Unacquainted Callers Can Predict Which Citizens Will Vote

Over and Above Citizens' Stated Self-Predictions

Todd Rogers

Harvard Kennedy School

Harvard University

Leanne ten Brinke

Haas School of Business & Department of Psychology

University of California, Berkeley

Dana R. Carney

Haas School of Business

University of California, Berkeley

To cite:
Rogers, T., ten Brinke, L., & Carney, D.R. (2016). Unacquainted callers can predict which citizens will vote over and above citizens' stated self-predictions. *Proceedings of the National Academy of Sciences*, *113*(23), 6449-6453.

Abstract

People are regularly asked to report on their likelihoods of carrying out consequential future behaviors, including complying with medical advice, completing their educations, and voting in upcoming elections. Yet, despite these stated self-predictions being notoriously unreliable, they are used to inform many strategic decisions. We report two studies examining stated self-prediction about whether citizens will vote. We find that most self-predicted voters do not actually vote despite saying they will, and that campaign callers can discern which self-predicted voters will not actually vote. In Study 1 (*N*=4,463), self-predicted voters rated by callers as "100% likely to vote" were two times more likely to *actually* vote than those rated unlikely to vote. Study 2 (*N*=3,064) replicated this finding and further demonstrated that callers' prediction accuracy was mediated by citizens' nonverbal signals of uncertainty and deception. Strangers can use nonverbal signals to improve predictions of follow through on self-reported intentions – an insight of potential value for politics, medicine, and education.

Keywords: self-prediction; human judgment; deception; nonverbal behavior; voting

Significance Statement

People are regularly asked to report on their likelihoods of carrying out consequential future behaviors, including complying with medical advice, completing their educations, and voting in elections; responses, however, are notoriously unreliable. For example, more than half of people who state a self-prediction that they will vote in an upcoming election actually do not. We find that untrained survey callers' predictions of who will vote meaningfully increases the statistical prediction of which respondents will follow-through on their stated self-predictions—over and above respondents' self-predictions. Callers accomplish this by attending to signals of uncertainty and deception conveyed in respondents' voices. These findings could improve political campaign resource allocation, and could improve the targeting of interventions in domains including health and education.

People are regularly asked to predict whether they will follow-through on important commitments. Responses are often interpreted as if they are accurate and interventions are directed toward those who self-predict being unlikely to follow through. This scenario can play out in medical treatment adherence, persistence in higher education, and showing up at the polls to vote. However, many individuals who state that they will follow-through on consequential behaviors actually do not—rendering people's stated self-predictions poor forecasts of actual future behaviors (*1*). For example, in some US elections the majority of respondents who self-predict that they will vote do not actually vote (i.e., they "flake out"; *2*).

"Flaking-out" on one's stated self-prediction may be attributable to a range of factors including uncertainty about the future, poor planning for carrying out the intended behavior (*3*), and providing false, but socially desirable, responses (*4*). In a context with strong situational demand, such as calls in which respondents are asked to predict whether they will vote in an upcoming election, the desire to appear like an active and responsible citizen is likely to outweigh variation in responses due to individual differences. In other words, there is strong situational pressure for respondents to engage in deception and say that they *will* vote when, in fact, they *will not* (*5*).

Despite the error-prone nature of voters' stated self-predictions, these data inform which campaign advertisements are developed, where and when these messages are aired, and which voters are targeted in get-out-the-vote (GOTV) campaigns (*6*). Though political candidates, analysts, and pundits make due with these inaccurate self-reports, social psychological research suggests that incorporating the human judgment of people other than the respondents might enhance the statistical accuracy of people's stated self-predictions. Humans are extremely adept at reading others' psychological states and traits (*7*). For example, based on a very brief interaction or “thin slice” of social information, people can accurately detect if someone is feeling happy, sad, angry or fearful (e.g. *8*), how agreeable, conscientious, and extraverted they are (*9*), how racially prejudiced they are (*10*), assess their sexual orientation (*11*), judge their interest in having children (*12*), and their past violent behavior (*13*). This impressive ability to evaluate others so quickly and accurately is presumed to foster efficient prediction of others’ behavior, and our social interactions with them (*14*). Although efforts by the American National Election Study have, for example, for a brief period attempted to substitute objective measures of citizens’ political knowledge with callers’ subjective impressions of the same (*15, 16*), research has not examined whether “thin slice” judgments of stated self-predictions can discriminate who will “flake out” on that action, and who will follow-though.

Here, we examine whether brief interactions allow observers to predict people’s future behavior in an important context with societal and financial implications. We ask whether callers can achieve enhanced predictive accuracy in discriminating which self-predicted voters will actually cast a vote from those who will not (Study 1). In “thin-slices” research, perceivers are shown or hypothesized to rely on nonverbal cues in addition to verbal statements to make accurate judgments about others’ mental states, performance, and future behavior (*7*). In Study 2, we examine which specific nonverbal behaviors callers rely upon to achieve enhanced predictive accuracy—exploring how nonverbal cues associated with uncertainty and deception predict “flaking out”.

Study 1

Study 1 tested the hypothesis that human judgment could improve upon the accuracy of people’s stated self-predictions about their future behavior. It examined whether callers could identify which citizens who stated that they would vote actually voted versus “flaked out.”

Method

During the week prior to the 2009 New Jersey General Election, a nonprofit organization ran a GOTV paid phone program targeting 2008 voters who had not voted in 2006, and were identified on the public voter file as either African American or Hispanic (see demographic information in Table S1). Callers first confirmed reaching pre-specified individual citizens, then asked, “Do you plan to vote on Tuesday?” If citizens answered yes, callers read a fifteen-second script encouraging participation in the election (see SOM). No other questions were asked of citizens, leaving only three brief opportunities for communication between callers and citizens: initial confirmation that the pre-specified citizens had been reached, the citizens’ responses to the callers’ self-prediction question, and whatever exchange occurred as the calls terminated. After each brief call with respondents who stated that they would vote (i.e., self-reported voters), callers were asked to estimate “How likely is this person to *actually cast a vote* in the 11/09 election?” Callers responded on a five-point scale ranging from a 0% to 100% (i.e., likelihood of follow-through, increasing in 25% increments). Callers were not asked to make this estimate for self-predicted nonvoters or unsure voters. A total of 4,487 citizens stated that they would vote and callers predicted turnout for 4,463 of these citizens (99% of sample). An additional 1,696 citizens self-predicted that they would not vote, and a further 644 were unsure.

To determine the accuracy of caller predictions, two datasets were merged post-election: that containing the responses of all self-predicted voters, and the publicly available voter file. This voter file reflected which citizens actually voted in the 2009 election, the demographics of the citizens, and whether citizens had voted in the 2007 general election. Because callers predicted turnout for only citizens who self-predicted that they would vote, the following analyses reflect only those citizens. UC Berkeley IRB determined that analysis of de-identified Study 1 voter data was not human subject research.

Results and Discussion

While only 47% of self-predicted voters actually cast a vote, callers' predictions of who would actually vote significantly predicted which self-predicted voters followed through versus flaked out: log-odds ratio = 2.75, $p < .001$, controlling for caller fixed-effects (excluding fixed effects does not affect results; Figure 1 & Table S2). Those citizens whom callers predicted were "100% likely to vote" were *nearly two times* as likely to actually vote as citizens whom callers predicted were "0% likely to vote." Citizens who self-predicted that they would vote, but were rated by callers as "0% likely to vote", flaked-out in 74% of cases—indicating high caller accuracy when determining that a citizen had falsely stated that they would vote (see Table S3). Though callers predicted that most self-predicted voters would follow through and cast a vote, when callers doubted citizens' stated self-predictions, callers were highly accurate. Overall, the logistic regression model presented in Table S2 (Model 1) was successful in predicting the actual voting behavior of 58.5% of self-predicted voters.

Although callers' ratings were affected by respondent demographics (Table S4), adding citizens' race, age, and gender to the models predicting voter turnout had no substantive effect on the predictive accuracy of callers' ratings (Table S2). These findings suggest that callers' predictive accuracy did not result from callers inferring citizens' memberships in groups with stereotypically relatively higher turnout (e.g., older, Hispanic) or lower turnout (e.g., younger, African American). Further, adding previous voting behavior to the model increased r^2, but did not decrease the predictive power of caller predictions, suggesting that callers were not simply picking up on citizens' trait-like qualities of being an engaged citizen (*17*), or of a particular demographic category when making their judgments. As such, it appears that callers attended to some nonverbal cues conveyed in the manner in which citizens said "yes" to achieve an increase in accuracy over and above self-stated predictions.

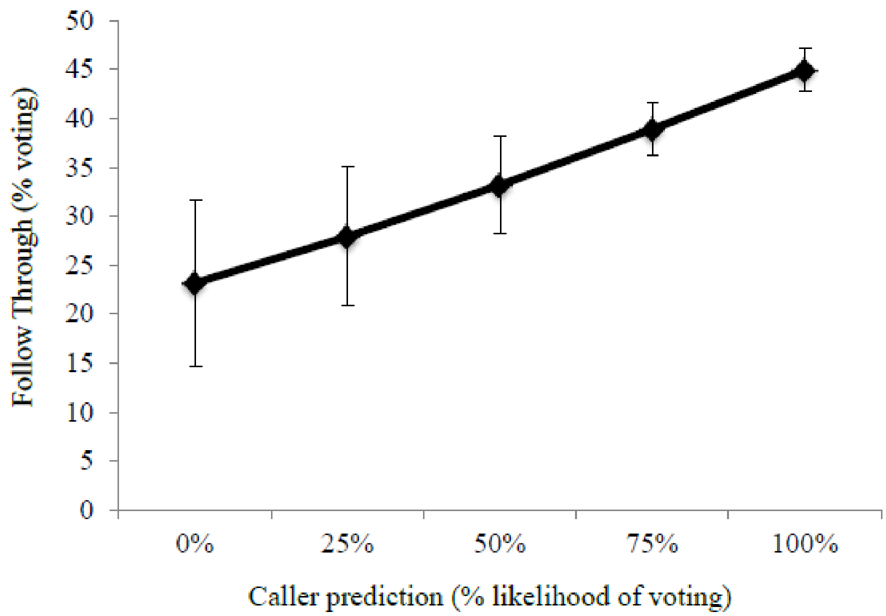

Figure 1. Caller predictions of whether self-predicted voters will vote, by actual turnout. Error bars represent 95% confidence intervals.

Study 2

Study 2 replicated Study 1's finding that untrained callers can discern which self-predicted voters will not actually vote. Extending Study 1, Study 2 also examined whether callers have the inverse predictive ability: can callers discern which self-predicted non-voters actually will vote (i.e., "flake-in")? Additionally, Study 2 tested which voice-related nonverbal cues callers relied upon to differentiate those who followed through from those who flaked-out. While self-predicted voters may flake-out for a number of reasons, two likely explanations are that when they self-predicted that they would vote they were: a) uncertain of their future behavior but offered a sincere best guess; and/or b) explicitly deceiving the callers – offering callers what citizens may have perceived to be socially desirable responses despite not really expecting to vote.

Social psychological research suggests that uncertainty and deception are likely to cause detectable nonverbal paralinguistic cues in citizens' voices. Humans can detect uncertainty in others' voices within milliseconds: it is generally associated with a quiet voice and a rise in pitch at the end of a statement (*18-20*). Deception is commonly associated with nonverbal signals of arousal and cognitive load, a term describing a "spent" mind, lacking in available cognitive resources (*21*). Arousal may increase due to feelings of guilt, fear of discovery, or duping delight (i.e., "glee" associated with deception), and may manifest in perceptible tension, nervousness, and increases in vocal pitch. Cognitive load occurs during acts of deception because creating a falsified event or intention is more difficult than recalling true details from memory. Cognitive load may manifest in slowed speech rate, speech fillers, and long pauses before responding (*22*). In addition to testing which nonverbal cues callers naturally use to predict whether self-predicted voters will follow through or "flake-out," Study 2 examined whether other nonverbal cues were left unused by callers that could have been leveraged to make even better predictions. Further, nonverbal cues that require human judgment were examined alongside those that can be assessed automatically by computer software, to examine whether an improved prediction model would require human observers (e.g. *23*).

Method

Study 2 involved a GOTV paid phone program conducted during the weekend prior to the 2010 Texas Gubernatorial Election (*N*=3,064). Unlike Study 1, individuals were not targeted on the basis of race; a diverse cross-section of the population was included (see Table S7 for demographic information). However, targets were selected because they were predicted to have a moderate likelihood of voting in the 2010 Texas Gubernatorial Election (e.g., they voted in the high-salience 2008 Presidential General Election, but had not voted in the much lower salience 2007 statewide Texas Constitutional Amendment election). After callers confirmed reaching the targeted citizens, callers asked “Do you plan on voting in this Tuesday’s election for Governor?” Calls proceeded in the same manner as in Study 1; however, in this study, callers predicted whether citizens would vote regardless of the citizens’ stated self-predictions (whereas in Study 1 callers only made predictions about self-predicted voters). After calls terminated, callers were asked “How likely is this person to actually cast a vote in this Tuesday’s election?” As in Study 1, callers responded on a five-point scale ranging from a 0% to 100% likelihood of follow-through. 2,049 citizens stated that they would vote, 695 stated that they would not vote, and 320 were unsure. Callers made follow-through predictions for all 3,064 of these individuals. To determine the accuracy of caller predictions, the dataset containing the responses of all citizens who offered a self-prediction was merged with the public voter file. The voter file reflected which citizens actually voted in the 2010 election, the demographics of the citizens, and whether citizens had voted in the 2006 general election.

Texas law permitted these calls to be recorded as long as: i) callers were aware they were being recorded; ii) calls were placed from within Texas; and iii) callers telephoned citizens within Texas. Following commonly used methods in social and personality psychology (*24*), the recordings were later coded by three research assistants who were blind to voting behavior and hypotheses. Coders were trained to listen to each audio clip and code the presence/absence and qualitative aspects of nonverbal behaviors related to uncertainty, cognitive load, and arousal. Pairs of coders were randomly assigned to code each behavior. After practicing the to-be-coded behavior on a separate set of audio stimuli, both coders coded a randomly determined subset of 10% of the audio clips. After reliability was established, one coder proceeded to code all stimuli on that behavior. Behaviors were coded one at a time over 3 months. Table S6 lists all behaviors coded, definitions, associated references to relevant research, coding scales, approach, and inter-rater reliability. New England IRB reviewed and approved the collection of audio recordings used in Study 2 on behalf of the partner organization that conducted the study, the Analyst Institute. UC Berkeley IRB determined that the coding of the de-identified audio calls and analysis of abstracted data was exempt.

Results and Discussion

Replicating Study 1, only 38% of self-predicted voters actually voted. Also consistent with Study 1 and our hypothesis, callers' predictions of citizen turnout was a significant predictor of which self-predicted voters would follow through versus flake-out: log-odds ratio = 3.09, $p < .001$, controlling for caller fixed-effect (excluding fixed effects does not affect results; Figure 2 & Table S7). Self-predicted voters whom callers predicted were "100% likely to vote" were, again, *two times as likely* to actually vote as self-predicted voters whom callers predicted were unlikely to vote. Overall, this logistic regression model (Table S7; Model 4) accurately predicted the actual voting behavior of 64.2% of self-predicted voters. Again, caller predictions were somewhat related to respondents' demographics (Table S8), but as in Study 1 adding respondents' age, race, gender, and previous voting behavior to the models predicting voter turnout had no substantive effect on the predictive accuracy of callers' ratings (Table S7). Results suggest that callers' were not simply relying on racial stereotypes, nonverbal indices of past voting behavior, or trait-like qualities to be accurate.

Unlike in Study 1, callers made ratings of the likelihood of voting for all respondents, regardless of whether they said yes, no, or were unsure whether they would vote in the upcoming election. As such, we can conduct a logistic regression model predicting actual voting behavior for all respondents, including stated self-prediction (yes, no, maybe) and caller ratings of voting likelihood as predictors (and controlling for caller fixed-effects). As expected, caller ratings explained unique variance in actual voting behavior (log-odds ratio = 2.12, $p < .001$), over and above respondents' stated self-predictions (Table S7; Model 7). In other words, any respondent, regardless of his or her stated self-prediction, is twice as likely to show up to the polls if they are rated as "100% likely to vote", versus those rated "0% likely to vote". However, when we restrict our sample to only those respondents who self-predicted no intention to vote ($N = 695$), or only those that who stated they were unsure ($N = 320$), callers' predictions were unrelated to actual voting behavior, log-odds ratio = .536, $p = .388$, and log-odds ratio = 2.62, $p = .225$, respectively, excluding caller fixed-effects to maximize sample size (Figure 2 and Table S9). In short, caller ratings improve upon respondents' stated self-predictions in predicting actual voting behavior. In particular, callers provide particular insight into which respondents will "flake-out" on their stated self-predictions to vote.

To determine how callers formed accurate predictions of which self-predicted voters would flake-out versus follow-through, a Brunswikian lens model was fit to the data. Figure 3 reveals valid nonverbal behaviors (i.e., correctly used cues that led to accuracy), invalid behaviors (i.e., cues used that did not lead to accuracy), and missed opportunities (i.e., valid cues that were not leveraged; *25*). Sounding uncertain, sounding insecure, and having longer latencies prior to responding to the self-prediction question were valid behaviors, meaning callers utilized these behaviors to make accurate judgments. In other words, these behaviors were correlated with both callers' *predictions* and citizens' *actual* voting behavior. Callers also interpreted sounding tense, and sounding nervous as signals that self-predicted voters would not vote, but these nonverbal behaviors were unrelated to actual voting behavior. Additionally, the more speech fillers self-predicted voters used the less likely they were to vote, though callers failed to use this diagnostic cue when predicting who would vote. Speech rate and mean and maximum vocal pitch were unrelated to both caller predictions and actual voting behavior.

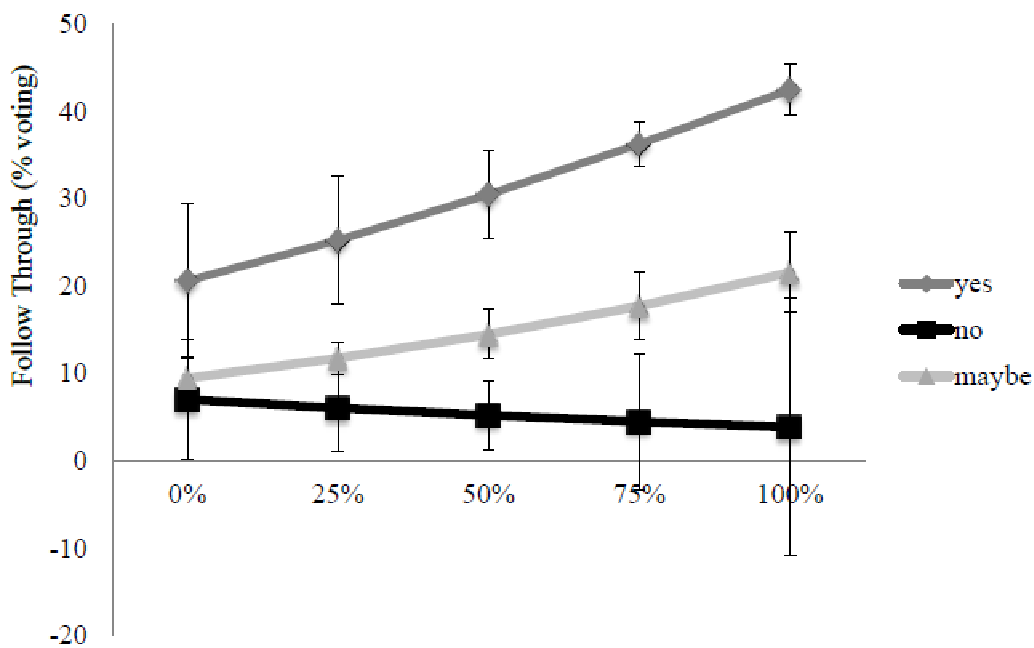

Figure 2. Caller predictions of whether self-predicted voters, self-predicted non-voters, and those reporting being unsure whether they would vote will vote, by actual turnout. Error bars represent 95% confidence intervals.

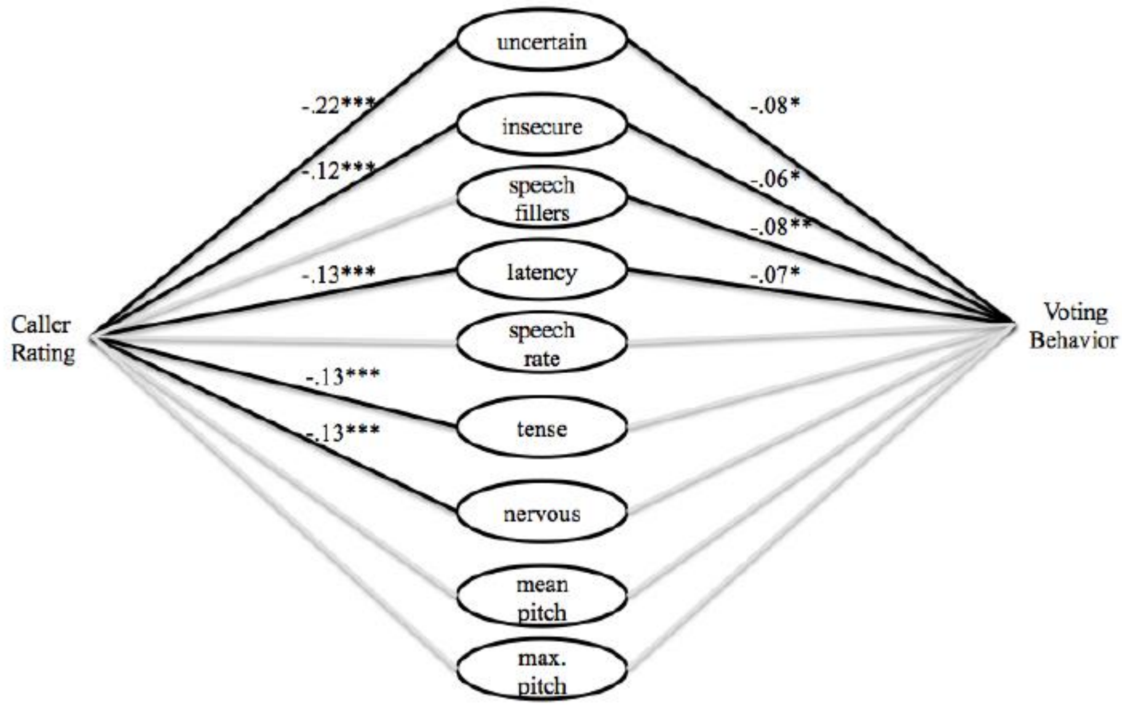

Figure 3. Correlations between nonverbal cues and respondents' voting behavior (on the right) and caller ratings of voting likelihood (on the left). Black lines indicate significant relationships (**p*<.05; ***p*<.01; ****p*<.001); grey lines indicate non-significant relationships, *p*s>.05. Callers correctly incorporate in their vote predictions respondents sounding more uncertain, more insecure, and having longer speech latency/voice onset (these attributes were related to actual respondent voting). Callers incorrectly incorporate in their vote predictions respondents sounding more tense and more nervous (these attributes were unrelated to actual respondent voting). Callers correctly did not incorporate in their vote predictions respondents' mean vocal pitch, maximum vocal pitch, and higher speech rate (these attributes were unrelated to actual respondents voting). Callers missed the opportunity to incorporate in their vote predictions respondents' use of speech fillers (this attribute was related to actual respondent voting).

General Discussion

Two large-scale studies (total N = 7,527 registered US voters) demonstrated that callers can discriminate which citizens will follow-through versus flake-out on their stated self-predictions to vote in upcoming elections. Practically, the findings suggest that human social judgments may be used to complement citizens' stated self-predictions about critical future behaviors in elections, as well as domains such as medicine and education. For example, this research directly suggests that human social judgment can complement other data sources (including stated self-prediction) to better target campaign messaging and GOTV efforts (*26*). Theoretically, this research suggests that accurate thin-slice predictions extend beyond psychological states and traits, to the prediction of follow-through on stated self-predictions of future behaviors. Study 2 indicates that callers relied upon signals of uncertainty and cognitive load to discern which self-predicted voters would actually vote. Specifically, coders' ratings of uncertainty, insecurity, and measurements of response latency were used by callers to correctly identify which citizens would "flake out." Speech fillers also predicted flaking out. While objective measures of response latency and speech fillers may be automated to improve follow through predictions, subjective measures of uncertainty and insecurity require human judgment (*23*, *27*).

The nonverbal cues that validly predicted which self-predicted voters will actually vote provide support for two potential explanations for why stated self-predictions may not accurately reflect future behavior. Nonverbal uncertainty may reveal that in the moments when individuals state their self-predictions, some are unsure about their desire or ability to complete the behaviors. This explanation is consistent with psychological research suggesting that helping self-predicted voters make a concrete plan can improve voter turnout (*3*). Further, that nonverbal cues of cognitive load—some of the most reliable indicators of deception—are valid indicators of follow through suggests that some individuals may lie when self-predicting their future behaviors. Callers correctly leveraged multiple signals of uncertainty to detect flake-outs. At the same time, though, callers relied upon a mixture of valid and invalid cues to deception, and missed the opportunity to use speech fillers to detect deceptive self-predictions. Future research should test whether explicit training about which cues are valid and which are invalid could improve caller accuracy. Additionally, future research should explore when stated self-predictions will and will not accurately predict behavior, as well as the prevalence and consequence of potential biases based of respondent demographics on caller predictions. These findings add to an emerging literature documenting circumstances in which humans can detect deception in real-time interactions at greater than chance levels (e.g. *28*, *29*).

In conclusion, ordinary, untrained human judges can significantly improve predictions of who will follow-through versus flake-out on important commitments. This knowledge could increase the efficiency of the allocation of campaign resources and is likely to be valuable in other domains as well. For example, it could be used to improve the targeting of costly interventions that increase patient compliance in medical care – a context in which billions of dollars are wasted due to patients' lack of follow-through (*30*) – and to better identify the students most at-risk of failing to follow-through on their college study and persistence plans. In short, the findings speak to a broad challenge in social life and suggest a simple input that leverages human social judgment to increase the accuracy of intervention targeting.

References

1. Ajzen I (1991) Theories of cognitive self-regulation: the theory of planned behavior. *Organ. Behav. Hum. Decis. Process*. 50: 179–211.

2. Rogers T, Aida M (2014) Vote self-prediction hardly predicts who will vote, and is (misleadingly) unbiased. *Am. Polit. Res*. 42: 503–528.

3. Nickerson D W, Rogers T (2010) Do you have a voting plan?: implementation intentions, voter turnout, and organic plan making. *Psychol. Sci*. 21: 194–199.

4. Tourangeau R, Rips L J, Rasinski K (2000) *The Psychology of Survey Response* (Cambridge University Press).

5. Cooper W H, Withey M J (2009) The strong situation hypothesis. *Pers. Soc. Psychol. Rev*. 13: 62–72.

6. Green D P, Gerber A S (2008) *Get Out the Vote: How to Increase Voter Turnout* (Brookings Institution Press).

7. Ambady N, Rosenthal R (1992) Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol. Bull*. 111: 256–274.

8. Ekman P (1993) Facial expression and emotion. *Am. Psychol*. 48: 384–392.

9. Carney D R, Colvin D R, Hall J A (2007) A thin slice perspective on the accuracy of first impressions. *J. Res. Pers*. 41: 1054–1072.

10. Shelton J N, Richeson J A (2005) Intergroup Contact and Pluralistic Ignorance. *J. Pers. Soc. Psychol*. 88: 91–107.

11. Ambady N, Hallahan M, Conner B (1999) Accuracy of judgments of sexual orientation from thin slices of behavior. *J. Pers. Soc. Psychol*. 77: 538–547.

12. Roney J R, Hanson K N, Durante K M, Maestripieri D (2006) Reading men’s faces: women’s mate attractiveness judgments track men’s testosterone and interest in infants. *Proc. Roy. Soc. Lond. B. Biol. Sci*. 273: 2169–2175.

13. Stillman T F, Maner J K, Baumeister R F (2010) A thin slice of violence: distinguishing violent from nonviolent sex offenders at a glance. *Evol. Hum. Behav*. 31: 298–303.

14. Ambady N, Bernieri F J, Richeson J A (2000) *Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream.* (Academic Press).

15. Zaller J R (1986) Analysis of information items in the 1985 ANES pilot study. ANES Pilot Study Report, No. nes002261

16. Ryan J B (2011) Accuracy and bias in perceptions of political knowledge. *Polit. Behav.* 33: 335-356.

17. Bryan C J, Walton G M, Rogers T, Dweck C S (2011) Motivating voter turnout by invoking the self. *Proc. Natl. Acad. Sci. Unit. States. Am*. 108: 12653–12656.

18. Barr P S, Glynn M A (2004) Cultural variations in strategic issue interpretation: relating cultural uncertainty avoidance to controllability in discriminating threat and opportunity. *Strat. Manag. J*. 25: 59–67.

19. Bolinger, D L M (1989) *Intonation and Its Uses: Melody in Grammar and Discourse* (Stanford University Press).

20. Jiang X, Pell M D (2015) On how the brain decodes vocal cues about speaker confidence. *Cortex*. 66: 9–34.

21. Zuckerman M, DePaulo B M, Rosenthal R (1981) Verbal and nonverbal communication of deception. *Adv. Exp. Soc. Psychol*. 14: 59.

22. DePaulo B M *et al.*, (2003) Cues to deception. *Psychol. Bull*. 129: 74-118.

23. Hauch V, Blandon-Gitlin I, Masip J, Sporer S L (2015) Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Pers. Soc. Psychol. Rev*. 19: 307–342.

24. Murphy N A (2005) Using thin slices for behavioral coding. *J. Nonverbal Behav.* 4: 235-246.

25. Brunswik E (1956) *Perception and the Representative Design of Psychological Experiments* (University of California Press).

26. Nickerson D W, Rogers T (2014) Political campaigns and big data. *J. Econ. Perspect*. 28: 51–73.

27. Bassili J N (1995) Response latency and the accessibility of voting intentions: what contributes to accessibility and how it affects vote choice. *Pers. Soc. Psychol. Bull*. 21: 686–695.

28. ten Brinke L, Oberholzer-Gee F, Carney D R (2015, in preparation) Some evidence for strategic use of nonverbal signals of emotion in high-stakes economic games.

29. von Hippel W, Sekaquaptewa D, Vargas P (1995) On the role on encoding processes in stereotype maintenance. *Adv. Exp. Soc. Psychol.* 27: 177–254.

Proceedings of the National Academy of Sciences, 113(23), 6449-6453.

30. Morgan T M (2001) The economic impact of wasted prescription medication in an outpatient population of older adults. *J. Fam. Pract.* 50: 779-781.