7162 (***h***)	Tuvalu 0.1487	8	apple 0.9075	milk 1.9335 (***h***)	tuna 0.1588
9	Bahamas 0.6005	Eritrea 0.4705	Ghana 1.9289 (***h***)	9	buckwheat 0.1740	tomato 1.9814 (***h***)	citrus 0.8445
10	Ukraine 0.7379	Jordan 1.9148 (***h***)	Zambia 0.3472	10	brandy 0.9506	pumpkin 0.5337	tomato 1.5156 (***h***)

B were obtained by drawing words at random from a list of ingredients scraped off the Epicurious cooking website (https://www.epicurious.com/).[11]

Participants were randomly assigned to the two versions: 48 subjects were assigned to Version A, and 43 subjects to Version B.[12] Average earnings were GBP 0.63 and 0.52 for Version A and B, respectively (not including subjects' participation fees). Before presenting the experimental results, we shall note that the order of the games was randomized across subjects. By contrast, the order of the three options—in a given game—was de-

[11]It is worth clarifying a general point in relation to our option sets. It is possible that—around the time our study was conducted—some options in our lists were being mentioned in the news or social media more often than usual, a fact that would not be immediately accounted for by the n-gram frequency at the time. Nevertheless, we believe that our *randomized* selection of labels overall controls for any such "random shocks."

[12]Given the odd number of subjects in Version B, one participant was assigned two partners (but received compensation for playing with either one, at random); the two partners were treated like any other participant.

termined prior to the experiment at random, and was *identical* across subjects. More precisely, in each game the strategy-labels were arranged in a column, with options X, Y, and Z of Table 1 being respectively displayed at the top, center, and bottom of the list. For the experimental instructions and screenshots, see Appendix B in the Online Supplementary Material (Sontuoso and Bhatia (2021)).

Results

The average participant in Study 1 chose the strategy associated with the highest, middling and lowest NGRAM value respectively 45.27%, 39.34%, and 15.39% of the time (specifically, subjects selected the option with the highest NGRAM 45.83% of the time in Version A, and 44.65% of the time in Version B on average; for a bar graph of the distributions of individual-level choices in each of the games, see Appendix A). These data patterns are clearly suggestive of a positive impact of the labels' n-gram frequency on strategic play.[13] The following tests provide extensive evidence in support of the strategic use of word frequency, while addressing a potential confound.

One may argue that labels that are displayed in a particular position might be perceived as more salient by some subjects. Thus, we now verify that the impact of word frequency is not confounded by the position of a label on the screen (i.e., top, center, or bottom of the list). To do so, we shall compute the frequency with which a subject chooses the nth position, in games where the nth label *does* and *does not* have the highest NGRAM value.

We start by considering the label displayed first (i.e., at the top of the list). This exercise reveals that when the top label has the highest NGRAM value, it is chosen 55.03% of the time on average; by contrast, the top label is chosen 45.36% of the time whenever it does not have the highest NGRAM value. A two-tailed Wilcoxon sign-ranked test shows that the difference is strongly significant ($N = 91$ obs., $z = 3.042$, $p = 0.002$).[14] Performing the same analysis with respect to the other positions corroborates the trend. In fact, when the center label has the highest NGRAM value it is chosen 36.44% of the time, compared with 21.08% whenever it does not have the highest value ($N = 91$ obs., $z = 5.305$, $p = 0.000$). Furthermore, when the bottom label has the highest NGRAM value it is chosen 43.40% of the time, compared with 15.75% whenever it does not have the highest value ($N = 91$ obs., $z = 6.421$, $p = 0.000$). In summary, *the nth strategy option is*

[13] Interestingly, these summary data are comparable to the behavioral patterns observed in games where one of the strategies is devised by the experimenter as the obvious "odd-one-out." For example, Mehta, Starmer, and Sugden (1994) designed four coordination games where the strategy options are represented by visual objects (i.e., questions no. 17–20, p. 669). In two of those games the odd-one-out was chosen by 44.4% of participants, whereas in the other two games it was chosen by about 64% of participants; in this regard, note that the Mehta, Starmer, and Sugden (1994) games feature a 15-element strategy set, which renders the odd-one-out more prominent (recognizably different) than it would have been in a 3-element strategy set such as ours.

[14] The test uses one observation per subject, consisting of the difference between the two above-described rates (i.e., the frequency with which a subject plays top, in games where the top label has the highest NGRAM value *and* in games where the top label does not have the highest value). Note that the Wilcoxon sign-ranked test is the non-parametric analog to the paired samples t-test.

chosen significantly more often when its label has the highest NGRAM value, compared with when it does not have the highest value.

Moving on, we note that since the three options (in any of our games) constitute symmetric strategies, here Harsanyi and Selten (1988) would argue that the rational solution is to play a "symmetry-invariant" equilibrium, assigning each option the same probability. In their view, such a solution has the benefit of being unique and it ensures that a renaming of the strategies *cannot* ever affect game play (Harsanyi and Selten (1988), pp. 70–74.) In this case, their proposed solution is not supported by the data, as our choice distribution significantly differs from the fully mixed equilibrium strategy profile that assigns equal probability to each strategy ($N = 91$ obs., $T^2 = 127.67$, $p = 0.000$ under a Hotelling's T-squared generalized means test, conducted on the *sample of per-subject mean choices*; note that the Hotelling's test is simply a multivariate generalization of the t-test).[15] Relatedly, we stress that the fully mixed equilibrium implies a coordination rate of 0.33, whereas the expected coordination rate resulting from our sample is roughly 0.50 on average (specifically, 0.57 for Version A and 0.49 for Version B; this means that the payoffs earned by participants in Version A and B are respectively 72% and 48% higher than the payoffs subjects would earn by coordinating on the fully mixed equilibrium).[16,17]

In concluding, we note that we designed Study 1 as a preliminary test for predicting behavior in coordination games. The method of analysis employed so far has involved mean observations, thereby discarding a fair amount of information. To shed light on the strategic use of word frequency, below we consider some between-subjects designs; we then perform a more sophisticated analysis in such a way to account for the characteristics of each triplet of labels.

4. Study 2

Demographics

The subject pool for Study 2 consisted of 160 US resident individuals. The average participant was 30 years old, and 58% of the subjects were male. As with Study 1, participants

[15]The sample of (per-subject) mean observations is obtained as follows. First, for each choice of subject i, assign a value of 1 or 0 to indicate if the option with the *highest* NGRAM was chosen or not; then, take the average across all the games played by subject i. Similarly, assign a value of 1 or 0 to indicate if the option with the *middling* NGRAM was chosen or not, and take the average across games. The same applies to the option with the *lowest* NGRAM.

[16]In keeping with previous studies, we report expected coordination rates (as opposed to actual frequencies of coordination; see, for example, Mehta, Starmer, and Sugden (1994), and Crawford, Gneezy, and Rottenstreich (2008)). In fact, actual frequencies of coordination are affected by the eventual pairing of partners, thereby resulting in a biased metric in smaller samples.

[17]The expected coordination rate gives the probability that two randomly drawn subjects choose the same strategy in a randomly selected game. For each version of Study 1, we calculate this rate using Monte Carlo methods. The pseudocode is as follows: (A) pick two participants at random; (B) pick one of the 10 games at random; (C) if both participants chose the same strategy consider it an instance of successful coordination, otherwise consider it unsuccessful; (D) repeat steps A–C 100,000 times, then calculate the relative frequency of successful coordination.

took less than 10 minutes to review the instructions and complete all the tasks; they received a 0.5 GBP participation fee, in addition to the payoffs earned in each game (if any), as specified below.

Design

The objective of the study is to test for the strategic use of word frequency by systematically varying players' incentives. To that end, we designed a few variants of the simple coordination game of Study 1, in such a way to incentivize or disincentivize coordination for either player. In order to check for replicability, Study 2 features the exact same option triplets as in Version A of Study 1 (see the left panel of Table 1 above).[18] That being said, Study 2 involves the following four roles/conditions.

a. ***Coordinate***: This is a replication of Version A of Study 1 that, among other purposes, serves to verify the robustness of our earlier results. Participants in this condition ("Coordinators") were paired with other participants in the same condition, and were so informed.

b. ***Pick***: Participants in this condition ("Pickers") were presented with the same labels as in the Coordinate condition, except that they were asked to merely pick one of the three given options. That is, participants were not assigned a partner, nor did they receive any additional payoffs on the basis of their choices; hence, they had no strategic incentive to select one label over another.

c. ***Seek***: This condition features the same strategy-labels as in the Coordinate condition, except that the incentive structure reflects the role of "Seeker" in the Hide-and-Seek game—Figure 2 below represents the payoff structure of this game. As can be

		Seeker		
		X	**Y**	**Z**
Hider	**X**	0.10 0	0 0.10	0 0.10
	Y	0 0.10	0.10 0	0 0.10
	Z	0 0.10	0 0.10	0.10 0

FIGURE 2. The Hide-and-Seek game structure in bimatrix form. The bottom-left and top-right numbers in each cell represent the monetary payoffs to the *Hider* and *Seeker*, respectively.

[18]Note that here we focused on Version A—that is, names of countries—simply to be economical; in Study 3 below, we will resume investigating the impact of word frequency in the context of food ingredients.

seen there, a Seeker receives GBP 0.10 if both members of a pair choose the same strategy, and nothing otherwise. Participants in this condition were paired with participants in the Hide condition below, and were so informed.

d. ***Hide***: Again, this condition features the same strategy-labels as in the Coordinate condition, but the incentive structure reflects the role of "Hider" in the Hide-and-Seek game. As can be seen in Figure 2, a Hider receives GBP 0.10 if members of a pair choose different strategies, and nothing otherwise. Participants in this condition were paired with participants in the Seek condition above, and were so informed.

Each subject completed 10 tasks in the same role/condition, with each task differing from the others only in the names of the three options (see the left panel of Table 1 above). In the Coordinate, Seek, and Hide conditions, each subject was assigned the same partner for all the (10) games, and was so informed.[19] *No feedback* was provided between games.

Before discussing our predictions, we note that (as with Study 1) the order of the tasks/games was randomized across subjects. By contrast, the order of the three options—in a given task—was determined prior to the experiment at random, and was *identical* across subjects and conditions. Specifically, the labels were arranged in a column, with options X, Y, and Z of Table 1 (left panel) being respectively displayed at the top, center, and bottom of the list. For the experimental instructions and screenshots, see Appendix B.

Since all our conditions involve the exact same labels, if the effect of word frequency were merely due to an automatic (or naïve) response, then we should observe similar choice distributions across conditions. If instead participants used word frequency in a strategic manner, then frequently-mentioned labels should be more (less) likely selected when there are incentives to match (mismatch) others, with the magnitude of the changes varying with subjects' strategic sophistication. Below we model our intuition with a particular specification of level-k reasoning (Nagel (1995); Stahl and Wilson (1995); Costa-Gomes, Crawford, and Broseta (2001); Camerer, Ho, and Chong (2004); Costa-Gomes and Crawford (2006)).

Level-k theories posit a hierarchy of player types defined by the level of sophistication with which each player reasons. Specifically, level-k types anchor their beliefs in a nonstrategic $L0$ type and adjust them via iterated best responses, so that $L1$ players best respond to $L0$ players, $L2$ best respond to $L1$, and so on.[20] In what follows, we implement this approach by formulating a set of assumptions that are relevant for our

[19] Given that in the *Hide* condition there were less participants than in the *Seek* condition, (for the mere purpose of calculating the payoffs of the extra Seekers) nine Hiders were matched with two Seekers, but received compensation for playing with either one at random.

[20] Note that some applications of level-k reasoning differ in their assumptions as to whether there are actually any players at $L0$ (as will be clear, such assumptions do not qualitatively affect our predictions). Applications further differ in their assumptions as to the players' randomizing behavior at $L0$. Another element in regard to which models differ is the players' depth of reasoning about other types; in particular, some models assume that players at each level above $L0$ best respond to a probability mix of the decisions of all levels below their own, as opposed to best responding to the one level immediately below. For discussion, see Crawford, Costa-Gomes, and Iriberri (2013) and Mauersberger and Nagel (2018).

setting; this exercise will formally generate our predictions. In short, we shall assume that:

(i) A nonstrategic $L0$ type in the *Coordinate*, *Seek*, or *Hide* conditions behaves like a participant in the *Pick* condition;

(ii) Players at levels *above* $L0$ believe that the distribution q of $L0$ choices has a peak at the option with the highest NGRAM value;

(iii) There are no players at $L3$ or higher.

A few comments are in order. We note that level-k theories commonly assume that $L0$ types do not engage in strategic reasoning, but simply randomize between options according to some probability distribution q. In particular, Crawford and Iriberri (2007) assumed that such a distribution is nonuniform—positing that $L0$ types are relatively more likely to select salient labels compared with other labels—without however defining salience in general terms.

For the purposes of generating hypotheses, we partly sidestep the issue of specifying the probability distribution q by defining it empirically (on the basis of Pickers' behavior), as per assumption (i). Since our subjects obviously do not observe that distribution, we then posit that players at levels *above* $L0$ believe that "$L0$ types are more likely to select the option with the highest NGRAM than to select any other option," as per assumption (ii). Lastly, we note that (iii) is a simplifying assumption, based on previous empirical evidence about subjects' strategic sophistication. For example, Arad and Rubinstein (2012) noted that level-k experiments have shown that «the most frequent types are usually $L1$ and $L2$, whereas higher-level types are rare» (p. 3561). For a similar point, see also Penczynski's (2016) analysis of Hide-and-Seek games.

What does the above entail in terms of behavioral predictions? As usual, $L1$ types will best respond to (their beliefs about) $L0$ behavior; $L2$ types will accordingly adjust their beliefs, and best respond to $L1$ types. Specifically, $L1$ Coordinators will best respond to their beliefs about $L0$ behavior (which in our case are defined by assumption (ii)) and, therefore, will select the option with the highest NGRAM with probability one. Then $L2$ Coordinators will best respond to $L1$ Coordinators, thereby selecting the option with the highest NGRAM (with probability one). Moving on, we note that Seekers wish to select the label believed to be the modal choice of Hiders at the level below, whereas Hiders wish to avoid the label believed to be the modal choice of Seekers at the level below: here, this implies that the option with the highest NGRAM will be respectively selected by $L1$ Hiders and $L1$ Seekers with probability zero and one; hence, the option with the highest NGRAM will be selected by both $L2$ Hiders and $L2$ Seekers with probability zero.

In summary, the experiment aims to verify if subjects' behavior is compatible with a strategic use of word frequency. If it were not, we should observe the same choice distributions across roles/conditions. If instead subjects used word frequency in a (boundedly) rational manner, then—based on the assumptions above—behavior should vary across roles as follows:

H1: *Coordinators select the most frequently-mentioned label as often as* (*or more often than*) *Pickers.*

H2: *Coordinators select the most frequently-mentioned label more often than Seekers.*

H3: *Seekers* (*and Pickers*) *select the most frequently-mentioned label more often than Hiders.*

Formally, the assumptions above imply that the "Coordinators' probability of selecting the most frequently-mentioned label" (i.e., the option with the highest NGRAM) is defined by $p_{\text{COORD}} = (l_0 \cdot q_H) + (l_1 \cdot 1) + (l_2 \cdot 1)$, where l_k denotes the share of Lk players in our subject pool while q_H denotes the frequency with which $L0$ types select the option with the highest NGRAM (as defined by assumption (i)). Thus, a term in the expression above represents the probability that the relevant share of Lk Coordinators in our subject pool select the option with the highest NGRAM. (Incidentally, we stress that $L0$ behavior by definition is the same across roles.) Next, the "Seekers' probability of selecting the option with the highest NGRAM" is given by $p_{\text{SEEK}} = (l_0 \cdot q_H) + (l_1 \cdot 1) + (l_2 \cdot 0)$, where the second and third terms refer to $L1$ and $L2$ Seekers, respectively. Further, the "Hiders' probability of selecting the option with the highest NGRAM" is given by $p_{\text{HIDE}} = (l_0 \cdot q_H) + (l_1 \cdot 0) + (l_2 \cdot 0)$, where the second and third terms, respectively, refer to $L1$ and $L2$ Hiders. Now, assuming that there are strategic players in our subject pool (i.e., $l_1, l_2 > 0$, with $l_0, l_1, l_2 \in (0, 1)$ and $l_0 + l_1 + l_2 = 1$), then—under the standard assumption that the distribution of types is the same across roles/conditions—the above entails that $p_{\text{COORD}} > p_{\text{SEEK}} > p_{\text{HIDE}}$ and $p_{\text{COORD}} \geq p_{\text{PICK}} \geq p_{\text{HIDE}}$, where p_{PICK} denotes the "Pickers' probability of selecting the option with the highest NGRAM," with $p_{\text{PICK}} \equiv q_H$ as per assumption (i).[21] Also, the above implies that the likelihood of choosing the option with the lowest NGRAM will possibly rise when moving from Coordinate (or Pick) to Seek to Hide.

Results

Table 2 presents mean choices in each of the four roles/conditions, given a classification of the strategy-labels based on their relative n-gram frequency, as per Table 1 above. (For a bar graph of the distributions of individual-level choices in each of the games and conditions, see Appendix A.) By taking a glance at Table 2, the reader will notice that the mean distribution of choices varies with each condition. In fact, the *most* frequently-mentioned label (i.e., the option with the highest NGRAM) was chosen less and less often when moving from Coordinate (or Pick) to Seek to Hide. Consequently, the *least* frequently-mentioned label (i.e., the option with the lowest NGRAM) was chosen more and more often when moving from Coordinate (5.71%) to Pick (13.33%) to Seek (22.27%) to Hide (30.29%). Taken together, these patterns seem to confirm that subjects

[21]The weak inequalities are due to the fact that the modeler has no a-priori knowledge of q_H, with $q_H \in [0, 1]$. On a different note, we stress that since our focus is on behavioral comparisons across roles (as opposed to identifying the empirical distribution of levels), we need not make any further assumptions in order to generate our hypotheses.

Table 2. (Per-subject) mean choices, given a classification of the labels based on their relative n-gram frequency; in parentheses is the standard deviation. Note: the number of triplets is obtained by multiplying the number of tasks/games (i.e., 10) by the number of participants in each role/condition.

Choice by word frequency	*Coordinate*	*Pick*	*Seek*	*Hide*
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	46.19 (0.1464)	48.46 (0.1646)	43.64 (0.1556)	36.86 (0.1548)
Strategy-label with *middling* NGRAM metric is chosen [f_M], %	48.10 (0.1596)	38.21 (0.1211)	34.09 (0.1661)	32.85 (0.1202)
Strategy-label with *lowest* NGRAM metric is chosen [f_L], %	5.71 (0.0914)	13.33 (0.1675)	22.27 (0.2044)	30.29 (0.2121)
Total, %	100	100	100	100
Total # triplets (1600)	420	390	440	350
Subjects (160)	42	39	44	35

used word frequency in a (boundedly) rational manner whereby the higher the word frequency of a label, the lower the likelihood of choosing that option when moving from Coordinate to Hide. Put differently, the lower the word frequency of a label, the higher the likelihood of choosing that option when moving from Coordinate to Hide. In what follows, we further examine these trends.

We begin by reporting a Kruskal–Wallis test, which confirms significant differences in the choice of the option with the *highest* NGRAM across conditions ($N = 160$ obs., $\chi^2_3 = 10.477$, $p = 0.014$, two-tailed; note that in order to satisfy the assumption of independence of observations, all of our non-parametric tests are conducted on the sample of per-subject mean choices, as described in footnote 15). Similarly, the same test confirms significant differences in the choice of the option with the *lowest* NGRAM across conditions ($N = 160$ obs., $\chi^2_3 = 34.353$, $p = 0.000$, two-tailed).[22,23]

In order to address our hypotheses, later on we report a formal econometric analysis that accounts for the characteristics of each and every label. Before doing so—to provide a rough outline of the data—we present pairwise non-parametric tests conducted

[22]Also, Hotelling's T-squared generalized means tests (conducted on the samples of per-subject mean choices) reveal that the distribution of choices in each of the Coordinate, Pick, and Seek conditions differs from the fully mixed equilibrium assigning equal probability to all strategies (for *Coordinate*: $N = 42$ obs., $T^2 = 389.71$, $p = 0.000$; for *Pick*: $N = 39$ obs., $T^2 = 55.72$, $p = 0.000$; for *Seek*: $N = 44$ obs., $T^2 = 20.66$, $p = 0.000$). However, the same test shows that the distribution of choices in the *Hide* condition does not differ from the fully mixed equilibrium, a fact that might be interpreted as indirect evidence in support of H3 (we shall test that hypothesis below).

[23]Incidentally, we note that the expected coordination rate resulting from our sample is 0.44 (for subjects in a *Coordinator* role). This implies that, on average, the payoff to our Coordinators is 33% higher than the payoff subjects would obtain by playing the fully mixed equilibrium. In the case of the Hide-and-Seek game, the average payoff to subjects in a *Seeker* role is about 2% higher than the payoff that would be obtained by a hypothetical Seeker who randomizes uniformly over labels; then, the average payoff to subjects in a *Hider* role is about 1% lower than the payoff that would be obtained by a hypothetical Hider who randomizes uniformly over labels.

on the sample of mean observations (i.e., the tests use one observation per participant). We start by comparing behavior in the *Coordinate* and *Pick* conditions: a one-tailed test allows us to check the "alternative hypothesis" that the most frequently-mentioned label (i.e., the option with the highest NGRAM) is selected strictly less often in Coordinate than in Pick.[24] In short, a one-tailed Wilcoxon–Mann–Whitney test shows no evidence of a significant decrease in the choice of the *most* frequently-mentioned label when moving from Pick to Coordinate ($N = 81$ obs., $Z = -1.008$, $p = 0.1567$). Conversely, a one-tailed Wilcoxon–Mann–Whitney test provides evidence of a significant decrease in the choice of the *least* frequently-mentioned label when moving from Pick to Coordinate ($N = 81$ obs., $Z = -2.726$, $p = 0.003$). These tests provide some preliminary evidence in support of H1.

A similar test shows no significant difference between the *Coordinate* and *Seek* conditions with respect to the choice of the *most* frequently-mentioned label. Yet, a one-tailed Wilcoxon–Mann–Whitney test provides evidence of a significant increase in the choice of the *least* frequently-mentioned label when moving from Coordinate to Seek ($N = 86$ obs., $Z = -4.235$, $p = 0.000$). The latter result might be viewed as indirect evidence in support of H2, which therefore warrants further testing: the econometric analysis will later clarify these patterns.

Non-parametric tests also show that the *most* frequently-mentioned label was selected more often in *Pick* than in *Hide*, providing some preliminary support for H3 ($N = 74$ obs., $Z = 3.101$, $p = 0.000$, one-tailed Wilcoxon–Mann–Whitney test). Similarly, the *most* frequently-mentioned label was selected significantly more often in *Seek* than in *Hide* ($N = 79$ obs., $Z = 1.978$, $p = 0.023$, one-tailed Wilcoxon–Mann–Whitney test), which again supports H3. (Unsurprisingly, the same test shows that the *least* frequently-mentioned label was selected significantly more often in Hide than in Pick, and significantly more often in Hide than in Seek.)

In summary, our non-parametric tests confirm a trend where the higher the word frequency of a label, the lower the likelihood of choosing that option when moving from Coordinate (or Pick) to Hide. Put differently, there is a trend where the lower the word frequency of a label, the higher the likelihood of choosing that option when moving from Coordinate to Pick to Seek to Hide. However, we note that since the above tests are conducted on the sample of (per-subject) mean choices, they do not account for differences in the actual magnitude of the labels' n-gram frequency across options.

For the reasons above, we shall corroborate our findings by reporting the results of an alternative-specific conditional logit analysis ("asclogit"; i.e., McFadden's choice model (1973)). This analysis will provide the ultimate test of our hypotheses: below we report the main results while we refer the reader to Appendix A for the full econometric tables. In brief, McFadden's (1973) choice model is a special case of conditional logit analysis where independent variables come in two forms: alternative- and case-specific. *Alternative-specific* variables represent attributes that may vary across each of the options in a choice task (e.g., the labels' numerical NGRAM value varies across any three

[24]Note that H1 says that Coordinators select the most frequently-mentioned label as often as (or more often than) Pickers. Because of the weak inequality, here we shall test against the alternative hypothesis that the most frequently-mentioned label is selected strictly less often in Coordinate than in Pick.

options). *Case-specific* variables, on the other hand, represent attributes that are common to each of the options in a choice task (e.g., the same experimental condition characterizes the three options a subject is provided in a choice task). That said, our random-utility model can be expressed as $\boldsymbol{u}_i = \mathbf{X}_i\boldsymbol{\beta} + (z_i\mathbf{A})' + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\beta}$ is a vector of alternative-specific coefficients and $\mathbf{A}$ is a matrix of case-specific coefficients.[25] In particular, we consider a model consisting of the following predictors: (i) the word frequency of the labels, as measured by their numerical NGRAM value; (ii) the experimental condition; (iii) the interaction of (i) and (ii). Note that the latter is an alternative-specific variable and is the key to testing our hypotheses.

The analysis confirms a significant positive effect of word frequency on choice behavior across conditions; that is, the higher the n-gram frequency of a label, the more likely it is for that option to be selected (regardless of the label's position on the screen). Further, when contrasting behavior in *Coordinate* against *Pick*, the model indicates no significant difference in the relative impact of word frequency between these conditions (see interaction variable *WFC* in model [1] of Table A1, in Appendix A). This implies that between Coordinators and Pickers there is no difference in the probability of choosing frequently-mentioned labels, which confirms the previous evidence in regards to H1.

Next, when contrasting behavior in *Coordinate* against *Seek*, we find a significant difference in the relative impact of word frequency on choice. Specifically, we find that a label with a higher NGRAM value is more likely to drive the choices of Coordinators than Seekers (coef. $= -0.119$, $z = -2.44$, $p = 0.015$, two-tailed asclogit conducted on the sample of individual observations, with standard errors adjusted for clustering on 86 subjects; see variable *WFC* in model [2] of Table A1). The result evidently supports H2.

When contrasting behavior in *Pick* against *Hide*, we find again a significant difference in the relative impact of word frequency on choice; that is, a label with a higher NGRAM value is more likely to drive the choices of Pickers than Hiders (coef. $= -0.201$, $z = -3.77$, $p = 0.000$, two-tailed asclogit conducted on the sample of individual observations, with standard errors clustered on 74 subjects; see variable *WFC* in model [3] of Table A2). Similarly, when contrasting behavior in *Seek* against *Hide*, we find a significant difference in the relative impact of word frequency on choice: a label with a higher NGRAM value is more likely to drive the choices of Seekers than Hiders (coef. $= -0.2145$, $z = -2.14$, $p = 0.033$, two-tailed asclogit conducted on the sample of individual observations, with standard errors clustered on 79 subjects; see variable *WFC* in model [4] of Table A2). The results support H3.

To conclude, the data provide strong support for our hypotheses (please refer to Appendix A for an extended commentary on the econometric analysis). Despite the fact that our conditions involve the same option triplets, frequently-mentioned labels were selected less often when moving from *Coordinate* to *Seek*, from *Pick* to *Hide*, and from *Seek* to *Hide*. These data patterns confirm that individuals select strategies that fulfill

[25] Assume that the modeler considers p alternative-specific variables; so, for a generic choice task i there is a $J \times p$ matrix ($\mathbf{X}_i$), with J denoting the number of labels in the task (i.e., $J = 3$). Further, assume that the modeler considers q case-specific variables, so for choice task i there is a $1 \times q$ vector (z_i). Thus, in the random-utility model above, $\boldsymbol{\beta}$ is a $p \times 1$ vector of alternative-specific coefficients and $\mathbf{A}$ is a $q \times J$ matrix of case-specific coefficients.

our characterization of prominence, and they do so in a (boundedly) rational manner consistent with our level-k specification. In the remainder of the article, we shall delve into the strategic use of labels in pure coordination games.

5. Study 3

Demographics

The subject pool for Study 3 consisted of 80 individuals, of which half were US residents and half were UK residents. Participants were recruited through the Prolific Academic platform, at the same time. In the US-residents sample, the average participant was 29 years old, and 75% of the subjects were male. In the UK-residents sample, the average participant was 31 years old, and 63% of the subjects were male. As with our previous studies, participants took less than 10 minutes to review the instructions and complete all the tasks; they received a 0.5 GBP participation fee, in addition to the payoffs earned in each game, as specified below.

Design

In what follows, we describe a set of treatments that are intended to further illuminate the strategic use of labels in coordination games. The present study involves exactly the same pure coordination structure as in Study 1 (see Figure 1 above), except that this time our games feature different labels than before. Unlike our previous studies—which featured randomly generated labels—in this case, we purposely selected triplets of labels so that the option with the highest word frequency differs between the American- and British-English vocabularies (as measured by the relevant NGRAM in the American- and British-English Google Books corpora).

Specifically, subjects played 10 instances of the coordination game in Figure 1, with each instance differing from the others only in the names of the three options (displayed below). Each subject was assigned the same partner for all the (10) games, and was so informed. *No feedback* was provided between games. The order of the games was randomized across subjects. Instead, the order of the three options—in a given game—was determined prior to the experiment at random, and was *identical* across subjects and conditions. More precisely, in each game the strategy-labels were arranged in a column, with options X, Y, and Z of Table 3 being respectively displayed at the top, center, and bottom of the list. For the experimental instructions and screenshots, see Appendix B.

As we previously noted, a key feature of this experiment is that we varied the cultural makeup of the subject pool (by recruiting samples of US residents and of UK residents). Additionally, we manipulated the subjects' perception of their counterpart's cultural/linguistic background: we did so by providing subjects with different information as to their partner's country of residence. That is, each participant—whether in the US or in the UK sample—was assigned to one of the following "information conditions."

a. ***NO-Info***: Participants in this condition received no information about their partner's country of residence (hence, this condition is identical to Study 1, except that

TABLE 3. The option sets for Study 3. The first and second row below the strategy-labels respectively report the American-English and British-English relative n-gram frequency of the corresponding word, for books published after the year 2000. For visual clarity, the option with the relatively highest NGRAM value is marked with an "h."[26]

	[option X]	[option Y]	[option Z]		[option X]	[option Y]	[option Z]
1	paprika	curry	chili	6	sardines	tuna	cod
US	0.5718	0.3859	0.6963 (***h***)	US	0.4648	0.6225 (***h***)	0.4439
UK	0.4281	0.6140 (***h***)	0.3036	UK	0.5351	0.3774	0.5560 (***h***)
2	Bordeaux	Chianti	Syrah	7	cheesecake	scones	tiramisu
US	0.4784	0.6574 (***h***)	0.5222	US	0.6852 (***h***)	0.4438	0.5849
UK	0.5215 (***h***)	0.3425	0.4777	UK	0.3147	0.5561 (***h***)	0.4150
3	venison	lamb	pork	8	burrito	panini	kebab
US	0.5240	0.4543	0.5816 (***h***)	US	0.7599 (***h***)	0.3145	0.3579
UK	0.4759	0.5456 (***h***)	0.4183	UK	0.2400	0.6854 (***h***)	0.6420
4	peach	pineapple	pear	9	parsnip	beetroot	shallot
US	0.5110	0.5731 (***h***)	0.4359	US	0.5492	0.2788	0.7001 (***h***)
UK	0.4889	0.4268	0.5640 (***h***)	UK	0.4507	0.7211 (***h***)	0.2998
5	blueberry	blackberry	gooseberry	10	oatmeal	porridge	granola
US	0.6741 (***h***)	0.4838	0.2700	US	0.5838	0.3529	0.8263 (***h***)
UK	0.3258	0.5161	0.7299 (***h***)	UK	0.4161	0.6470 (***h***)	0.1736

we used different strategy-labels). The instructions in this condition stated: *"Please choose one option. Each of you and your partner receive* £0.10 *if you both choose the same option,* £0 *otherwise."*

b. ***Know-UK***: Participants in this condition were told that their partner resided in the UK. Specifically, subjects were shown the following message: *"Please choose one option. Each of you and your partner receive* £0.10 *if you both choose the same option,* £0 *otherwise. Your partner is a Prolific worker who resides in the UK. Your partner may or may not know where you reside."*

c. ***Know-US***: Participants in this condition were told that their partner resided in the US. Specifically, subjects were shown the following message: *"Please choose one option. Each of you and your partner receive* £0.10 *if you both choose the same option,* £0 *otherwise. Your partner is a Prolific worker who resides in the US. Your partner may or may not know where you reside."*[27]

[26]For any given label, the number reported in the first row ("US") below the label is obtained by performing the following operations: (i) divide the American-English NGRAM value of the label by the sum of the values of the three labels in the game; (ii) perform the same operation as before, except this time use British-English NGRAM values; (iii) divide the outcome of (i) by the sum of the outcomes of (i) and (ii). The rationale behind step (iii) is to ensure that our measure of word frequency is a function of both vocabularies, and hence applies to all the subjects, regardless of their information condition. Finally, note that—for each label—the numbers reported in the first and second row immediately below the label add up to 1.

[27]The reason we tell subjects that their partner may or may not know where they reside is the following. For the purpose of calculating coordination rates via Monte Carlo methods, such a wording permits us to

In summary, we recruited samples of US and UK residents; each subject was then randomly assigned to one of the three information conditions, irrespective of the subject's own residence. We stress that if the effect of word frequency in coordination games were due to an automatic (or naïve) response, then we should observe similar choice distributions across conditions and countries. If however subjects used word frequency in a strategic manner, then we should find that options with higher NGRAM values (in the subjects' respective vocabularies) are more often selected when subjects think that their partner is culturally alike. In that case, subjects would realize that culturally alike people view the problem through the same lens: if a label comes to mind easily to a subject, then she may realize that it will also come to mind easily to others with the same vocabulary. This leads to the following hypotheses:

H4: *Choice behavior differs between US and UK residents, and it is positively related to the labels' n-gram frequency in the vocabulary of the respective countries.*

H5: *Subjects who are informed that their partner resides in the same country select the most frequently-mentioned label* (*in their vocabulary*) *more often than subjects who are unaware of their partner's residence. In turn, the latter select the most frequently-mentioned label more often than subjects who are informed that their partner resides in a different country.*

We note that the hypotheses above are consistent with a level-k specification allowing for multiple (alternative) $L0$ types, whereby the modal choice of each $L0$ type corresponds to the option with the highest NGRAM value in that type's vocabulary. Given this, players at $L1$ best respond to a convex combination of their beliefs about each $L0$ type; as usual, players at $L2$ then best respond to (their beliefs about) $L1$ behavior, and so on.

Results

Table 4 presents mean choices, given a classification of the strategy-labels based on their n-gram frequency as follows. The left panel provides summary data by pooling all the choices, across our two country samples and three information conditions. Specifically, note that the left panel classifies strategy options based on the subjects' vocabularies (i.e., *American*- and *British*-English, respectively, for US and UK participants, as shown in Table 3), regardless of their information condition. Thus, the table gives us an idea of the overall impact of word frequency: as can be seen in the left panel, a plurality of the choices (41.25%) consists of the option with the highest NGRAM value in the responding subject's vocabulary. A Hotelling's T-squared generalized means test (conducted on the sample of per-subject mean choices) confirms that the distribution of choices differs from the fully mixed equilibrium assigning equal probability to all the strategies ($N = 80$ obs., $T^2 = 15.14$, $p = 0.001$). Moreover, a two-tailed Wilcoxon signed-rank test indicates

virtually match each subject with any participants in the relevant country (regardless of those participants' information condition).

TABLE 4. (Per-subject) mean choices, given a classification of the labels based on their n-gram frequency; in parentheses is the standard deviation. Note: the number of triplets equals the number of games (i.e., 10) times the number of participants in each condition. The left panel presents summary data by pooling choices from the US and UK samples (in each country sample, strategy options are ranked by word frequency in the respective vocabulary). For the sole purpose of comparing distributions across samples, the right panel breaks down the data by country, given a classification of the labels based on the American-English corpus.

	Labels' word frequency refers to the relevant vocabulary (i.e., *American*- and *British*-English for US and UK residents, resp.)	Labels' word frequency refers to the *American*-English vocabulary	
Choice by word frequency	Pooled country samples	US residents	UK residents
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	41.25 (0.1951)	47.25 (0.1986)	38.75 (0.1785)
Strategy-label with *middling* NGRAM metric is chosen [f_M], %	28.13 (0.1599)	30.25 (0.1656)	26.00 (0.1532)
Strategy-label with *lowest* NGRAM metric is chosen [f_L], %	30.62 (0.1871)	22.50 (0.1597)	35.25 (0.1739)
Total, %	100	100	100
Total # triplets	800	400	400
Subjects	80	40	40

that the frequency of choice of the option with the highest NGRAM value is significantly different from chance ($N = 80$ obs., $z = 3.467$, $p = 0.000$).

For the purpose of comparing choice distributions between countries, the right panel of Table 4 classifies strategy options based on the labels' word frequency in the *American-English* vocabulary only. By taking a look at the right panel (Table 4), the reader will notice that the distribution of choices varies by country. In particular, the higher the n-gram frequency of a label in the American-English vocabulary, the more likely it is for the associated option to be selected by participants in the US rather than in the UK (remarkably, the option with the highest American-English NGRAM was chosen 47.25% and 38.75% of the time by US and UK residents, resp.). A two-group Hotelling's T-squared generalized means test (conducted on the samples of per-subject mean choices) confirms that the US and UK choice distributions differ from each other ($N = 80$ obs., $T^2 = 11.665$, $p = 0.004$). (The econometric analysis below will delve into the British-English vocabulary as well.)

To sum up, the above provides evidence in support of H4: behavior differs between US and UK residents and, overall, it is positively related to the labels' n-gram frequency in the vocabulary of the respective countries. (For a more granular breakdown of the data with respect to both the *American*- and *British*-English vocabularies, see Table A3 in Appendix A.) Later on, we will corroborate our findings by discussing some robustness checks.

We move on to address H5, which concerns differences across information conditions (as opposed to differences across country samples). To that end, Table 5 below

TABLE 5. (Per-subject) mean choices, given a classification of the labels based on the relevant n-gram frequency (i.e., with respect to the American- and British-English vocabularies for US and UK residents, respectively, regardless of their information condition). In parentheses is the standard deviation. Note: "Know-SAME" includes US residents in the know-US condition, and UK residents in the know-UK condition. Participants in "NO-Info" received no information about the partner's country of residence. "Know-OTHER" includes US residents in the know-UK condition, and UK residents in the know-US condition.

Choice by word frequency	*Know-SAME*	*NO-Info*	*Know-OTHER*
Strategy-label with *highest* NGRAM metric is chosen [f_H], %	50.00 (0.1809)	41.62 (0.1818)	30.50 (0.1904)
Strategy-label with *middling* NGRAM metric is chosen [f_M], %	29.13 (0.1621)	27.29 (0.1627)	28.50 (0.1598)
Strategy-label with *lowest* NGRAM metric is chosen [f_L], %	20.87 (0.1311)	31.09 (0.1882)	41.00 (0.1889)
Total, %	100	100	100
Total # triplets	230	370	200
Subjects	23	37	20

divides the data into three mutually exclusive groups. The first group ("know-SAME") consists of participants who were informed that their assigned partner resided in the same country (i.e., US participants in the *know-US* condition, and UK participants in the *know-UK* condition). The second group consists of subjects who were not informed about their partner's residence, and corresponds to all the participants in the *No-info* condition. The third group ("know-OTHER") consists of participants who were informed that their assigned partner resided in a different country (i.e., US participants in the *know-UK* condition, and UK participants in the *know-US* condition).

As can be seen in Table 5 above, the mean distribution of choices varies with each group. A quick glance reveals that the *most* frequently-mentioned label (i.e., the option with the highest NGRAM in the responding subject's vocabulary) was chosen less and less often when moving from know-SAME (50.00%) to No-info (41.62%) to know-OTHER (30.50%). A Kruskal–Wallis test (conducted on the sample of per-subject mean choices) confirms significant differences in the choice of the option with the highest NGRAM across groups ($N = 80$ obs., $\chi^2_2 = 10.597$, $p = 0.005$, two-tailed). Further, Table 5 reveals that the *least* frequently-mentioned label (i.e., the option with the lowest NGRAM) was chosen more and more often, moving from know-SAME (20.87%) to No-info (31.09%) to know-OTHER (41.00%), with such differences being again significant ($N = 80$ obs., $\chi^2_2 = 12.032$, $p = 0.002$, two-tailed Kruskal–Wallis test).[28] The above provides some preliminary evidence in support of H5.

[28]Also, the distribution of choices in each of the know-SAME and No-info groups differs from the fully mixed equilibrium assigning equal probability to all the strategies, as is confirmed by Hotelling's T-squared generalized means tests conducted on the samples of mean choices (for *know-SAME*: $N = 23$ obs., $T^2 = 26.91$, $p = 0.000$; for *No-info*: $N = 37$ obs., $T^2 = 9.21$, $p = 0.018$). However, we find no such difference in *know-OTHER* ($N = 20$ obs., $T^2 = 3.73$, $p = 0.199$). The latter suggests that subjects picked an option at

The tests presented so far provide some insights into the patterns that emerge from a dataset consisting of average choices (i.e., to satisfy the assumption of independence of observations, the tests above are conducted on the sample of per-subject mean choices, as described in footnote 15). We now proceed to corroborate our findings by conducting an econometric analysis of the full sample of individual observations, while adjusting standard errors for clustering on the subjects.

We start by investigating the relationship between one's use of word frequency and one's knowledge of the counterpart's country of residence. To that end, we coded a "knowledge" ordinal variable as follows: this variable takes on value 1 if a subject is in the *know-OTHER* group (i.e., a subject knows that the partner resides in a different country); it takes on value 2 if a subject is in the *NO-info* condition (i.e., a subject receives no information about the partner's country of residence); it takes on value 3 if a subject is in the *know-SAME* group (i.e., a subject knows that the partner resides in the same country). We then coded a binary "NGRAM prediction" variable, which takes on value 1 if a subject selects the option with the highest NGRAM value in her own vocabulary, and takes on value 0 otherwise. A logit model consisting of the *NGRAM prediction* as the binary dependent variable (and of the knowledge variable as the sole predictor) confirms a significant positive effect of the knowledge variable on the choice of the option with the highest NGRAM (for the *US sample*, coef. $= 0.4529$, $z = 2.65$, $p = 0.008$, two-tailed logit with standard errors clustered on 40 subjects; for the *UK sample*, coef. $= 0.3851$, $z = 2.44$, $p = 0.015$, two-tailed logit with standard errors clustered on 40 subjects). The above confirms that—in each of our two country samples—subjects used word frequency in a strategic manner. The more they had reason to believe that their partner was alike, the more likely they were to select the most frequently-mentioned label in their respective vocabularies. So, these results strengthen our previous evidence in support of H5.

We turn to the next robustness checks. (Here, we report some key findings while we refer the reader to Appendix A for the econometric tables and further commentary.) In order to control for the magnitude of the labels' NGRAM value, we performed an alternative-specific conditional logit analysis (McFadden's choice model (1973); see footnote 25 above) as with Study 2. The analysis confirms a positive effect of word frequency on choice behavior: the higher the n-gram frequency of a label in the responding subject's vocabulary, the more likely it is for that option to be selected (regardless of the label's position on the screen). Moreover, when comparing behavior in the *know-SAME* and *know-OTHER* groups, the model confirms significant differences in the relative impact of word frequency; specifically, if one is informed that the partner resides in a different country, one is less likely to select options with higher NGRAM values in one's own vocabulary (see Table A4 in Appendix A).

We conclude this section by discussing coordination rates. For this purpose, we compared the (expected) coordination rates that would be obtained if participants in different subsamples were paired with each other, using Monte Carlo methods (see footnote 17). In short, whereas picking at random implies a coordination rate of 0.33, our

random only if they did not have a compelling reason to rely on the labels' word frequency in their own vocabularies.

subjects' choice behavior implies an expected coordination rate of 0.55 and 0.43, respectively, for US and UK participants *knowingly* paired with their compatriots. Notably, these rates drop to 0.47 (0.39) in the case of US (UK) participants *unknowingly* paired with their compatriots. We finally considered the case in which US and UK participants were knowingly and unknowingly paired with each other, and found that the expected coordination rates were respectively 0.37 and 0.38. In summary, our results show that successful coordination is more likely when subjects are knowingly paired with partners from their own country, as opposed to when they are knowingly or unknowingly paired with partners from a different country.

6. Conclusion

We have presented a set of studies that test whether the frequency with which labels are mentioned in everyday language may affect game play. In the first study, we found that the labels' frequency of occurrence in the vocabulary of the subject (quantified by the Google Books n-gram frequency) is a good predictor of choice behavior in coordination games. Our second study verifies if subjects utilize word frequency strategically rather than naïvely. To do so, we contrasted participants in coordination games with participants in three alternative roles, namely, "Pickers," "Hiders," and "Seekers." The data reveal that Pickers are as likely as Coordinators to select the most frequently-mentioned label; instead, Hiders are less likely than Seekers and, in turn, Seekers are less likely than Coordinators. This pattern suggests a (boundedly) rational use of frequently-mentioned labels. Our third study delves into the strategic use of word frequency in coordination games, by contrasting culturally diverse participants. To that end, we recruited samples of US and UK participants, and then varied their knowledge of the counterpart's country of residence. Consistent with our predictions, we found that behavior differs across US and UK residents, and it is positively related to the labels' n-gram frequency in the subject's own vocabulary. Further, a subject is less likely to rely on word frequency as a means to guiding her behavior *if* she knows that the counterpart resides in a different country.

Our approach was inspired by previous evidence on the use of word frequency as a cue in nonstrategic tasks (e.g., Goldstein and Gigerenzer (2002); Dougherty, Franco-Watkins, and Thomas (2008)). Here, we have shown that word frequency has a role in strategic reasoning, too. Remarkably, our results imply that subjects are consciously aware as to how labels might be perceived by culturally alike counterparts, and accordingly adjust their strategies. In this regard, our results are related to a stream of research suggesting that individuals are aware of the communication difficulties that arise when groups with different "conversational codes" merge with each other (Weber and Camerer (2003); Feiler and Camerer (2010)). It is also worth noting a connection with *rational speech act theory*, which formalizes how participants in conversational interactions make inferences about the meaning of utterances so as to achieve "coordination of meaning," based on their knowledge of the counterpart and context (Goodman and Frank (2016)).

Before concluding, we note that the past decade has seen the growth of large-scale online datasets and, with it, unique opportunities to investigate human behavior and its cultural correlates. In particular, the Google Books corpus has been used to analyze trends in stereotypes and wellbeing across cultures (e.g., Garg et al. (2018); Hills et al. (2019)). More generally, internet data on human activity have been used in fields as diverse as public health (Hawn (2009)), cognitive science (Griffiths (2015)), and management (George, Osinga, Lavie, and Scott (2016)). Our paper shows that such data may inform theories of strategic reasoning as well. Relatedly, we note that although this paper has focused on strategic problems with incentives *to* and *not to* coordinate, our characterization of prominence may have wider application; in fact, subjects' exposure to alternative labels—ceteris paribus—can inform behavioral predictions in any class of games with strategic uncertainty.

To conclude, this paper has proposed and tested an a-priori measurable proxy for prominence that explicitly rests on players' culture. The results provide very robust evidence in support of our characterization of prominence. In doing so, the results contribute to shedding light on the relationship between culture, (bounded) rationality and coordination, which plays an important role in several interactions among economic agents.

References

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003), "Fractionalization." *Journal of Economic Growth*, 8 (2), 155–194. [284]

Alesina, A. and P. Giuliano (2015), "Culture and institutions." *Journal of Economic Literature*, 53 (4), 898–944. [283]

Alós-Ferrer, C. and C. Kuzmics (2013), "Hidden symmetries and focal points." *Journal of Economic Theory*, 148 (1), 226–258. [285]

Anderson, J. R. and L. J. Schooler (1991), "Reflections of the environment in memory." *Psychological Science*, 2 (6), 396–408. [285]

Arad, A. and A. Rubinstein (2012), "The 11-20 money request game: A level-k reasoning study." *American Economic Review*, 102 (7), 3561–3573. [295]

Bacharach, M. (1993), "Variable universe games." In *Frontiers of Game Theory* (K. Binmore et al., eds.). MIT Press, Cambridge, MA. [285]

Bacharach, M. (2006), *Beyond Individual Choice: Teams and Frames in Game Theory* (N. Gold and R. Sugden, eds.). Princeton University Press. [284]

Bacharach, M. and M. Bernasconi (1997), "The variable frame theory of focal points: An experimental study." *Games and Economic Behavior*, 19 (1), 1–45. [284, 285]

Bacharach, M. and O. Dale Stahl (2000), "Variable-frame level-n theory." *Games and Economic Behavior*, 32 (2), 220–246. [285]

Balota, D. A. and D. H. Spieler (1999), "Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency." *Journal of Experimental Psychology: General*, 128 (1), 32–55. [285]

Bardsley, N., J. Mehta, C. Starmer, and R. Sugden (2010), "Explaining focal points: Cognitive hierarchy theory versus team reasoning." *Economic Journal*, 120 (543), 40–79. [287]

Bardsley, N. and A. Ule (2017), "Focal points revisited: Team reasoning, the principle of insufficient reason and cognitive hierarchy theory." *Journal of Economic Behavior & Organization*, 133, 74–86. [286]

Blume, A. (2000), "Coordination and learning with a partial language." *Journal of Economic Theory*, 95 (1), 1–36. [285]

Blume, A. and U. Gneezy (2010), "Cognitive forward induction and coordination without common knowledge: An experimental study." *Games and Economic Behavior*, 68 (2), 488–511. [287]

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004), "A cognitive hierarchy model of games." *Quarterly Journal of Economics*, 119 (3), 861–898. [294]

Casajus, A. (2000), "Focal points in framed strategic forms." *Games and Economic Behavior*, 32 (2), 263–291. [285]

Charness, G., U. Gneezy, and M. A. Kuhn (2013), "Experimental methods: Extra-laboratory experiments–extending the reach of experimental economics." *Journal of Economic Behavior & Organization*, 91, 93–100. [288]

Charness, G. and A. Sontuoso (2019), "The doors of perception." ESI Working Paper 19-32. [285]

Cooper, R. W., D. V. DeJong, R. Forsythe, and T. W. Ross (1990), "Selection criteria in coordination games: Some experimental results." *American Economic Review*, 80 (1), 218–233. [286]

Costa-Gomes, M. A., V. P. Crawford, and B. Broseta (2001), "Cognition and behavior in normal-form games: An experimental study." *Econometrica*, 69 (5), 1193–1235. [287, 294]

Costa-Gomes, M. A. and V. P. Crawford (2006), "Cognition and behavior in two-person guessing games: An experimental study." *American Economic Review*, 96 (5), 1737–1768. [294]

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013), "Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications." *Journal of Economic Literature*, 51 (1), 5–62. [294]

Crawford, V. P., U. Gneezy, and Y. Rottenstreich (2008), "The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures." *American Economic Review*, 98 (4), 1443–1458. [285, 292]

Crawford, V. P. and H. Haller (1990), "Learning how to cooperate: Optimal play in repeated coordination games." *Econometrica*, 58 (3), 571–595. [285]

Crawford, V. P. and N. Iriberri (2007), "Fatal attraction: Salience, naivete, and sophistication in experimental hide-and-seek games." *American Economic Review*, 97 (5), 1731–1750. [295]

Dougherty, M. R., A. M. Franco-Watkins, and R. Thomas (2008), "Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics." *Psychological Review*, 115 (1), 199–211. [286, 306]

Easterly, W. and R. Levine (1997), "Africa's growth tragedy: Policies and ethnic divisions." *Quarterly Journal of Economics*, 112 (4), 1203–1250. [284]

Einav, L. and J. Levin (2014), "Economics in the Age of Big Data." *Science*, 346 (6210), 1243089. [284]

Faillo, M., A. Smerilli, and R. Sugden (2017), "Bounded best-response and collective-optimality reasoning in coordination games." *Journal of Economic Behavior & Organization*, 140, 317–335. [286]

Feiler, L. and C. F. Camerer (2010), "Code creation in endogenous merger experiments." *Economic Inquiry*, 48 (2), 337–352. [306]

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou (2018), "Word embeddings quantify 100 years of gender and ethnic stereotypes." *Proceedings of the National Academy of Sciences*, 115 (16), E3635–E3644. [285, 288, 307]

George, G., E. C. Osinga, D. Lavie, and B. A. Scott (2016), "Big data and data science methods for management research." *Academy of Management Journal*, 59 (5), 1493–1507. [307]

Gigerenzer, G. and D. G. Goldstein (1996), "Reasoning the fast and frugal way: Models of bounded rationality." *Psychological Review*, 103 (4), 650–669. [286]

Ginsburgh, V. and S. Weber (2020), "The economics of language." *Journal of Economic Literature*, 58 (2), 348–404. [284]

Goldstein, D. G. and G. Gigerenzer (2002), "Models of ecological rationality: The recognition heuristic." *Psychological Review*, 109 (1), 75–90. [306]

Goodman, N. D. and M. C. Frank (2016), "Pragmatic language interpretation as probabilistic inference." *Trends in Cognitive Sciences*, 20 (11), 818–829. [306]

Griffiths, T. L. (2015), "Manifesto for a new (computational) cognitive revolution." *Cognition*, 135, 21–23. [307]

Hargreaves Heap, S., D. Rojo Arjona, and R. Sugden (2014), "How portable is level-0 behavior? A test of level-k theory in games with non-neutral frames." *Econometrica*, 82 (3), 1133–1151. [287]

Hargreaves Heap, S., D. Rojo Arjona, and R. Sugden (2017), "Coordination when there are restricted and unrestricted options." *Theory and Decision*, 83 (1), 107–129. [287]

Harsanyi, J. C. and R. Selten (1988), *A General Theory of Equilibrium Selection in Games*. MIT Press, Cambridge, MA. [285, 292]

Hawn, C. (2009), "Take two aspirin and tweet me in the morning: How Twitter, Facebook, and other social media are reshaping health care." *Health Affairs*, 28 (2), 361–368. [307]

Hedden, T. and J. Zhang (2002), "What do you think I think you think?: Strategic reasoning in matrix games." *Cognition*, 85 (1), 1–36. [286]

Hertwig, R., S. M. Herzog, L. J. Schooler, and T. Reimer (2008), "Fluency heuristic: A model of how the mind exploits a by-product of information retrieval." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34 (5), 1191–1206. [286]

Hertwig, R., T. Pachur, and S. Kurzenhäuser (2005), "Judgments of risk frequencies: Tests of possible cognitive mechanisms." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31 (4), 621–642. [286]

Hills, T. T., E. Proto, D. Sgroi, and C. I. Seresinhe (2019), "Historical analysis of national subjective wellbeing using millions of digitized books." *Nature Human Behaviour*, 3, 1–5. [285, 288, 307]

Janssen, M. C. W. (2001), "Rationalizing focal points." *Theory and Decision*, 50 (2), 119–148. [285]

Jescheniak, J. D. and W. J. M. Levelt (1994), "Word frequency effects in speech production: Retrieval of syntactic information and of phonological form." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20 (4), 824–843. [285]

Kapferer, J.-N. (2012), *The New Strategic Brand Management: Advanced Insights and Strategic Thinking*. Kogan Page, London. [284]

Kossoudji, S. A. (1988), "English language ability and the labor market opportunities of Hispanic and East Asian immigrant men." *Journal of Labor Economics*, 6 (2), 205–228. [284]

Lang, K. (1986), "A language theory of discrimination." *Quarterly Journal of Economics*, 101 (2), 363–382. [284]

Lazear, E. P. (1999), "Culture and language." *Journal of Political Economy*, 107 (6), S95–S126. [284]

Lewis, D. K. (1969), *Convention: A Philosophical Study*. Harvard U. Press, Cambridge, MA. [284]

Lichtenstein, S., P. Slovic, B. Fischhoff, M. Layman, and B. Combs (1978), "Judged frequency of lethal events." *Journal of Experimental Psychology: Human Learning and Memory*, 4 (6), 551–578. [286]

Luce, S. D. and H. Raiffa (1957), *Games and Decisions: Introduction and Critical Survey*. John Wiley & Sons. [284]

Matz, S. C. and O. Netzer (2017), "Using big data as a window into consumers' psychology." *Current Opinion in Behavioral Sciences*, 18, 7–12. [284]

Mauersberger, F. and R. Nagel (2018), "Levels of reasoning in Keynesian beauty contests: A generative framework." In *Handbook of Computational Economics*, Vol. 4 (C. Hommes and B. LeBaron, eds.). Elsevier, Amsterdam. [294]

McFadden, D. (1973), "Conditional logit analysis of qualitative choice behavior." In *Frontiers in Econometrics* (P. Zarembka, ed.). Academic Press, New York, NY. [298, 305]

Mehta, J., C. Starmer, and R. Sugden (1994), "The nature of salience: An experimental investigation of pure coordination games." *American Economic Review*, 84 (3), 658–673. [291, 292]

Melitz, J. (2008), "Language and foreign trade." *European Economic Review*, 52 (4), 667–699. [284]

Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. Lieberman Aiden (2011), "Quantitative analysis of culture using millions of digitized books." *Science*, 331 (6014), 176–182. [284, 288]

Morrison, C. M. and A. W. Ellis (1995), "Roles of word frequency and age of acquisition in word naming and lexical decision." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (1), 116–133. [285, 286]

Nagel, R. (1995), "Unraveling in guessing games: An experimental study." *American Economic Review*, 85 (5), 1313–1326. [287, 294]

Penczynski, S. P. (2016), "Strategic thinking: The influence of the game." *Journal of Economic Behavior & Organization*, 128, 72–84. [295]

Rubinstein, A. and Y. Salant (2016), "'Isn't everyone like me?': On the presence of self-similarity in strategic interactions." *Judgment and Decision Making*, 11 (2), 168–173. [286]

Rubinstein, A., A. Tversky, and D. Heller (1997), "Naive strategies in competitive games." In *Understanding Strategic Interaction* (W. Albers et al., eds.). Springer, Berlin. [285]

Schelling, T. C. (1960), *The Strategy of Conflict*. Harvard U. Press, Cambridge, MA. [284]

Schwarz, N., H. Bless, F. Strack, G. Klumpp, H. Rittenauer-Schatka, and A. Simons (1991), "Ease of retrieval as information: Another look at the availability heuristic." *Journal of Personality and Social Psychology*, 61 (2), 195–202. [286]

Sebanz, N., G. Knoblich, and W. Prinz (2003), "Representing others' actions: Just like one's own?" *Cognition*, 88 (3), B11–B21. [286]

Seidenberg, M. S. and J. L. McClelland (1989), "A distributed, developmental model of word recognition and naming." *Psychological Review*, 96 (4), 523–568. [286]

Sontuoso, A., and S. Bhatia (2021), "Supplement to 'A notion of prominence for games with natural-language labels'." *Quantitative Economics Supplemental Material*, 12, https://doi.org/10.3982/QE1212. [291]

Stahl, D. O. and P. W. Wilson (1995), "On players' models of other players: Theory and experimental evidence." *Games and Economic Behavior*, 10 (1), 218–254. [287, 294]

Sugden, R. (1995), "A theory of focal points." *Economic Journal*, 105 (430), 533–550. [285]

Tversky, A. and D. Kahneman (1973), "Availability: A heuristic for judging frequency and probability." *Cognitive Psychology*, 5 (2), 207–232. [286]

Tversky, A. and D. Kahneman (1974), "Judgment under uncertainty: Heuristics and biases." *Science*, 185 (4157), 1124–1131. [286]

Van Huyck, B. J., R. C. Battalio, and R. O. Beil (1990), "Tacit coordination games, strategic uncertainty, and coordination failure." *American Economic Review*, 80 (1), 234–248. [286]

Varian, H. R. (2014), "Big data: New tricks for econometrics." *Journal of Economic Perspectives*, 28 (2), 3–28. [284]

Weber, R. A. and C. F. Camerer (2003), "Cultural conflict and merger failure: An experimental approach." *Management Science*, 49 (4), 400–415. [306]

Yin, Z., L. M. Sulieman, and B. A. Malin (2019), "A systematic literature review of machine learning in online personal health data." *Journal of the American Medical Informatics Association*, 26 (6), 561–576. [284]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 24 September, 2018; final version accepted 28 September, 2020; available online 7 October, 2020.