Anomalous Responses on Amazon Mechanical Turk: An Indian Perspective*

Published in *Research & Politics*

William O'Brochta†
Department of Political Science
Washington University in St. Louis

Sunita Parikh
Department of Political Science
Washington University in St. Louis

What can researchers do to address anomalous survey and experimental responses on Amazon Mechanical Turk (MTurk)? Much of the anomalous response problem has been traced to India, and several survey and technological techniques have been developed to detect foreign workers accessing U.S.-specific surveys. We survey Indian MTurkers and find that 26% pass survey questions used to detect foreign workers, and 3% claim to be located in the United States. We show that restricting respondents to Master Workers and removing the U.S. location requirement encourages Indian MTurkers to correctly self-report their location, helping to reduce anomalous responses among U.S. respondents and to improve data quality. Based on these results, we outline key considerations for researchers seeking to maximize data quality while keeping costs low.

Keywords: Amazon Mechanical Turk, data quality, crowdwork, anomalous responses, India
(1936 words)

*We thank Mathangi Krishnamurthy, Chris Lucas, and Bryant Moy for their comments and suggestions. We acknowledge financial support from the Weidenbaum Center on the Economy, Government, and Public Policy.

†Corresponding Author: One Brookings Drive, Campus Box 1063, St. Louis, Missouri 63130, 314-935-5852, obrochtawj@wustl.edu

Amazon Mechanical Turk (MTurk) is a popular and inexpensive way for scholars, particularly those interested in American politics, to gather survey responses. Much of MTurk's value, however, rests on drawing respondents from the target population of individuals located in the U.S. and obtaining high quality responses. In 2018, researchers noticed an increase in responses to these Human Intelligence Tasks (HITs) that were a mix of anomalous responses that did not fit requester expectations and/or work from foreign, primarily Indian workers using virtual private servers (VPS) to circumvent survey location restrictions (Chmielewski and Kucker, 2020). Indians comprise a major part of the MTurk workforce, so addressing this problem is of paramount importance.

Two sets of solutions have been developed to detect and resolve these issues: question batteries designed to uncover anomalous responses (Barends and de Vries, 2019; Buchanan and Scofield, 2018) and location-based methods to block IP addresses of foreign workers attempting to access U.S.-specific HITs (Ahler, Roush and Sood, 2019; Kennedy et al., 2020). We know that foreign workers often answer these question batteries in the same ways as do U.S. respondents (Kennedy et al., 2020). Further, we also know that foreign workers successfully circumvent location-based detection methods, including VPS detection (Dennis, Goodson and Pearson, 2020; MacInnis, Boss and Bourdage, 2020). Therefore, neither of these solutions can successfully identify foreign workers who provide non-anomalous responses and appear to be located in the U.S. (Dennis, Goodson and Pearson, 2020).

Foreign respondents who access U.S.-specific surveys undetected may respond randomly, adding random noise to survey results (Kennedy et al., 2020). What is more likely is that the cultural experiences of foreign workers result in responses that are systematically different than those of U.S. workers. For example, we asked survey respondents to list the three most important government positions. While most U.S. respondents listed the President, Vice President, and Speaker of the House, 10% of Indian respondents listed the head of the post office. Had these Indian respondents accessed a U.S.-based survey, we would have incorrectly interpreted U.S. public opinion. As such, ensuring that only U.S. workers complete U.S.-

specific HITs is a high priority for maintaining MTurk data quality.

We survey Indian MTurkers ($N = 305$) and show that 26% of respondents do not provide anomalous responses and, therefore, could access U.S.-specific surveys without being identified, thereby contaminating the MTurk U.S. survey pool. Indeed, 3% of our respondents on Indian MTurk claimed to be taking the survey from a U.S. state, suggesting that they regularly access U.S.-specific surveys. We identify characteristics of workers who provide anomalous responses.

Since Indian MTurkers can pass question batteries and circumvent IP address detection, we develop a simple solution to improve U.S.-specific HIT quality: using Master Workers, removing HIT location restrictions, and asking respondents to self-identify their country of residence. We implement this solution, alongside a battery of pilot tested questions designed to induce different answers among U.S. and foreign respondents. The results ($N = 302$) indicate that over 87% of Indian MTurkers honestly self-report their location. Only four respondents used a VPS, and at most 3% of respondents with U.S.-specific IP addresses were potentially foreign respondents, compared to the 20 to 30% of suspected foreign respondents detected in surveys with U.S.-specific location restrictions (Ahler, Roush and Sood, 2019; Dennis, Goodson and Pearson, 2020). Based on our results, we provide guidance for researchers designing MTurk HITs who face trade-offs between data quality and cost.

Anomalous Responses

We conducted a survey asking Indian MTurkers about their descriptive characteristics, work on MTurk, and communication patterns.[1] Embedded in our survey were five measures of anomalous responses (Table 1), which we argue represent the main anomalous responses requesters observe when conducting surveys and experiments on MTurk.[2] Thirty-six percent of respondents failed our attention check question. We developed two measures to assess

[1] Institutional Review Board Approval #201911090.

[2] See the Supplemental Information (SI) for more details about each part of this article.

whether respondents provided consistent answers: 18% reported working more hours on MTurk per week than the total number of hours worked, and 42% reported having completed fewer HITs than were required to be eligible for the survey. Fully 66% of respondents did not write a coherent description of their typical day working on MTurk, our measure of effort. There is also evidence that Indian MTurkers do use location-based technology to access U.S. surveys: 11% obfuscated their location and 3% actually claimed to live in a U.S. state, even though the survey was only available to MTurkers located in India. Overall, while 74% of Indian MTurkers failed at least one of these five measures, 26% did not, meaning that their non-anomalous responses would go undetected if they accessed U.S.-specific surveys.

Table 1: Measures of Anomalous Responses

Type	Measure	Anomalous	Anomalous %
Inattentive	Attention Check	Anomalous response if fails a standard attention check question that asked respondents to choose "3" on a 5-point Likert scale.	36
Inconsistent	Hours Worked	Anomalous response if answers questions asking total weekly hours worked and total weekly hours worked on MTurk such that the total hours worked are fewer than the total hours worked on MTurk.	18
Inconsistent	HITs	Respondents had to have completed 1000 total HITs to be eligible to take this survey. Anomalous response if answers question about the total number of HITs completed with a number less than 1000.	42
Low effort	Description	Respondents were asked to describe a typical day working on MTurk in an open text box. Anomalous response if description provided in no way answered the question.	66
Obfuscating	State	Respondents were asked to report the state where they lived in an open text box. Anomalous response if the respondent did not report their state (instead saying "India" or the "United States").	11

Five measures of four dichotomous types of anomalous responses. Criteria for being counted as an anomalous response shown along with the percentage of anomalous responses for each measure. Mean of 1.68 anomalous responses per respondent.

What is more, respondents who provided one or more anomalous responses were more likely to engage in other forms of inattentive behavior like satisficing and hurrying (see SI.2).

Taken together, these results emphasize the need to identify and to reduce the number of foreign respondents in U.S.-based surveys: they are not the target population, and a high number of anomalous responses can negatively impact data quality.

Explaining Anomalous Responses

Who provides anomalous responses? We use Bayesian Model Averaging (BMA) (Montgomery and Nyhan, 2010) to explore relationships between respondent demographic and workplace characteristics and anomalous responses. BMA generates all possible model specifications and determines the proportion of models where each independent variable is included (the posterior inclusion probability — PIP).

There are three key variables that differentiate Indian MTurkers providing anomalous and non-anomalous responses (see SI.3). Respondents who receive help starting on MTurk and those who have household members working on MTurk were more likely to provide an anomalous response. Respondents who had worked on MTurk longer were less likely to provide an anomalous response. Experienced workers know how to navigate the cutthroat MTurk system by forming relationships, sharing information only with key friends, and treating work on MTurk as a way earn some extra money while doing HITs of their choosing. Much of the work available to Indian MTurkers is repetitive, particularly low paying, and does not value worker experience. Hence, these MTurkers are more likely to be interested in and to have the skills to be able to navigate the process of completing U.S.-specific surveys undetected.

Addressing Anomalous Responses

How can requesters incentivize high-quality Indian workers who pass all five measures of anomalous responses not to take U.S.-specific surveys when they could do so successfully? Based on our experience working with Indian MTurkers, we propose limiting the respondent

pool to Master Workers and removing all location-specific requirements (see SI.4). While foreign respondents use increasingly advanced methods to circumvent technology designed to prevent them from accessing U.S.-based surveys, workers in our approach have no incentive to obfuscate their location. Requesters can then filter out non-U.S.-specific responses after the HIT is complete.

We test this design by fielding a survey among Master Workers with no location requirement ($N = 302$). The survey was similar to our first survey with the addition of questions asking about the respondents' country of residence and six pilot tested questions designed to elicit different responses among U.S. and Indian workers. See SI.5 for details. Only four respondents (1.32%, three in the U.S. and one in India) used a VPS. Among Indian respondents, 87.95% self-reported that they were in India. None of the five Indian respondents who claimed to live in the U.S. masked their IP address, so their location was easily detected using IP address detection techniques.

We want to know whether any respondents whose IP address locates them in the U.S. are actually Indian respondents who circumvented VPS and anomalous responses checks. Note that existing IP address detection methods cannot detect these respondents because their IP address indicates that they are in the U.S. 81.43% percent of respondents with IP addresses in the U.S. passed all four measures of anomalous responses and all six items used to differentiate U.S. and Indian respondents. Of the 15.71% who passed nine of ten measures, there was no pattern in these errors, suggesting that they were in fact respondents in the U.S. who made a mistake.

Six respondents (3.51%) passed eight or fewer questions. This is a substantially smaller percentage of potentially foreign respondents than identified in previous surveys (Ahler, Roush and Sood, 2019; Dennis, Goodson and Pearson, 2020). These respondents largely passed our items designed to determine if they were U.S. residents or not, but received a lower score because they provided anomalous responses to other questions. There is some chance that these respondents are foreign workers taking U.S.-specific surveys and circumventing

VPS detection, but it is more likely that they are either inattentive U.S. citizens or people not fluent in English.

Discussion and Conclusion

Limiting survey respondents to Master Workers and removing location requirements disincentivizes Indian respondents from claiming to be located in the U.S. and from gaining access to U.S.-specific work. Though few Indian MTurkers are Master Workers, opening surveys designed for U.S. participants to foreign workers trades off increased confidence that U.S. respondents can be successfully identified for increased cost. Table 2 describes common ways researchers restrict HITs to certain types of workers and use analytic techniques to filter out non-U.S. based respondents (see SI.6 for details). The Table shows that more stringent analytic techniques cost more while increasing confidence that U.S.-based respondents can be successfully identified.

We focus on the last three rows of the Table, where confidence is highest. Surveys open to all respondents will have more anomalous responses than surveys restricted to Master Workers, both because non-Master Workers are less attentive and because foreign workers can access U.S.-based surveys. These factors make it worth restricting survey respondents to Master Workers, who Loepp and Kelly (2020) show are not demographically different from non-Master Workers. Dropping the location restriction means collecting responses from foreign workers, which increases costs, but provides the surest way to discourage foreign workers from obfuscating their location or circumventing location requirements. This technique also allows requesters to compare self-identified foreign respondents to anomalous U.S. responses in order to see whether anomalous U.S. responses are likely to be from foreign workers, which validates the quality of the sample.

Requesters should think carefully about the trade-off between confidence and cost when selecting a survey approach. We test a method that removes existing incentives for foreign

Table 2: Strategies to Reduce Anomalous Responses

HIT Restrictions	Analytic Techniques	Cost pp.	Confidence U.S.-Based Respondents Can Be Identified
None	None	$1.20	None
U.S.	None	$1.20	Low
U.S.	Anomalous data response checks	$1.71	Medium Low
U.S.	IP address detection	$1.41	Medium
U.S.	Anomalous data response checks and IP address detection	$2.18	Medium High
U.S., Master	Anomalous data response checks and IP address detection	$1.62	High
Master	Anomalous data response checks and IP address detection	$2.19	Very High

HIT Restrictions are Amazon features to restrict survey eligibility. We assume all HITs are restricted to an approval percentage above 98%. Analytic techniques are requester strategies to filter out anomalous responses. We assume that IP address detection occurs after the survey is complete. Cost pp. is the estimated cost per non-anomalous, U.S. response. Cost estimate is based on $1.00 payment, 30% of non-Master and 20% of Master Workers providing anomalous responses, 3% of workers using a VPS, and 80% of Master Workers being from the U.S. 20% Amazon Fee plus 5% Master Worker fee.

workers to defeat location blocking technologies. As such, requesters need not continue to invest in trying to block foreign workers from U.S.-based surveys and instead can focus on building positive, long-lasting relationships with reliable MTurk workers. Obtaining foreign survey responses for requesters only interested in U.S. respondents is an added cost that may not be feasible for some requesters, but the ability to compare responses from self-identified foreign workers to anomalous U.S.-based respondents adds a new layer of previously unidentified data quality checks. Our work suggests that carefully considering the incentives and motivations of foreign respondents to access U.S.-based surveys can lead to new data quality solutions that are beneficial for both requesters and workers.

References

Ahler, Douglas J, Carolyn E Roush and Gaurav Sood. 2019. Micro-Task Market for Lemons. In *Meeting of the Midwest Political Science Association*. Chicago, Illinois: .

Barends, Ard J. and Reinout E. de Vries. 2019. "Noncompliant Responding: Comparing Exclusion Criteria in MTurk Personality Research to Improve Data Quality." *Personality and Individual Differences* 143:84–89.

Buchanan, Erin M. and John E. Scofield. 2018. "Methods to Detect Low Quality Data and Its Implication for Psychological Research." *Behavior Research Methods* 50:2586–2596.

Chmielewski, Michael and Sarah C. Kucker. 2020. "An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results." *Social Psychological and Personality Science* 11(4):464–473.

Dennis, Sean A., Brian M. Goodson and Christopher A. Pearson. 2020. "Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures." *Behavioral Research in Accounting* 32(1):119–134.

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Ryan Jewell and Philip Waggoner. 2020. "The Shape of and Solutions to the MTurk Quality Crisis." *Political Science Research and Methods* OnlineFirst.

Loepp, Eric and Jarrod T. Kelly. 2020. "Distinction without a Difference? An Assessment of MTurk Worker Types." *Research & Politics* 7(1):1–8.

MacInnis, Cara C., Harrison C.D. Boss and Joshua S. Bourdage. 2020. "More Evidence of Participant Misrepresentation on Mturk and Investigating Who Misrepresents." *Personality and Individual Differences* 152.

Montgomery, Jacob M. and Brendan Nyhan. 2010. "Bayesian Model Averaging: Theoretical Developments and Practical Applications." *Political Analysis* 18(2):245–270.

Supplemental Information: Anomalous Responses on Amazon Mechanical Turk: An Indian Perspective

Contents

Replication data and code for all empirical analysis is provided on the Harvard Dataverse.

SI.1: Anomalous Responses

We built our survey around four types of anomalous responses. Previous research has shown that Indian MTurkers produce significantly lower quality work compared to American workers (Antin and Shaw, 2012; Kazai, Kamps and Milic-Frayling, 2012). After identifying issues with the quality of crowdsourced surveys, most researchers have turned to technical or data driven solutions. These solutions involve debates about attention check or screening questions (Abbey and Meloy, 2017; Hydock, 2018), restricting samples to high reputation workers (Loepp and Kelly, 2020; Matherly, 2019), and building new survey panels (Sharpe Wessling, Huber and Netzer, 2017). Another strand of research has compared MTurk samples to other convenience samples (typically students) and has concluded that MTurk data is generally better quality than such samples (Anson, 2018; Necka et al., 2016). Even with these assurances and potential corrections, MTurk workers may still provide anomalous responses (Barends and de Vries, 2019). In addition, dropping subjects who fail attention checks or other items introduces significant biases (Aronow, Baron and Pinson, 2019) and leads to an even more unrepresentative sample (Berinsky, Margolis and Sances, 2014). We focus on how demographic and workplace characteristics influence Indian respondents' tendency to provide anomalous responses in order to develop more effective solutions to the quality problem that do not rely exclusively on circumventable technological fixes.

First, we include a standard attention check as a measure of inattentiveness. The question asks respondents to select option "3" on a 5 point scale. Fully thirty-six percent of respondents failed this attention check, indicating that at least that percentage of respondents were simply not paying attention while taking the survey. This large proportion of inattentive respondents is in line with existing published work that uses an Indian MTurk sample. Researchers have dropped up to 50% of their initial Indian MTurk sample because

of failed attention checks, unusual responses, or attrition (Chandler and Paolacci, 2017; Lawson et al., 2010; Spears, 2013; Yudkin et al., 2016). Interestingly, we found some evidence of satisficing in the incorrect responses to this question. Only 5 respondents selected "1" or "2," but 105 respondents selected "4" or "5" potentially in order to show that they thought the requester wanted to see agreement to their questions.

Second, we asked two questions meant to test whether respondents provided inconsistent answers to survey questions. Inconsistencies in survey responses are usually how requesters discern that there are anomalous responses in their survey. For example, many surveys include multiple items designed to test the same theoretical argument. If the correlation between these items is extremely low, some respondents may be providing inconsistent answers. We asked two sets of questions that required different types of consistency. Our first set of questions asked about the number of hours the respondent worked in a given week, one question about the number of hours worked total and the other question about the number of hours worked on MTurk. Obviously, the number of hours worked on MTurk cannot exceed the total number of hours worked. Eighteen percent of respondents reported working more hours per week on MTurk than they worked overall. This signals a problem with internal consistency: the hours worked questions appeared one after another, so it is unlikely that the respondent forgot what she answered to the overall hours question such that she would provide an inconsistent response to the MTurk hours question.

We constructed a similar indicator and test for the number of HITs a respondent completed. In this question, we asked "about how many HITs have you completed?" In order to be qualified to take the survey, a respondent needed to have completed a minimum of 1000 HITs, so anyone reporting fewer than 1000 provided an inconsistent response. Forty-two percent of respondents failed to answer at least 1000 HITs. In order to avoid inconsistencies, the respondent needed to know that the survey was only open to those who had completed at least 1000 HITs and also to know the number of HITs she had completed. While this information is readily available, respondents needed to recall more information in order to provide a consistent response to this question.

Inconsistencies and inattentiveness may signal a larger issue with respondents failing to put in the required effort to answer the survey questions in an expected way. Nowhere is lack of effort more apparent and obvious than in open text responses. To successfully complete an open text response, respondents need to fully understand the question, to formulate an answer, and to write the answer. Indian respondents attempting to complete work meant for U.S. respondents may end up providing an anomalous response because of lack of effort on one or more of these requirements. Instructions often do not make sense, especially when respondents are trying to interpret American English as quickly as possible (Feitosa, Joseph and Newman, 2015; Milland, 2017). Additionally, even if respondents fully understand what is being asked of them, they must also be able to communicate their response, another often challenging task when English is not the respondents' first language (Hauser, Paolacci and Chandler, 2019).

Respondents were asked to "describe their typical day working on MTurk" and were provided a text box in which to complete this task. Sixty percent of respondents failed to provide a suitable answer. Anomalous responses ranged from those copied from the Internet to nonsensical responses like "good." We also observed 66 responses (22%) that were very similar to or duplicates of other responses. These responses sometimes, but not always, came

from a duplicated IP address. Clearly some coordination was going on among respondents. We think of this indicator as a measure of performance: does the respondent put in sufficient effort to write a meaningful answer to this question in English?

Despite pleas from requesters to answer truthfully and to spend time answering carefully, Indian workers have few incentives to do so because pay is low and work is anonymous (Miura and Kobayashi, 2016; Sharpe Wessling, Huber and Netzer, 2017). This is especially relevant if Indian MTurkers are gaining access to work restricted to U.S. respondents. Workers may, therefore, adopt a persona that more closely aligns to the type of respondent they believe that requesters expect (Kaufmann, Schulze and Veit, 2011; MacInnis, Boss and Bourdage, 2020). Alternatively, workers may decide not to disclose personal information simply because they have no incentive to do so, as the anonymous nature of MTurk surveys means that most demographic information cannot be independently verified. We asked one question about the state or union territory in which respondents lived. Though we specified "state or union territory" to ensure respondents did not misunderstand and state "India," a number of respondents provided such a response (8%). Other respondents (3%) responded that they lived in the United States. We cross-referenced the states provided with IP address geolocation, but because IP address geolocation is often inaccurate, we only counted respondents as providing an anomalous response if they reported living in the United States, India (without listing their Indian state), or some other response that did not make sense. Those respondents who reported living in the United States were clearly misrepresenting themselves.

We dichotomize each of these five measures of the four types of anomalous responses where 1 is an anomalous response and 0 otherwise. We also construct an indicator of whether a respondent provided at least one anomalous response where 1 is at least one anomalous response and 0 is no anomalous responses.

SI.2: Survey Details

In order to assess who provided anomalous responses on Indian MTurk we fielded a survey on MTurk from December 18-23, 2019.[1] To obtain a balance of MTurk Masters and regular workers, we restricted the first 82 respondents to only MTurk Masters. Otherwise, respondents needed to have HIT approval rates above 98% and to have completed at least 1000 HITs. Respondents were paid $0.75 for completing this 10 minute survey, and we obtained 305 completed responses.

- Pre-Survey Questions:
    1. ResponseNumber: Number of the response
    2. StartDate: Date/time survey was started
    3. EndDate: Date/time survey ended
    4. IPAddress: Respondent reported IP address, redacted per journal instructions
    5. Progress: 100 if survey was finished and submitted; 98 if finished, but not submitted
    6. Duration: Duration the survey was active in seconds
    7. Finished: 1 if progress is 100
    8. Latitude: Reported respondent location
    9. Longitude: Reported respondent location
- MTurk Usage:
    10. HoursWorked: How many hours do you work per week? (Enter number of hours)
    11. HoursWorkedMTurk: Of the total number of hours you work per week, how many hours per week do you work on MTurk? (Enter number of hours)
    12. MTurkPrimaryJob: Do you consider working on MTurk to be your primary job? (1-Yes, 0-No)
        - PrimaryJobText: If no: How would you describe your primary job? (Enter primary job information)
        - PrimaryJobCoded: PrimaryJobText recoded into: Various (works in many jobs to support family), Invalid (invalid response), Sales, Manager, IT (including software development), Engineer, Teacher (including professor), Employee (general employee — not in management — at a private company), Self Employed
    13. HITsOneTime: How many HITs do you typically complete at one time? (Enter a number)
- MTurk Origin:

[1] This study was approved by the Institutional Review Board #201911090.

14. FindOut: How did you first learn about MTurk? (Choose one: 1-friend, 2-someone you work with, 3-family, 4-job website, 5-online forum, 6-news article, 7-web search, 8-other)
    - FindOutText: If other: Please specify how you first heard about MTurk
    - FindOutCoded: Recode FindOutText to Invalid (invalid response): all respondents who selected other in FindOut provided invalid responses in FindOutText.
15. Months: How many months have you worked on MTurk? (Enter a number of months)
16. StartMotivation: Which of the following was the reason you were most motivated to start working on MTurk? (Check one: 1-sense of purpose, 2-to kill time, 3-to have fun, 4-to make extra money, 5-to do interesting tasks)
17. StartHelp: Did anyone help you get started working on MTurk? (0-no, 1-yes)
    - StartHelpText: If yes: Please describe how you knew the person who helped you get started working on MTurk and what they did to help you
    - StartHelpCoded: Recode StartHelpText into Social Media (including Twitter and online forums), Invalid, Friend, Family, Neighbor, MTurk Business (a supervisor or business related to MTurk), Don't Remember, Colleague (from work), Many People
    - StartHelpFlag: Response in StartHelpText seems duplicated from another survey respondent. 1 if flagged
    - StartHelpFlagMatch: ResponseNumber of matching/duplicated responses
18. HITsCompleted: About how many HITs have you completed? (Enter a number)

- MTurk Networking:

19. KnowOnMTurk: How many people do you know that work on MTurk? (Enter number of people)
20. TalkMTurk: How many people have you talked to about working on MTurk? (Enter number of people)
21. TalkFrequency: About how frequently do you talk with people about your work on MTurk? (1-less than once per month, 2-once every few weeks, 3-several times each week, 4-almost every day, 5-more than once per day)
22. HelpJoinNumber: How many people have you helped join or get started working on MTurk? (Enter number of people)
    - If at least 1 person: Who have you helped join MTurk? (Choose all that apply: friend, family, neighbor, someone you work with, other)
        * If other: Please describe your relationship to the people you have helped join MTurk
    - Recoded "Who have you helped join MTurk?" into separate columns: HelpFriend, HelpFamily, HelpNeighbor, HelpColleague, and HelpOther where 1 indicates

that the respondent helped that person join MTurk. HelpOtherText is the text field for HelpOther. HelpOtherCoded is either Invalid or Acquaintances and Strangers. Response 270 entered Friend, which was recoded to HelpFriend and coded as Invalid in HelpOtherCoded

23. WorkerFriendships: I have developed friendships with other MTurk workers (1-strongly disagree to 5-strongly agree)
24. How do you usually find out about HITs? (Choose all that apply: WhatsApp, Facebook, Reddit, friend, family member, MTurk website, SMS message, other)
    - If other: Please describe how you find out about HITs
    - Recoded into separate columns: DiscoverWhatsApp, DiscoverFacebook, DiscoverReddit, DiscoverFriend, DiscoverFamily, DiscoverMTurk, DiscoverSMS, and DiscoverOther
    - DiscoverOtherText is taken from the other text field. DiscoverOtherCoded is Invalid because respondents filling out DiscoverOtherText simply repeated the choices that they had selected.
25. DiscoverThisHIT: How did you find out about this HIT? (Choose one: 1-WhatsApp, 2-Facebook, 3-Reddit, 4-friend, 5-family member, 6-MTurk website, 7-SMS message, 8-other)
    - DiscoverThisHITText: If other: Please describe how you found out about this HIT. No respondents selected Other for DiscoverThisHIT, so this question was not displayed to any respondents
26. Do you belong to any of the following groups? (Choose all that apply: WhatsApp, Facebook, Reddit, other group)
    - If other: Please write other groups related to MTurk of which you are a part
    - Recoded into separate columns: GroupWhatsApp, GroupFacebook, GroupReddit, GroupOther, and GroupNone
    - GroupOtherText is taken from the other text field. GroupOtherCoded is Telegram, SMS, Panda, E-mail, Instagram, Crowdworkers, Turkerview, Mturkcrowd
    - Recode respondent 33 to add GroupFacebook, 136 to add GroupReddit, 248 to add GroupNone and remove GroupOther, and 278 to add GroupReddit and remove GroupOther.
27. ShareFrequency: How often do you share information with others about an MTurk HIT or about working on MTurk? (1-never to 5-very frequently)
    - If at least a 2: With whom do you share information about an MTurk HIT or about working on MTurk? (Choose all that apply: friend, family member, neighbor, someone at work, other)
    - Recode into separate columns: ShareFriend, ShareFamily, ShareNeighbor, ShareColleague, and ShareOther
28. AttentionCheck: The answer to this item should be 3 or neither agree nor disagree so we know to keep your data. (1-strongly disagree to 5-strongly agree)

- MTurk Workplace:

  29. DeviceUsed: What device do you usually use to complete MTurk HITs? (Choose one: 1-phone, 2-tablet, 3-laptop, 4-desktop)
  30. DeviceOwned: Do you own the device you use to complete MTurk HITs? (0-no, 1-yes)
  31. DeviceShared: Does anyone else use the same device to work on MTurk? (0-no, 1-yes)
  32. Location: Where do you usually complete MTurk HITs? (Choose one: 1-at home, 2-at work, 3-at a coffee shop/cafe, 4-other location)
      - LocationOtherText: If other location: Where do you complete MTurk HITs?
      - LocationOtherCoded: All responses were Invalid. Recode respondent 91 to Location=1 and 148 to Location=2
  33. LocationOtherWorkers: Do you complete MTurk HITs at a location where other MTurk workers are working? (0-no, 1-yes)
  34. LocationOthersHousehold: Do any members of your household also work on MTurk? (0-no, 1-yes)

- MTurk Tasks:

  35. AcademicSurveys: What percent of MTurk HITs that you complete are academic surveys? (Enter a percentage)
  36. Please describe a typical day working on MTurk like you might summarize your workday to a friend. We are not asking for any personal or identifying information. We are interested in how you day goes, the kinds of people you interact with during the day, and the way you work on MTurk. Feel free to write anything that comes to mind. You will be asked to write for at least one minute. (Free response for one minute)
      - TimingStart: Time after the page was displayed that the respondent clicked on the page for the first time
      - TimingEnd: Time after the page was displayed that the respondent clicked on the page for the last time
      - TimingSubmit: Time after the page was displayed that the respondent clicked the next button (at least 60 seconds)
      - TimingClicks: Number of clicks on the page
      - TimingText: Text response
      - TimingInvalid: Responses that do not answer the question. Coded 1 if invalid
      - TimingFlag: Coded 1 if the response seems substantively similar to other responses
      - TimingFlagMatch: Row numbers of similar responses
      - TimingTimeofDay: All Day (including both morning and afternoon), Free Time, Morning (including early morning), Evening (including after work), Morning and Evening

  - TimingDuration: Number of hours worked per day in hours
  - TimingActivities: What respondents report doing while working on MTurk. Coded as either Television or Internet
  - TimingInteraction: Who respondents collaborate with about MTurk during the day. Coded as Friend (one friend), Friends (multiple friends), Family, Colleagues (from work), Neighbor, Social Media, or None (explicitly mentions working alone)
  - TimingMoney: Coded as Highest paid (respondent selects highest paying HITs first), Good pay (respondent reports being paid well on MTurk), or low pay (respondent reports low pay on MTurk).
  - TimingTasks: The tasks the respondent likes doing. Coded as Surveys, Batches (including image processing, data entry, and translation tasks), Surveys and Batches, or Anything

- Demographics:

37. Female: What is your gender? (0-male, 1-female)
38. Age: What is your age? (Enter your age in years). Recoded those who reported a year of birth instead of age
39. Married: Are you married? (0-no, 1-yes)
40. Education: What is the highest educational level you have attained? (Choose one: 1-no formal education, 2-incomplete primary school, 3-completed primary school, 4-incomplete intermediate school, 5-complete intermediate school, 6-Bachelors, 7-above Bachelors)
41. IncomeLadder: Indicate your household's income group where 1 indicates the lowest income group in India and 10 the highest income group in India. (1-lowest income group to 10-highest income group)
42. Caste: Are you: (Choose one: 1-Brahmin, 2-General/Forward, 3-SC, 4-ST, 5-OBC, 6-non-Hindu/other religion)
43. State: In what state or union territory do you live? (Enter state or union territory name)
    - StateCoded: Standardized version of State (reflecting what the respondents said, not necessarily corresponding to the latitude and longitude where they were located or their IP address). Options are Tamil Nadu, Telangana, Kerala, Karnataka, Delhi, West Bengal, Maharashtra, Haryana, Gujarat, and Andhra Pradesh. Additionally, there is an Invalid option, an option for respondents who said "India" instead of the state or union territory, and an option for respondents who claimed to be in the United States
44. NativeLanguage: What is your native language? (Enter native language)
    - NativeLanguageCoded: Coded version of self-reported native language: Tamil, English, Tamil and English, Malayalam, Hindi, Gujurati, Bengali, Urdu, Telugu, Marathi, and Invalid

45. RandomID: Generated by Qualtrics to complete the survey
46. MasterWorker: Respondents recorded before 9:30AM on December 21 were required to be MTurk Masters

Tables SI.2.1 and SI.2.2 display descriptive statistics for the main independent variables used in the analysis. We ran these descriptive statistics among both all respondents and those responses who passed the five anomalous response checks. It is a bit unclear how we should interpret the answers to the survey questions for respondents providing anomalous responses. It is of course possible that respondents misrepresented themselves, satisficed, or chose random responses to some if not all of these questions. Therefore, it will be difficult to interpret any differences between how all respondents answer a question and how only those respondents who passed the anomalous response checks. For now, we focus on the descriptive statistics for valid respondents, as they are more conservative than the descriptive statistics for all respondents.

Respondents communicate about HITs with a small group of friends. Friends told respondents about the MTurk platform, and respondents helped a couple of friends join MTurk. At least a fifth of respondents belonged to an online group like those on WhatsApp or Facebook where they can learn about HITs or share information. More commonly, over a third of respondents report sharing information with friends about MTurk HITs with information sharing occurring semi-regularly. This seems to have led to relationships forming with respondents helping friends working on MTurk and talking to a number of other MTurk workers. Most respondents work from home, though a much larger proportion of respondents providing anomalous responses reported working from the office and locations with other workers.

Table SI.2.1: Descriptive Statistics for All Responses

Variable	Min	Max	SD	Mean	Median
Female	0.00	1.00	0.42	0.23	0.00
MasterWorker	0.00	1.00	0.46	0.30	0.00
Age	2.00	62.00	6.20	30.34	28.00
Married	0.00	1.00	0.47	0.67	1.00
Months	1.00	310.00	36.45	38.22	24.00
GroupWhatsApp	0.00	1.00	0.50	0.48	0.00
GroupFacebook	0.00	1.00	0.48	0.36	0.00
ShareFriend	0.00	1.00	0.50	0.53	1.00
ShareFamily	0.00	1.00	0.40	0.20	0.00
ShareFrequency	1.00	5.00	1.16	2.93	3.00
AcademicSurveys	1.00	100.00	30.45	66.85	75.00
LocationHome	0.00	1.00	0.50	0.53	1.00
LocationOtherWorkers	0.00	1.00	0.48	0.66	1.00
LocationOthersHousehold	0.00	1.00	0.49	0.59	1.00
StartMotivationMoney	0.00	1.00	0.50	0.50	0.00
TalkFrequency	1.00	5.00	1.17	2.83	3.00
MTurkPrimaryJob	0.00	1.00	0.33	0.88	1.00
FindOutFriends	0.00	1.00	0.49	0.40	0.00
FindOutWeb	0.00	1.00	0.40	0.20	0.00
KnowOnMTurk	0.00	100000.00	5725.52	370.87	10.00
TalkMTurk	0.00	800.00	66.57	24.61	8.00
HelpJoinNumber	0.00	1000.00	72.67	16.98	4.00
HelpFamily	0.00	1.00	0.47	0.32	0.00
HelpFriend	0.00	1.00	0.44	0.73	1.00
WorkerFriendships	1.00	5.00	1.07	4.22	5.00
DiscoverFriend	0.00	1.00	0.48	0.36	0.00
DiscoverWhatsApp	0.00	1.00	0.41	0.21	0.00
CasteOBC	0.00	1.00	0.49	0.40	0.00
IncomeLadder	1.00	10.00	2.08	4.76	5.00
NativeLanguageEnglish	0.00	1.00	0.46	0.30	0.00
StartHelp	0.00	1.00	0.41	0.79	1.00

Table SI.2.2: Descriptive Statistics for Responses Without Anomalous Items Only

Variable	Min	Max	SD	Mean	Median
Female	0.00	1.00	0.43	0.24	0.00
MasterWorker	0.00	1.00	0.50	0.59	1.00
Age	23.00	62.00	8.26	35.00	33.50
Married	0.00	1.00	0.42	0.77	1.00
Months	5.00	310.00	42.99	67.76	60.00
GroupWhatsApp	0.00	1.00	0.40	0.19	0.00
GroupFacebook	0.00	1.00	0.39	0.18	0.00
ShareFriend	0.00	1.00	0.49	0.37	0.00
ShareFamily	0.00	1.00	0.38	0.17	0.00
ShareFrequency	1.00	5.00	1.09	2.41	2.00
AcademicSurveys	1.00	100.00	34.35	39.92	30.00
LocationHome	0.00	1.00	0.32	0.88	1.00
LocationOtherWorkers	0.00	1.00	0.42	0.22	0.00
LocationOthersHousehold	0.00	1.00	0.42	0.22	0.00
StartMotivationMoney	0.00	1.00	0.46	0.69	1.00
TalkFrequency	1.00	5.00	1.09	2.06	2.00
MTurkPrimaryJob	0.00	1.00	0.49	0.60	1.00
FindOutFriends	0.00	1.00	0.50	0.47	0.00
FindOutWeb	0.00	1.00	0.45	0.28	0.00
KnowOnMTurk	0.00	100000.00	11318.66	1343.55	3.50
TalkMTurk	0.00	450.00	53.41	18.77	5.00
HelpJoinNumber	0.00	500.00	56.42	8.73	2.00
HelpFamily	0.00	1.00	0.39	0.18	0.00
HelpFriend	0.00	1.00	0.46	0.69	1.00
WorkerFriendships	1.00	5.00	1.42	3.37	4.00
DiscoverFriend	0.00	1.00	0.42	0.23	0.00
DiscoverWhatsApp	0.00	1.00	0.29	0.09	0.00
CasteOBC	0.00	1.00	0.50	0.42	0.00
IncomeLadder	1.00	7.00	1.48	4.76	5.00
NativeLanguageEnglish	0.00	1.00	0.29	0.09	0.00
StartHelp	0.00	1.00	0.50	0.49	0.00

In addition to describing respondents who provided anomalous and non-anomalous responses, we also checked to see the extent to which respondents engaged in two common inattentive behaviors: satisficing and hurrying. As mentioned in SI.1, our first evidence for satisficing comes from the attention check question, where we asked respondents to select "3" on a Likert scale. Respondents who failed the attention check tended to systematically select a 4 or 5 more frequently than a 1 or 2, suggesting that they thought the researchers wanted "more agreeable" responses. To investigate this phenomenon further, we combined three survey questions that all use the same 5 point Likert scale: TalkFrequency, WorkerFriendships, and ShareFrequency. While it makes sense that the responses to these questions are correlated, we have no reason to believe that respondents who engage in these behaviors should systematically be more likely to provide anomalous responses. However, the grand mean of these three survey items is 10.72 for respondents who provided at least one anomalous response and 7.85 for respondents who did not provide any anomalous responses (t=8.19, p-value=0.00). The same result holds when comparing Master Workers and non-Master Workers (Master Worker mean=8.53, non-Master Worker mean=10.59, t=5.97, p=0.00). Therefore, we have additional evidence of satisficing to add to our observations from the attention check question.

We also assessed whether individuals providing anomalous responses were more likely to rush through the survey. There was no significant difference in survey duration between anomalous and non-anomalous respondents (anomalous mean=894 seconds, non-anomalous mean=939, t=0.34, p=0.74) or between Master and non-Master workers (Master Worker mean=923 seconds, non-Master Worker mean=898, t=0.18, p=0.86). Recall that the survey contained an open text box response where respondents were asked to describe their typical day working on MTurk. Because this question required respondents to be on the page for at least 60 seconds before proceeding, some respondents left the survey and returned much later to complete it. For this reason, the overall duration of the survey is not a particularly good measure of whether respondents hurried or not. We think that the number of words that respondents provided to the question asking them to describe their typical day working on MTurk is a better measure because it indicates effort. Respondents who did not provide anomalous responses provided longer answers than respondents who did provide anomalous responses (anomalous mean=19 words, non-anomalous mean=47 words, t=7.97, p=0.00); the same was true for Master and non-Master Workers (Master Worker mean=39 words, non-Master Worker mean=20, t=5.71, p=0.00).

SI.3: Empirical Analysis and Results

We want to see how respondent demographics and workplace behavior influences the likelihood of providing an anomalous response. Our survey contains a large number of potential independent variables that could predict anomalous responses. We do not have strong theoretical claims about how every survey item will predict anomalous responses, just that demographics and workplace behavior will generally influence anomalous responses. Including each survey item as an independent variable will guarantee overfitting.

To address this issue, we employ Bayesian Model Averaging (BMA) for the set of variables listed below. These variables include a mix of demographic characteristics, closed ended Likert scale questions, and questions where respondents were asked to select different mediums of communication or to provide text responses. We selected these independent variables based on the variation we saw in the descriptive statistics presented earlier. For example, most respondents who belonged to a social media group belonged to a WhatsApp and or Facebook group. Therefore, including other types of groups asked in the survey is relatively uninformative. Some questions were totally uninformative: the vast majority of respondents discovered this HIT on the MTurk website, so this question was not helpful in distinguishing respondents from one another. The variables listed below were included in the BMA.

Variables included in BMA: "MTurkPrimaryJob," "FindOutFriends," "FindOutWeb," "Months," "StartMotivationMoney," "StartHelp," "KnowOnMTurk," "TalkMurk," "TalkFrequency," "HelpJoinNumber," "HelpFamily," "HelpFriend," "WorkerFriendships," "DiscoverFriend," "DiscoverWhatsApp," "GroupWhatsApp," "GroupFacebook," "ShareFrequency," "ShareFriend," "ShareFamily," "LocationHome," "LocationOtherWorkers," "LocationOthersHousehold," "AcademicSurveys," "Female," "Age," "Married," "IncomeLadder," "CasteOBC," "NativeLanguageEnglish," "MasterWorker."

Bayesian Model Averaging is a technique that uses criteria based model selection to identify independent variables that are important predictors of the dependent variable. In essence, we are able to identify independent variables that predict the dependent variable by engaging in a full and complete search of all possible combinations of independent variables instead of haphazardly guessing which set of independent variables should be included. We input the above list of independent variables and run BMA on each of the five measures of anomalous responses. Output consists of a list of independent variables that predict a given anomalous response type. All other independent variables were dropped.

Table SI.3.1 displays the characteristics with a posterior inclusion probability (PIP) exceeding 50% along with their posterior mean and standard deviation. The PIP is the estimate of the percentage of model runs where the variable was included; a PIP higher than 50% is generally considered a variable worth including (Montgomery and Nyhan, 2010).

Table SI.3.1: Characteristics and Anomalous Responses

	PIP	Post Mean	Post SD
Attention			
WorkerFriendships*	97.10	0.63	0.24
ShareFrequency*	86.00	-0.35	0.19
Months#	85.30	-0.01	0.01
LocationOthersHousehold	82.70	0.79	0.47
AcademicSurveys%	58.20	0.01	0.01
StartHelp	55.50	0.62	0.66
Hours			
DiscoverFriend	98.30	-1.27	0.43
MasterWorker	81.10	-1.19	0.76
StartHelp	58.80	0.91	0.92
Months#	57.50	-0.01	0.01
HITs			
Age#	100.00	-0.16	0.04
LocationHome	100.00	-1.31	0.33
StartMotivationMoney	100.00	-1.25	0.32
ShareFriend	98.50	-1.15	0.40
StartHelp	88.70	1.23	0.63
HelpFamily	74.50	0.66	0.48
FindOutFriends	68.60	-0.61	0.51
Description			
KnowOnMTurk#	100.00	-0.00	0.00
LocationOtherWorkers	91.10	1.10	0.51
WorkerFriendships*	88.20	0.42	0.22
Age#	85.10	-0.07	0.04
MTurkPrimaryJob	84.50	1.34	0.80
LocationOthersHousehold	66.10	0.58	0.50
State			
Months	67.90	-0.01	0.01
LocationOthersHousehold	55.40	0.72	0.76

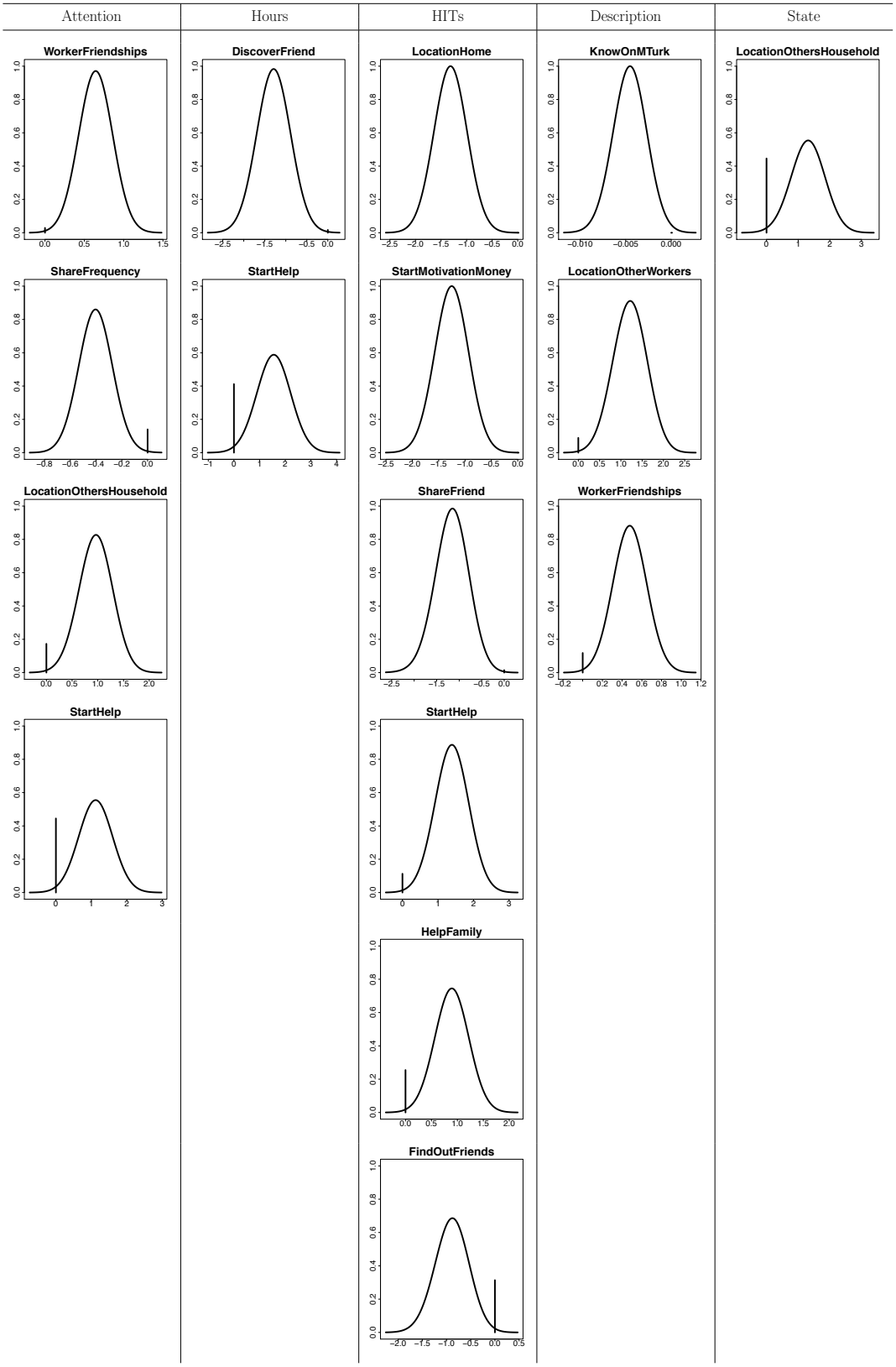
Dichotomous variables except: * 5 point Likert scale with 5-strongly agree. # Integer. % Percent.

Table SI.3.2 shows posterior distributions resulting from BMA for independent variables with posterior inclusion probabilities exceeding 50%. Only non-demographic variables are shown.

Figure SI.3.1 shows each dependent variable with colors indicating the sign of the posterior mean for each BMA model.

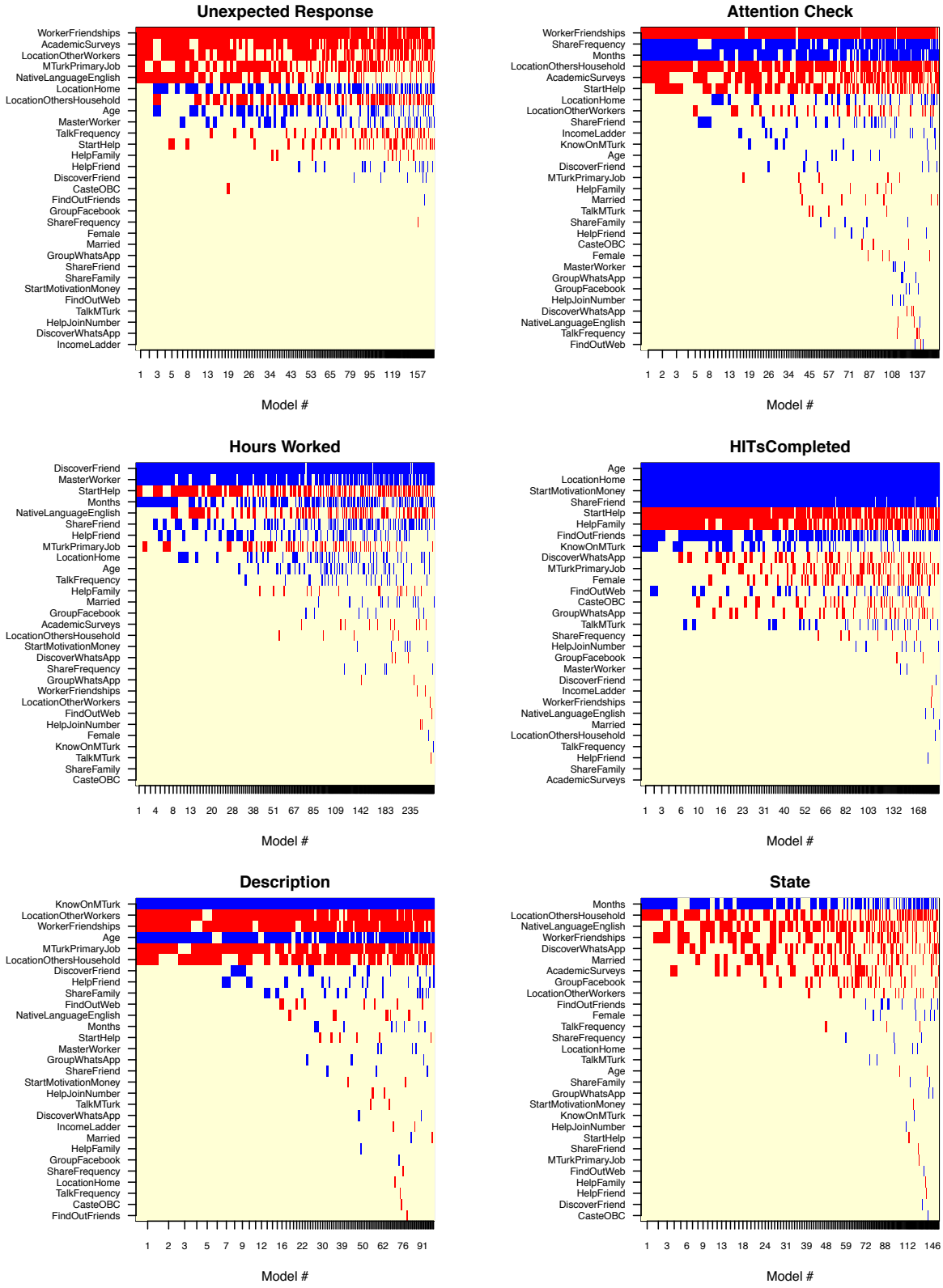
Tables SI.3.3, SI.3.4, SI.3.5, SI.3.6, SI.3.7, and SI.3.8 display Bayesian Model Averaging resuts sorted by posterior inclusion probability (PIP) or the overall importance of a particular independent variable in predicting a given type of anomalous response.

Table SI.3.2: BMA Posterior Distributions



Posterior distributions from BMA models with posterior inclusion probability exceeding 50%.

Figure SI.3.1: BMA Model Plots



Red indicates positive estimate, blue indicates negative estimate. Results from BMA.

Table SI.3.3: BMA Results for Any Anomalous Response

	PIP	Post Mean	Post SD
WorkerFriendships	95.60	0.56	0.22
AcademicSurveys	79.40	0.02	0.01
LocationOtherWorkers	61.30	0.86	0.78
MTurkPrimaryJob	55.20	0.92	0.95
NativeLanguageEnglish	50.80	0.73	0.83
LocationHome	49.60	-0.58	0.67
LocationOthersHousehold	49.50	0.57	0.65
Age	32.40	-0.03	0.05
MasterWorker	19.60	-0.21	0.46
TalkFrequency	19.20	0.08	0.18
StartHelp	17.40	0.17	0.42
HelpFamily	5.10	0.04	0.21
HelpFriend	4.80	-0.04	0.23
DiscoverFriend	1.80	-0.01	0.12
CasteOBC	1.20	0.01	0.09
FindOutFriends	0.30	-0.00	0.04
GroupFacebook	0.20	0.00	0.04
ShareFrequency	0.20	0.00	0.01
Female	0.00	0.00	0.00
Married	0.00	0.00	0.00
GroupWhatsApp	0.00	0.00	0.00
ShareFriend	0.00	0.00	0.00
ShareFamily	0.00	0.00	0.00
StartMotivationMoney	0.00	0.00	0.00
FindOutWeb	0.00	0.00	0.00
TalkMTurk	0.00	0.00	0.00
HelpJoinNumber	0.00	0.00	0.00
DiscoverWhatsApp	0.00	0.00	0.00
IncomeLadder	0.00	0.00	0.00

Table SI.3.4: BMA Results for Attention

	PIP	Post Mean	Post SD
WorkerFriendships	97.10	0.63	0.24
ShareFrequency	86.00	-0.35	0.19
Months	85.30	-0.01	0.01
LocationOthersHousehold	82.70	0.79	0.47
AcademicSurveys	58.20	0.01	0.01
StartHelp	55.50	0.62	0.66
LocationHome	19.10	-0.12	0.29
LocationOtherWorkers	17.80	0.18	0.43
ShareFriend	12.90	-0.10	0.28
IncomeLadder	5.60	-0.01	0.03
KnowOnMTurk	3.70	-0.00	0.00
Age	3.00	-0.00	0.01
DiscoverFriend	3.00	-0.01	0.09
MTurkPrimaryJob	2.80	0.03	0.23
HelpFamily	2.60	0.01	0.08
Married	2.30	0.01	0.08
TalkMTurk	2.00	0.00	0.00
ShareFamily	1.70	-0.01	0.07
HelpFriend	1.60	-0.01	0.06
CasteOBC	1.50	0.00	0.05
Female	1.30	0.00	0.05
MasterWorker	0.90	-0.00	0.04
GroupWhatsApp	0.90	-0.00	0.03
GroupFacebook	0.90	-0.00	0.03
HelpJoinNumber	0.90	-0.00	0.00
DiscoverWhatsApp	0.90	0.00	0.03
NativeLanguageEnglish	0.90	0.00	0.03
TalkFrequency	0.80	0.00	0.01
FindOutWeb	0.80	-0.00	0.03

Table SI.3.5: BMA Results for Hours

	PIP	Post Mean	Post SD
DiscoverFriend	98.30	-1.27	0.43
MasterWorker	81.10	-1.19	0.76
StartHelp	58.80	0.91	0.92
Months	57.50	-0.01	0.01
NativeLanguageEnglish	35.30	0.28	0.43
ShareFriend	34.60	-0.28	0.43
HelpFriend	27.00	-0.22	0.41
MTurkPrimaryJob	24.00	0.48	1.01
LocationHome	20.40	-0.16	0.36
Age	13.40	-0.01	0.03
TalkFrequency	6.10	-0.02	0.07
HelpFamily	4.90	0.03	0.15
Married	2.90	-0.02	0.11
GroupFacebook	2.80	-0.02	0.12
AcademicSurveys	2.50	0.00	0.00
LocationOthersHousehold	2.50	0.01	0.10
StartMotivationMoney	1.80	-0.01	0.08
DiscoverWhatsApp	1.40	0.01	0.08
ShareFrequency	1.20	-0.00	0.03
GroupWhatsApp	0.40	0.00	0.04
WorkerFriendships	0.40	0.00	0.02
LocationOtherWorkers	0.30	0.00	0.03
FindOutWeb	0.30	0.00	0.03
HelpJoinNumber	0.20	0.00	0.00
Female	0.10	-0.00	0.01
KnowOnMTurk	0.10	-0.00	0.00
TalkMTurk	0.10	0.00	0.00
ShareFamily	0.00	0.00	0.00
CasteOBC	0.00	0.00	0.00

Table SI.3.6: BMA Results for HITs

	PIP	Post Mean	Post SD
Age	100.00	-0.16	0.04
LocationHome	100.00	-1.31	0.33
StartMotivationMoney	100.00	-1.25	0.32
ShareFriend	98.50	-1.15	0.40
StartHelp	88.70	1.23	0.63
HelpFamily	74.50	0.66	0.48
FindOutFriends	68.60	-0.61	0.51
KnowOnMTurk	28.30	-0.00	0.00
DiscoverWhatsApp	27.30	0.23	0.43
MTurkPrimaryJob	22.90	0.46	1.00
Female	20.50	0.15	0.34
FindOutWeb	17.70	-0.17	0.41
CasteOBC	15.50	0.10	0.27
GroupWhatsApp	14.40	0.11	0.30
TalkMTurk	12.90	-0.00	0.00
ShareFrequency	2.70	0.01	0.05
HelpJoinNumber	2.20	-0.00	0.00
GroupFacebook	0.70	0.00	0.04
MasterWorker	0.50	-0.00	0.04
DiscoverFriend	0.40	-0.00	0.03
IncomeLadder	0.40	0.00	0.01
WorkerFriendships	0.30	0.00	0.01
NativeLanguageEnglish	0.30	-0.00	0.02
Married	0.20	-0.00	0.01
LocationOthersHousehold	0.20	-0.00	0.02
TalkFrequency	0.20	0.00	0.01
HelpFriend	0.20	-0.00	0.02
ShareFamily	0.00	0.00	0.00
AcademicSurveys	0.00	0.00	0.00

Table SI.3.7: BMA Results for Description

	PIP	Post Mean	Post SD
KnowOnMTurk	100.00	-0.00	0.00
LocationOtherWorkers	91.10	1.10	0.51
WorkerFriendships	88.20	0.42	0.22
Age	85.10	-0.07	0.04
MTurkPrimaryJob	84.50	1.34	0.80
LocationOthersHousehold	66.10	0.58	0.50
DiscoverFriend	11.90	-0.07	0.21
HelpFriend	11.80	-0.08	0.24
ShareFamily	9.70	-0.06	0.20
FindOutWeb	5.50	0.03	0.16
NativeLanguageEnglish	5.10	0.03	0.14
Months	4.60	-0.00	0.00
StartHelp	3.80	0.02	0.13
MasterWorker	2.00	-0.01	0.07
GroupWhatsApp	1.90	-0.01	0.07
ShareFriend	1.70	-0.00	0.05
StartMotivationMoney	1.10	0.00	0.04
HelpJoinNumber	1.10	0.00	0.00
TalkMTurk	1.00	0.00	0.00
DiscoverWhatsApp	1.00	-0.00	0.04
IncomeLadder	0.90	0.00	0.01
Married	0.80	0.00	0.03
HelpFamily	0.60	-0.00	0.03
GroupFacebook	0.50	-0.00	0.02
ShareFrequency	0.50	0.00	0.01
LocationHome	0.50	0.00	0.02
TalkFrequency	0.50	0.00	0.01
CasteOBC	0.50	0.00	0.02
FindOutFriends	0.40	0.00	0.02

Table SI.3.8: BMA Results for State

	PIP	Post Mean	Post SD
Months	67.90	-0.01	0.01
LocationOthersHousehold	55.40	0.72	0.76
NativeLanguageEnglish	49.60	0.49	0.56
WorkerFriendships	41.80	0.31	0.42
DiscoverWhatsApp	31.90	0.29	0.48
Married	24.10	0.21	0.43
AcademicSurveys	24.00	0.00	0.01
GroupFacebook	12.40	0.09	0.27
LocationOtherWorkers	5.20	0.06	0.28
FindOutFriends	3.30	-0.02	0.12
Female	1.70	-0.01	0.09
TalkFrequency	1.30	0.00	0.04
ShareFrequency	1.10	-0.00	0.03
LocationHome	0.80	-0.00	0.06
TalkMTurk	0.80	-0.00	0.00
Age	0.50	0.00	0.00
ShareFamily	0.50	-0.00	0.04
GroupWhatsApp	0.40	-0.00	0.04
StartMotivationMoney	0.30	0.00	0.02
KnowOnMTurk	0.30	-0.00	0.00
HelpJoinNumber	0.30	-0.00	0.00
StartHelp	0.30	0.00	0.04
ShareFriend	0.20	0.00	0.02
MTurkPrimaryJob	0.20	0.00	0.06
FindOutWeb	0.20	-0.00	0.03
HelpFamily	0.20	0.00	0.02
HelpFriend	0.20	0.00	0.02
DiscoverFriend	0.20	-0.00	0.02
CasteOBC	0.20	-0.00	0.02

Table SI.3.9 displays a robustness check where we employ logistic regression with the independent variables identified in the BMA and Table SI.3.10 displays a robustness check with all variables included. As expected, the significant predictors in the BMA remain significant in the logistic regression model.

Table SI.3.9: Significant Predictors of Anomalous Responses

	Dependent variable:					
	Anomalous	Attention Check	Hours Worked	HITsCompleted	Description	State
	(1)	(2)	(3)	(4)	(5)	(6)
AcademicSurveys	0.020***				0.015**	
	(0.006)				(0.006)	
Age		−0.081***		−0.168***		
		(0.031)		(0.040)		
LocationOtherWorkers	1.561***	0.975***				
	(0.389)	(0.356)				
LocationOthersHousehold		0.805**			0.968***	1.477***
		(0.321)			(0.308)	(0.503)
MTurkPrimaryJob	1.328**	1.430**				
	(0.564)	(0.598)				
ShareFriend				−1.027***		
				(0.340)		
LocationHome				−1.396***		
				(0.317)		
StartMotivationMoney				−1.346***		
				(0.314)		
FindOutFriends				−0.813**		
				(0.326)		
KnowOnMTurk		−0.005**		−0.004**		
		(0.002)		(0.002)		
WorkerFriendships	0.644***	0.431***			0.608***	
	(0.185)	(0.166)			(0.203)	
NativeLanguageEnglish	1.505***					
	(0.559)					
MasterWorker			−1.278**			
			(0.523)			
Months			−0.018**		−0.018***	−0.022**
			(0.008)		(0.005)	(0.009)
DiscoverFriend			−1.447***			
			(0.372)			
HelpFamily				0.971***		
				(0.322)		
StartHelp			1.390**	1.484***		
			(0.644)	(0.458)		
ShareFrequency					−0.390***	
					(0.127)	
Constant	−5.006***	−1.152	−1.478**	5.393***	−3.187***	−2.383***
	(0.881)	(1.416)	(0.681)	(1.297)	(0.946)	(0.525)
Observations	305	305	305	305	305	305
Note:						*p<0.1; **p<0.05; ***p<0.01

Logistic regression models where 1 indicates an anomalous response. Significant covariates resulting from Bayesian Model Averaging presented in rows.

Table SI.3.10: Models with All Predictors

	Dependent variable:					
	Anomalous	Attention Check	Hours Worked	HITsCompleted	Description	State
	(1)	(2)	(3)	(4)	(5)	(6)
AcademicSurveys	0.027***	0.015**	0.005	0.003	0.027***	0.016
	(0.009)	(0.007)	(0.008)	(0.008)	(0.007)	(0.011)
Age	0.012	−0.037	−0.036	−0.105**	−0.033	0.051
	(0.046)	(0.040)	(0.045)	(0.051)	(0.038)	(0.055)
LocationOtherWorkers	0.453	−0.209	−0.872	1.600***	1.046**	0.545
	(0.637)	(0.491)	(0.585)	(0.553)	(0.447)	(0.818)
LocationOthersHousehold	0.887	0.530	−0.176	−0.581	0.739**	0.900
	(0.549)	(0.378)	(0.457)	(0.437)	(0.369)	(0.651)
MTurkPrimaryJob	1.219*	0.017	1.086	1.210	0.826	−0.507
	(0.702)	(0.921)	(1.192)	(1.184)	(0.671)	(1.307)
ShareFriend	−0.451	−0.123	−0.712	−1.636***	−0.712	0.773
	(0.766)	(0.434)	(0.490)	(0.518)	(0.490)	(0.634)
LocationHome	−0.733	−0.393	−0.715	−0.847**	0.121	0.097
	(0.601)	(0.343)	(0.436)	(0.388)	(0.381)	(0.494)
StartMotivationMoney	−0.144	−1.169***	−0.169	−1.095***	0.538	0.695
	(0.533)	(0.321)	(0.397)	(0.356)	(0.359)	(0.443)
FindOutFriends	0.063	−0.688**	−1.139**	−0.871**	0.291	−0.590
	(0.498)	(0.330)	(0.443)	(0.367)	(0.343)	(0.481)
KnowOnMTurk	−0.006***	−0.0001	−0.0001	−0.003	−0.004**	−0.004
	(0.002)	(0.0003)	(0.001)	(0.002)	(0.002)	(0.004)
WorkerFriendships	0.539**	0.533**	0.060	−0.028	0.339*	0.731**
	(0.243)	(0.230)	(0.253)	(0.240)	(0.188)	(0.348)
NativeLanguageEnglish	2.485***	0.007	0.696*	−0.612	0.258	1.239***
	(0.831)	(0.342)	(0.405)	(0.393)	(0.374)	(0.464)
MasterWorker	−0.846	−0.516	−1.310**	0.185	−0.154	1.841***
	(0.586)	(0.415)	(0.576)	(0.462)	(0.405)	(0.568)
Months	−0.024***	−0.010*	−0.012	−0.035***	−0.007	−0.027**
	(0.007)	(0.006)	(0.009)	(0.009)	(0.006)	(0.011)
DiscoverFriend	−1.170**	−0.217	−1.256***	−0.346	−0.956***	−0.138
	(0.582)	(0.341)	(0.433)	(0.398)	(0.370)	(0.474)
HelpFamily	1.022*	0.458	0.586	1.089***	−0.092	−0.049
	(0.573)	(0.326)	(0.388)	(0.373)	(0.351)	(0.423)
StartHelp	0.884	1.129**	1.371**	0.428	0.644	0.491
	(0.588)	(0.528)	(0.697)	(0.557)	(0.443)	(0.763)
ShareFrequency	0.203	−0.371*	0.051	0.270	0.219	−0.407
	(0.349)	(0.192)	(0.217)	(0.221)	(0.218)	(0.268)
Constant	−3.704	−1.505	−1.061	2.856	−4.044**	−8.448***
	(2.260)	(1.861)	(2.207)	(2.206)	(1.786)	(2.806)
Observations	305	305	305	305	305	305

Note: *p<0.1; **p<0.05; ***p<0.01

Logistic regression models where 1 indicates an anomalous response.

SI.4: Understanding Indian MTurkers' Motivations

That some Indian MTurkers provide anomalous responses should not be a surprise, given that these workers operate in a highly competitive environment with low pay and incentives to complete work as quickly as possible (Boas, Christenson and Glick, 2020; Gray and Suri, 2019). Yet, not all workers fall into this category. Experienced workers know how to navigate the cutthroat MTurk system by forming relationships, sharing information only with key friends, and treating work on MTurk as a way earn some extra money while doing HITs of their choosing. These workers have incentives to access U.S.-specific surveys undetected.

Given these findings that different types of Indian MTurkers are more or less likely to provide anomalous responses, we were driven to learn more about how Indian MTurkers approached and interacted with MTurk and what their motivations were. Part of our method involved interviewing Indian MTurkers to get their perspectives on the platform and how it shapes their lives. While the full contents of our investigation is outside of the purview of this article, we want to highlight one key finding that informs our proposed strategy of opening MTurk surveys to Master Workers with no location requirements.

Work on Indian MTurk is scarce and particularly poorly paid, even in comparison to pay on U.S. MTurk and adjusting for differences in purchasing power. Indian MTurkers are motivated to complete as much work as possible and must do so in order to have any hope of supplementing their income through MTurk. As such, they are strongly motivated to accept any work paying a fair wage on Indian MTurk. Fair paying work is viewed very favorably, and Indian MTurkers (particularly Master Workers) take these work assignments very seriously, producing excellent results. When fair paying work is not available on Indian MTurk, Indian MTurkers are incentivized to access U.S. MTurk because workers are acutely aware that most HITs on Indian MTurk are underpaid. Additionally, the disparity between U.S. and Indian MTurk is often viewed from a colonial lens: by accessing U.S. MTurk work, workers are engaging in a form of work-based resistance. This mindset carries over to how Indian MTurkers complete work on U.S. MTurk. Workers seek to provide non-anomalous responses so that they can be paid, but worker effort in such cases is rightfully low.

Taken together, then, we suggest that opening MTurk surveys to all Master Workers provides the optimal incentive structure to encourage Indian MTurkers to self-identify and improves researcher confidence that U.S. respondents can be identified.

SI.5: Addressing Anomalous Responses

We test our expectation that removing location requirements and requiring Master Workers will result in Indian MTurkers self-identifying as Indian by fielding a survey with many of the same questions as our first survey, but adding several questions that U.S. and Indian respondents should answer differently. We pilot tested these questions on MTurk on November 9, 2020 with U.S.-based respondents ($N = 51$) and found that almost all respondents answered the questions as expected.[2] The question wording of these new questions is shown below.

- Date: What is today's date?
  - DateUS=1 if the date was written 11/18/2020, November 18, 2020, or something similar.
  - DateUS=0 if the date was written 18/11/2020, 18.11.2020, 11.18.2020, or something similar.
- ElectedOffices: What would you say are the three most important elected offices?
  - ElectedOfficesUS=1 if the respondent listed political positions in the U.S. Those listing fewer than three positions, but positions that are in the U.S. were okay, as were respondents listing unelected positions in the U.S. Examples: President, Vice President, Senator, Representative, Governor, Speaker of the House, Attorney General, Mayor, Judge, Sheriff, Prime Minister of the UK (this was acceptable because "of the UK" was specified, indicating that the respondent knew that this was a foreign leader).
  - ElectedOfficesUS=0 if the respondent refused, copied a response from the Internet, or provided a response that made sense only in a foreign context. Examples: CM, MP, MLA, Bank, Post Office, Railways, Election Commission, Audit General, Local Council Members, General of the Army.
- Height: About how tall are you?
  - HeightUS=1 if height was written 5"5', 5'5", 5 feet 5 inches, 61 inches, or something similar.
  - HeightUS=0 if height was written 5.11, 150, 1.70 m, 165CM, 5.4", or something similar.
- Country: This survey is open to respondents from around the world. In what country do you live?
- State: Within that country, in what state do you live?

[2]Of course, it is impossible to fully rule out that an Indian respondent could have circumvented the U.S.-based location requirement and taken the survey. However, we included all exclusion criteria offered by Cloud Research to obtain only U.S. respondents, and the responses we received aligned with our expectations about how U.S.-based respondents would answer these questions.

- LargestCity: What is the city in your state with the highest population?
  - LargestCityCorrect=1 if the respondent listed the largest city, by population in the state that they reported living in. LargestCityUS=1 if said city was the largest city in a U.S. state.
  - LargestCityClose=1 i the respondent listed one of the top five largest cities, by population in the state that they reported living in. LargestCityCloseUS=1 if said city was one of the top five largest cities in a U.S. state.
- LargestCitySize: About how many people live in the city in your state with the highest population?
  - LargestCitySizeUS=1 if the respondent listed the population in hundreds of thousands or millions.
  - LargestCitySizeUS=0 if the respondent refused to guess or listed the population in Lakh or Crore.
- TimingUSEnglish: Using the existing question asking respondents to describe their typical day working on MTurk, we coded this variable as 1 if the respondent used American English and 0 otherwise. Examples of non-American English phrases: "is some," "leave for market," "if there are some work," "take lunch and some rest," "new works."

The questions asking about anomalous responses from our first survey remained and were coded in the following way:

- HoursWorkedAnomalous: 1 if Hours worked on MTurk were more than the total number of hours worked.
- AttentionCheckAnomalous: 1 if the respondent did not select "neither agree nor disagree" on the attention check question.
- HITsCompletedAnomalous: 1 if the respondent reported completing fewer than 1000 HITs.
- TimingAnomalous: 1 if the respondent refused to answer the question, copied text from a website, or otherwise did not describe their day in a meaningful way.

We then created a ten point scale representing the likelihood that a respondent was in the U.S. (called *USScore*) based on these survey questions by summing one minus each dummy variable representing an anomalous response plus each of the variables representing whether the respondent answered a given question like a U.S. respondent would.

We conducted this survey ($N = 302$) on November 18 and 19, 2020. The survey was open to any Master Worker who had completed more than 1000 HITs with an approval rate above 98%. The survey launched at 12:30PM Indian Standard Time (2:00AM Eastern Time) in order to maximize the number of Indian respondents.

In addition to analyzing the survey responses, we also conducted checks on respondents' IP addresses to determine their location and whether they were using a VPS. Our checks used Proxycheck.io, IP Hub, IP Intel, and Qualtrics' IP locator. We used IP Hub to identify countries, Proxycheck to identify states, and IP Intel to identify VPS'.

Our survey had respondents (based on their IP address) from Brazil (1), Canada (1), India (82), Libya (1), Mexico (1), Oman (1), Singapore (1), Sri Lanka (1), Sweden (1), Thailand (1), and the United States (210). Of these respondents, ten Indian respondents obfuscated their location: 5 claimed to be in the U.S., 1 claimed to be in Singapore, and 4 did not provide a response (listing: "200," "united state," "united status," and "YES").

We found that four respondents used a Virtual Private Server. One person living in India used a VPS in Sweden (for a total of 83 Indian respondents), and the other three respondents were living in the U.S. and used a VPS to change their location to another location in the U.S.

Taken together, there were 5 respondents who claimed to be in the U.S. Keep in mind that all of these respondents were detected using their IP address, so we know that they are Indian respondents just by this fact. Therefore, these respondents will not impact the quality of the U.S. data because they did not use a VPS to mask their true location. Nevertheless, we might be interested in whether these respondents' *USScore* was lower than for respondents actually in the U.S. Two of these respondents scored a 4, one a 5, one a 7, and one a 9. None of these respondents provided responses to the question about their typical day on MTurk that indicated that they were fluent in American English. Therefore, even if these respondents used a VPS to access U.S.-based surveys, their lack of English fluency would have alerted researchers.

Table SI.5.1 shows the individual dummy variables that make-up *USScore*. The clear pattern is that Indian respondents are substantially more likely to provide an anomalous response to a question whose answer relies on knowledge about the U.S. or fluency in American English. Indian respondents also scored lower on measures of anomalous responses, but these items alone are not enough to filter out Indian respondents.

Table SI.5.1: Components of *USScore*

	India		United States	
	Expected	Anomalous	Expected	Anomalous
HoursWorked	78	5	202	8
AttentionCheck	78	5	208	2
HITsCompleted	78	5	209	1
Timing	56	27	208	2
TimingUSEnglish	11	72	206	4
ElectedOfficesUS	12	71	202	8
LargestCityUS	2	81	198	12
DateUS	55	28	210	0
HeightUS	30	53	208	2
LargestCitySizeUS	58	25	200	10

The following is the distribution of *USScore*:

- United States: 5 (1), 7 (1), 8 (4), 9 (33), 10 (171).
- India: 3 (8), 4 (11), 5 (25), 6 (19), 7 (12), 8 (5), 9 (3).
- Others: 5 (Sri Lanka), 6 (Singapore), 7 (Libya, Mexico, Oman, Thailand), 8 (Brazil, Canada, Sweden). The person in Sweden used a VPS and was located in India.

We now turn our attention to respondents whose IP addresses were located in the United States and check to determine whether we think that they may actually be Indian respondents who are obfuscating their location. If this is occurring, existing methods have no way of filtering out these respondents especially because the three respondents in the U.S. using a VPS all scored a perfect 10 on *USScore*, so it is extremely unlikely that they are foreign respondents.

The overwhelming majority of respondents whose IP addresses located them in the U.S. scored a perfect 10 on *USScore*. Thirty-three respondents scored a 9. The one item that each person missed is listed below. From the list, it is fairly evident that these respondents either misunderstood the question, made a mistake, or just responded in a non-standard manner.

- HoursWorked: 5 respondents said they worked more hours on MTurk than they did overall each week. Three of these respondents claimed to work zero hours overall, so they may have misread the question.
- HITsCompleted: 1 respondent said that they completed 80 HITs.
- TimingUSEnglish: 1 respondent had non-standard English skills, saying "When you are looking at the computer do a few turks to have money sent to your Amazon gift card. I also buy cool stuff with the little extra cash."
- ElectedOfficesUS: 6 respondents listed incorrect answers to this question. Four said "none" or "not sure;" one said "President, governor, dog catcher;" and one said "wealth gap, social justice, immigration."

- LargestCityCloseUS: 10 respondents listed the incorrect largest city (or top five largest city) in their state. Nine listed what appeared to be their city of residence. One was in California, but listed New York.

- HeightUS: 2 respondents did not list their height in a standard way. One said "6-2" and 1 said "5,12."

- LargestCitySizeUS: 8 respondents did not correctly list the size of the largest city in their state. All eight said that they were unsure.

There were 6 respondents who scored an 8 or less. The details on these respondents are below

- HoursWorked: 3 respondents said they worked more hours on MTurk than they did overall each week.

- AttentionCheck: 2 failed it.

- Timing: 2 did not describe their day working on MTurk. One person refused to answer and one person said "various task i do on mturk."

- TimingUSEnglish: 3 wrote in non-standard English.

- ElectedOfficesUS: 2 did not list normal elected offices. One said "president, vp, finance ministry," and one said "I don't know."

- LargestCityCloseUS: 2 did not list the largest city in their state. One entered a space and one said the town that they likely lived in.

- LargestCitySizeUS: 2 did not list the largest city size in their state. One said "not sure," and one said "I don't know."

Of these respondents, none used a VPS, none were flagged by IP Intel as having a bad IP address, and none were marked as "risky" by Proxycheck. Two respondents reported living in a different state than the one they listed. One person's IP address was recorded as being in Georgia, while they reported being in Alabama. Another person's IP address reported that they were in Illinois, while they reported being in Michigan. We searched lists of suspicious and spam IP addresses for these IP addresses, but none of them were listed as suspicious on any sites. These respondents could be Indian workers who are using a VPS undetected, native U.S. workers who are responding carelessly, or immigrant U.S. workers who are not fluent in English. Even if all of these respondents were Indian workers, they make up less than 3% of our sample of U.S.-based respondents. Additionally, VPS users in the United States make up less than 1% of the sample. This is compared to the between 20 and 30% of potentially foreign respondents identified as using a VPS in prior work (Ahler, Roush and Sood, 2019; Dennis, Goodson and Pearson, 2020).

SI.6: Comparing Anomalous Response Solutions

In the main text, Table 2 describes several strategies requesters can use to increase confidence that there are no non-U.S. respondents in their survey. Of course, there are almost infinite combinations of HIT restrictions and analytic techniques that requesters could use, and we compare particularly popular and relevant choices in the Table.

Our cost per non-anomalous U.S. response is based on a number of estimates from our surveys. We base our estimates on a per respondent payment rate of $1.00, which is reasonable for a five to ten minute survey. First, we estimate that 30% of non-Master and 20% of Master Workers provide anomalous responses. The 20% figure is from our survey analysis; based on prior work, we know that non-Master Workers are more likely to provide anomalous responses. Second, we estimate that 15% of non-Master and 3% of Master Workers use a VPS. Ahler, Roush and Sood (2019) and Dennis, Goodson and Pearson (2020) identify between 20 and 30% of respondents who use a VPS when accessing U.S.-based surveys. These respondents are potentially foreign respondents. Costs could be reduced if IP address detection was conducted before allowing the respondent to complete the survey as in Kennedy et al. (2020). Some Institutional Review Boards have restrictions on the ability to filter out potential respondents in this way. Three percent of Master Workers used a VPS in our survey.

Finally, we assume that 80% of Master Workers are from the U.S. Recall that we set the start time of our survey to maximize the number of foreign (particularly Indian) respondents. Hence, we base our estimate of the percentage of respondents from the U.S. on the percentage of U.S.-based respondents in our survey beginning at the more reasonable time of 6AM Eastern Time.

We focus on comparing the last three strategies in Table 2. Requesters may want to use non-Master Workers if they can conduct IP address detection before respondents take the survey, therefore avoiding having to pay for and subsequently throw out data from respondents who use a VPS. Keeping a survey available to all workers also increases the pool of potential respondents, reducing survey completion time and allowing more casual MTurk users who are less likely to be professional survey takers to take the survey. At the same time, any non-Master Worker sample is more likely to be accessed by foreign workers who circumvent location requirements because foreign workers are primarily non-Master Workers. Requesters can also opt to add in other screening questions like those we identify as most likely for foreign workers to miss. Adding additional questions increases survey length and the percentage of responses identified as anomalous, but can help improve data quality.

A viable alternative is to restrict the HIT to U.S. respondents who are also Master Workers. The increased cost of paying for Master Workers is outweighed by fewer Master Workers using a VPS and providing anomalous responses. Since there are fewer foreign Master Workers, the number of foreign workers who could potentially access surveys restricted to U.S.-based Master Workers is lower. However, Master Workers are also more skilled at using MTurk, so they may have the technical ability to circumvent location requirements that foreign, non-Master Workers lack. There is really no way to be certain that no foreign workers circumvented location requirements, especially because Master Workers are likely to pass all anomalous response checks.

Removing the location requirement increases requester confidence that foreign workers correctly self-report their location and do not attempt to circumvent location restrictions. Getting caught circumventing location restrictions is potentially costly for Master Workers, so they are incentivized to complete as much work as possible without circumventing location restrictions.[3] For example, Indian workers will check Indian MTurk for HITs before accessing U.S. MTurk. Additionally, each HIT has a qualifications section that lists whether there is a location requirement. If an Indian worker is on U.S. MTurk and sees a HIT with no location requirement, the worker will know that they can access and complete that HIT on Indian MTurk. In addition to foreign workers being able to complete a HIT without circumventing location requirements, this approach also provides requesters with a comparison group of self-identified foreign workers whose responses they can compare to workers self-reporting being in the U.S. The comparison group means that requesters can better adjudicate whether a self-reported U.S. respondent who provides anomalous responses answers survey questions in a similar way to a foreign respondent or if the respondent is simply an inattentive U.S.-based respondent.

One consequence of removing location requirements is that any eligible MTurk worker from around the world can complete the HIT. Exactly what this would mean for Institutional Review Board approval of an MTurk study is likely to vary widely depending on the requester's institution. We forsee one of two options occurring based on our experience with Institutional Review Board approval. First, many Institutional Review Boards have special policies and procedures for studies conducted using MTurk. These policies often mean that MTurk surveys without location restrictions do not require any special review. Alternatively, some Institutional Review Boards may require that MTurk studies undergo special review based on the anticipated country make-up of the respondent pool. For MTurk studies, this means being reviewed for cultural applicability in India, as workers not from the U.S. or India are rare. This additional review procedure is not particularly difficult or time consuming, especially for opt-in online surveys.

Table SI.6.1 displays the same strategies discussed in the main text with a new column describing the total number of responses required in order to achieve a desired number of non-anomalous U.S. responses. For example, if the desired $N = 300$, then requesters would only need 300 respondents if they used no HIT restrictions or analytic techniques, but they would need $300 + (300 \times 0.43) = 429$ respondents to obtain 300 non-anomalous U.S. responses if they implemented a U.S. location requirement and anomalous data response checks.

[3]Workers identified circumventing location requirements are usually permanently banned from MTurk without warning.

Table SI.6.1: Strategies to Reduce Anomalous Responses

HIT Restrictions	Analytic Techniques	Cost pp.	N Increase (%)	Confidence U.S.-Based Respondents Can Be Identified
None	None.	\$1.20	0	None.
U.S.	None.	\$1.20	0	Low. Circumventing location requirements is easy.
U.S.	Anomalous data response checks.	\$1.71	43	Medium Low. Circumventing location restrictions is easy, and only some workers will fail anomalous response checks.
U.S.	IP address detection.	\$1.41	18	Medium. Circumventing location restrictions using a non-traceable method is non-trivial.
U.S.	Anomalous data response checks and IP address detection.	\$2.18	82	Medium High. Circumventing location restrictions using a non-traceable method is non-trivial. Only some workers will fail anomalous response checks.
U.S., Master	Anomalous data response checks and IP address detection.	\$1.62	30	High. Master Workers are much more reliable. Circumventing location restrictions using a non-traceable method is non-trivial. Few Master Workers will fail anomalous response checks.
Master	Anomalous data response checks and IP address detection.	\$2.19	75	Very High. Master Workers are much more reliable. Few Master Workers will fail anomalous response checks. No location restrictions to circumvent.

HIT Restrictions are Amazon features to restrict survey eligibility. We assume all HITs are restricted to an approval percentage above 98%. Analytic techniques are requester strategies to filter out anomalous responses. We assume that IP address detection occurs after the survey is complete. Cost pp. is the estimated cost per non-anomalous, U.S. response. Cost estimate is based on \$1.00 payment, 30% of non-Master and 20% of Master Workers providing anomalous responses, 3% of workers using a VPS, and 80% of Master Workers being from the U.S. 20% Amazon Fee plus 5% Master Worker fee. N Increase is percentage of additional respondents needed to obtain the desired number of non-anomalous, U.S. responses.

Also included in the Table are additional details about the confidence researchers should have that all U.S.-based survey respondents can be successfully identified. As mentioned in the main text, the first four strategies are widely recognized as insufficient for ensuring data quality. We recognize that these strategies are generally cheaper than the last three strategies listed, but after ongoing issues with MTurk data quality, the onus is on the researcher to show that her MTurk collected data is of high quality.

Indeed, the ability to validate anomalous U.S. survey responses with those from self-identified foreign workers is one of the biggest strengths of fielding surveys among Master Workers without location restrictions. In this way, responses from self-identified foreign workers are not thrown out; they are analyzed and compared to anomalous responses from self-identified U.S. workers. After making this comparison, researchers can be much more confident that the anomalous responses among self-identified U.S. workers are not the result of foreign workers circumventing IP address detection and completing U.S.-based work undetected. In essence, researchers can audit the quality of their own data using self-identified foreign responses. We feel that this provides researchers with a major advantage because it helps them defend their work against criticisms about the quality of MTurk data.

In cases where data quality is less of a concern or cost is of paramount importance, other methods may be more appropriate. Using U.S.-based Master Workers can be a good com-

promise between cost and data quality if requesters are comfortable with more uncertainty about whether foreign workers are completing U.S.-based HITs undetected.

References

Abbey, James D. and Margaret G. Meloy. 2017. "Attention by Design: Using Attention Checks to Detect Inattentive Respondents and Improve Data Quality." *Journal of Operations Management* 53-56(1):63–70.

Ahler, Douglas J, Carolyn E Roush and Gaurav Sood. 2019. Micro-Task Market for Lemons. In *Meeting of the Midwest Political Science Association*. Chicago, Illinois: .

Anson, Ian G. 2018. "Taking the Time? Explaining Effortful Participation among Low-Cost Online Survey Participants." *Research & Politics* 5(3).

Antin, Judd and Aaron Shaw. 2012. Social Desirability Bias and Self-Reports of Motivation: A Study of Amazon Mechanical Turk in the US and India. In *SIGCHI Conference on Human Factors in Computing Systems*. Austin, TX: ACM Press pp. 2925–2934.

Aronow, Peter M., Jonathon Baron and Lauren Pinson. 2019. "A Note on Dropping Experimental Subjects Who Fail a Manipulation Check." *Political Analysis* 27(4):572–589.

Barends, Ard J. and Reinout E. de Vries. 2019. "Noncompliant Responding: Comparing Exclusion Criteria in MTurk Personality Research to Improve Data Quality." *Personality and Individual Differences* 143:84–89.

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.

Boas, Taylor C., Dino P. Christenson and David M. Glick. 2020. "Recruiting Large Online Samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics." *Political Science Research and Methods* 8(2):232–250.

Chandler, Jesse J. and Gabriele Paolacci. 2017. "Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors." *Social Psychological and Personality Science* 8(5):500–508.

Dennis, Sean A., Brian M. Goodson and Christopher A. Pearson. 2020. "Online Worker Fraud and Evolving Threats to the Integrity of MTurk Data: A Discussion of Virtual Private Servers and the Limitations of IP-Based Screening Procedures." *Behavioral Research in Accounting* 32(1):119–134.

Feitosa, Jennifer, Dana L. Joseph and Daniel A. Newman. 2015. "Crowdsourcing and Personality Measurement Equivalence: A Warning about Countries Whose Primary Language Is Not English." *Personality and Individual Differences* 75:47–52.

Gray, Mary L. and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt.

Hauser, David, Gabriele Paolacci and Jesse J. Chandler. 2019. Common Concerns with MTurk as a Participant Pool: Evidence and Solutions. In *Handbook in Research Methods in Consumer Psychology*. New York: Routledge pp. 319–337.

Hydock, Chris. 2018. "Assessing and Overcoming Participant Dishonesty in Online Data Collection." *Behavior Research Methods* 50(4):1563–1567.

Kaufmann, Nicolas, Thimo Schulze and Daniel Veit. 2011. More than Fun and Money. Worker Motivation in Crowdsourcing – A Study on Mechanical Turk. In *17th Annual Conference on Information Systems*. Detroit, Michigan: .

Kazai, Gabriella, Jaap Kamps and Natasa Milic-Frayling. 2012. The Face of Quality in Crowdsourcing Relevance Labels: Demographics, Personality and Labeling Accuracy. In *21st ACM International Conference on Information and Knowledge Management*. Maui, Hawaii: pp. 2583–2586.

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Ryan Jewell and Philip Waggoner. 2020. "The Shape of and Solutions to the MTurk Quality Crisis." *Political Science Research and Methods* OnlineFirst.

Lawson, Chappell, Gabriel S. Lenz, Andy Baker and Michael Myers. 2010. "Looking Like a Winner: Candidate Appearance and Electoral Success in New Democracies." *World Politics* 62(4):561–593.

Loepp, Eric and Jarrod T. Kelly. 2020. "Distinction without a Difference? An Assessment of MTurk Worker Types." *Research & Politics* 7(1):1–8.

MacInnis, Cara C., Harrison C.D. Boss and Joshua S. Bourdage. 2020. "More Evidence of Participant Misrepresentation on Mturk and Investigating Who Misrepresents." *Personality and Individual Differences* 152.

Matherly, Ted. 2019. "A Panel for Lemons? Positivity Bias, Reputation Systems and Data Quality on MTurk." *European Journal of Marketing* 53(2):195–223.

Milland, Kristy. 2017. "The Unsupported Crowd: Exclusion of Indian Workers in Amazon Mechanical Turk Communities.".

Miura, Asako and Tetsuro Kobayashi. 2016. "Survey Satisficing Inflates Stereotypical Responses in Online Experiment: The Case of Immigration Study." *Frontiers in Psychology* 7.

Montgomery, Jacob M. and Brendan Nyhan. 2010. "Bayesian Model Averaging: Theoretical Developments and Practical Applications." *Political Analysis* 18(2):245–270.

Necka, Elizabeth A., Stephanie Cacioppo, Greg J. Norman and John T. Cacioppo. 2016. "Measuring the Prevalence of Problematic Respondent Behaviors among MTurk, Campus, and Community Participants." *PLOS ONE* 11(6).

Sharpe Wessling, Kathryn, Joel Huber and Oded Netzer. 2017. "MTurk Character Misrepresentation: Assessment and Solutions." *Journal of Consumer Research* 44(1):211–230.

Spears, Dean. 2013. "Poverty and Probability: Aspiration and Aversion to Compound Lotteries in El Salvador and India." *Experimental Economics* 16(3):263–284.

Yudkin, Daniel A., Tobias Rothmund, Mathias Twardawski, Natasha Thalla and Jay J. Van Bavel. 2016. "Reflexive Intergroup Bias in Third-Party Punishment." *Journal of Experimental Psychology: General* 145(11):1448–1459.