



Published in final edited form as:

*J Econ Sci Assoc.* 2017 July ; 3(1): 1–11. doi:10.1007/s40881-017-0035-0.

## Turking Overtime: How Participant Characteristics and Behavior Vary Over Time and Day on Amazon Mechanical Turk

Antonio A. Arechar<sup>1</sup>, Gordon Kraft-Todd<sup>1</sup>, and David G. Rand<sup>1,2,3</sup>

<sup>1</sup>Department of Psychology, Yale University, 2 Hillhouse Ave. New Haven, CT 06511

<sup>2</sup>Department of Economics, Yale University, 2 Hillhouse Ave. New Haven, CT 06511

<sup>3</sup>School of Management, Yale University, 2 Hillhouse Ave. New Haven, CT 06511

### Abstract

Online experiments allow researchers to collect datasets at times not typical of laboratory studies. We recruit 2,336 participants from Amazon Mechanical Turk to examine if participant characteristics and behaviors differ depending on whether the experiment is conducted during the day versus night, and on weekdays versus weekends. Participants make incentivized decisions involving prosociality, punishment, and discounting, and complete a demographic and personality survey. We find no time or day differences in behavior, but do find that participants at nights and on weekends are less experienced with online studies; on weekends are less reflective; and at night are less conscientious and more neurotic. These results are largely robust to finer grained measures of time and day. We also find that those who participated earlier in the course of the study are more experienced, reflective, and agreeable, but less charitable than later participants.

### Keywords

cooperation; honesty; decision-making; time of day; mturk; self-control

### JEL Classification

C80; C90

## 1 Introduction

Online survey platforms are an increasingly popular tool for studying human behavior in the social sciences. Since the appearance of Amazon Mechanical Turk (MTurk), a plethora of studies have validated their use by successfully replicating classic findings from economics and psychology (Paolacci et al. 2010; Horton et al. 2011; Amir et al. 2012; Berinsky et al. 2012; Rand 2012; Arechar et al. 2016). In comparison to other methods, online surveys permit quick and affordable collection of large volumes of data.

Another feature of these online studies is that they make it easy to collect data at any time and, unlike studies conducted in the laboratory or in other face-to-face environments, participation can easily occur late at night or on weekends. This is possible because researchers commonly leave a single study continuously open for a week or longer, allowing participation at whichever time suits participants.

A potential issue arising from this practice, however, is heterogeneity in participants' characteristics based on time of participation. There is evidence in support of such heterogeneity; for example, people who work in traditional white collar jobs may be unavailable to complete studies during regular business hours. As a result, studies run during those hours may be more likely to recruit "professional" participants who use MTurk as a primary source of income – and thus may have more prior experience (Casey et al. 2016), make fewer errors (Chandler et al. 2015), and complete studies more quickly (Deetlefs et al. 2015). Additionally, participants recruited when a study is first posted may differ from those recruited later, as in college samples where there is evidence that students differ depending on whether they sign up to complete studies at the beginning versus the end of the semester (Aviv et al. 2002). Indeed, in an unincentivized survey study, Casey et al. (2016) explore the demographic and personality differences of participants who took part in surveys at different times on MTurk. Notably, they find that experienced participants were more likely to complete tasks earlier in the day, and that participants tend to be older, less neurotic and more conscientious earlier in the data collection.

Still, little is known about how participants' *behavior* may vary based on time of participation, and this is crucial knowledge for accurately interpreting the results of online studies. To shed light on this issue, we ran an incentivized study at regular intervals over two weeks to explore how participation at day versus night, and on weekdays versus the weekend, affects incentivized behavior in common economic paradigms, as well as the demographics and personality of those who self-select to participate.

Participants took part in a series of tasks presented in randomized order. They made seven incentivized decisions: a dictator game, a one-shot prisoner's dilemma game, and a third-party punishment game with prosocial punishment of selfishness and antisocial punishment of fairness, as well as an honesty task, a charitable giving decision, and a time discounting task. In addition to these incentivized measures, they also completed unincentivized measures of reflectiveness (a modified version of the cognitive reflection test, CRT; Frederick (2005)), the Big-5 personality traits of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (Gosling et al. 2003), and basic demographics.

We do not find significant differences in decisions in any of the incentivized behavioral measures. However, we do find that people participating at night are less experienced, take more time to complete tasks, are less conscientious, and more neurotic than their daytime fellows; and that people participating on weekends are less experienced and reflective. We also examine behavioral and demographic differences based on participation order. We find no differences in any of the incentivized measures, with the exception of charitable giving, where people participating earlier on in the study give less. We also find that such participants are more experienced, reflective, and agreeable than later ones. Of course, our

results cannot speak to causality. A person's characteristics could be influencing when they select into participation in studies on MTurk, or there could be a causal effect such that the same person tends to be, for example, less reflective on the weekend compared to weekdays. Although this distinction is important for understanding the psychological basis of our observations, the direction of causality does not have particular bearing on the practical implications for experimenters interested in running experiments on nights and weekends using MTurk.

In sum, our results suggest that incentivized economic behavior on MTurk is robust to the time of day and the day of the week, while there is some variation in participants' personality and prior experience across these recruitment times.

## 2 Experimental design and procedure

We recruited participants via MTurk, restricting their geographical location to the USA. A total of 2,336 American participants completed the study; average age was 34 years (range: 18–77), and 50% were female. Participants completed the task in an average of 15 minutes and they received a flat fee of \$1 for participating, plus an additional variable payment (average \$0.52, range: \$0.02–\$60) depending on their choices in the study – both amounts were in range of what was common MTurk practice at the time. We prevented repeated participation by excluding an additional 90 observations from duplicate Amazon worker IDs or IP addresses.

We collected data over a span of two separate weeks in November and December 2014, launching a total of 84 sessions.<sup>1</sup> We classified participation time as *day (night)* if the study was completed between 8am and 8pm (8pm and 8am). We classified participation day as *weekend* if the study was completed between the start of Friday *night* and the end of Sunday *day*, and *weekday* otherwise. In total, 844 participants took part during weekday-day, 819 during weekday-night, 345 during weekend-day, and 328 during weekend-night.<sup>2</sup>

We analyzed time of participation using participants' *experienced* time. To achieve this, we retrieved the participants' locations from their IP addresses, and adjusted their timestamp for their time zone (data were timestamped in the Eastern Time Zone because we are located there). As Figure 1 shows, most of our data (75%) originates from locations in the Eastern and Central Time Zones, which is consistent with 2014 Census estimates and recent evidence showing that MTurk can be more representative than in-person convenience samples (Berinsky et al. 2012).

All participants first took part in a battery of seven incentivized decisions: cooperation in the Prisoner's Dilemma (PD);<sup>3</sup> interpersonal altruism in a Dictator Game (DG) with a \$0.50

<sup>1</sup>Each session was closed after 30 participants accepted the HIT or 1 hour had elapsed, and participants had a maximum of one hour to complete the study. The first week (11/19-11/15) had 28 sessions launched every 6 hours starting at 00:00 EST; the second week (12/8-12/15) had 56 sessions launched every 3 hours starting at 09:00 EST. This difference in granularity is not relevant for our analyses of day versus night, which uses 12-hour blocks.

<sup>2</sup>Unless otherwise stated, we found qualitatively similar results when the 12-hours *night* was defined as beginning at 7pm or at 9pm, or if we define *weekend* as the time between the start of Saturday *day* and the start of Monday *day*.

<sup>3</sup>We used a continuous implementation of the PD (as in Capraro et al. 2014) such that each player received a \$0.40 endowment and chose how much to transfer to the other person, with any transfer doubled by the experimenters.

endowment; charitable giving (CH) where participants choose how much of \$60 to donate to the charity Oxfam International ([www.oxfam.org](http://www.oxfam.org)), with one participant selected at random to have their choice implemented; third-party punishment of selfishness (3P) and of fairness (AP);<sup>4</sup> honesty (HO) in a measure where participants guessed which random number between 1 and 20 would be generated by the computer and then self-reported accuracy, with more reported accuracy leading to higher earnings (up to \$0.50); and time discounting (TD).<sup>5</sup> To account for potential income effects, we randomized the order in which each task was presented at the individual level and informed participants that only one of the tasks would be randomly selected for payment after all were completed. All materials used neutral wording and the economic games included comprehension questions. See the online appendix for a copy of the instructions.

Finally, participants completed a 10-item version of the Big-5 measure capturing five dimensions of personality (O, openness; C, conscientiousness; E, extroversion; A, agreeableness; N, neuroticism (reverse-coded); from Gosling et al. 2003), a modified version of the cognitive reflection test to assess intuitive versus deliberative cognitive style (a set of three math problems with intuitively compelling but incorrect answers; original introduced by Frederick (2005), modified by Shenhav et al. 2012), and a set of standard demographic questions.

## 3 Results

### 3.1 Time and Day

**3.1.1 Incentivized Behaviors**—We begin with our central (null) result: Figure 2 shows the difference in mean behavior in each of the seven incentivized decisions between day and night, and between weekday and weekend.<sup>6</sup> For the games with comprehension questions—DG, PG and 3P—we exclude participants who answered incorrectly, but the results are qualitatively similar if included.

Although DG giving and donations to charity tend to be larger at nights (uncorrected  $p=0.032$  and  $p=0.015$ , respectively), and antisocial punishment tends to be larger on weekends (uncorrected  $p=0.018$ ), none of these differences survive even a modest Bonferroni correction for seven simultaneous tests (which would require  $p<.007$ ), let alone a more stringent correction for 21 tests that accounts for the 3 coefficients in each model.<sup>7</sup>

<sup>4</sup>In the third-party punishment game, Player 1 chose whether or not to evenly split \$0.50 with Player 2. The participant, in the role of Player 3, then chose how much of a \$0.10 endowment to spend on punishing Player 1 (with each cent reducing Player 1's payoff by 3 cents) if Player 1 did not (3P) or did (AP) split the \$0.50. Participants in our study played only in the role of the third player (which was our decision of interest). We did not deceive participants, however – a small number of Players 1 and 2 were recruited separately and repeatedly matched with Player 3s (as per Stagnaro et al. 2017).

<sup>5</sup>We used a short version of the discounting task developed by Kirby et al. (1999), where participants chose 9 different monetary allocations between a smaller reward and a larger, delayed reward (e.g. “Would you rather have \$25 today or \$60 in 14 days”). Log-transformed values reported in all analyses. One participant was selected at random to have one of their choices implemented. Because of the instructions stating “At the end of the study one participant and one question will be selected randomly. The winner will receive the associated bonus according to the choice made”, we had assumed that participants understood “today” to mean “at the end of the study” On reflection, we realize that this (unintentional) poor execution on our part might have been misunderstood by the participants.

<sup>6</sup>We report only main effects because preliminary ANOVAs reveal no significant interaction between a dummy for night versus day and weekend versus weekday. See Appendix Table A1 for significance levels of all the variables and Appendix Figure A1 for their distributions.

As our central findings are null results, we also conducted power analysis calculations. Setting the default power to 0.80 for three levels of alpha, based on the degree of conservativeness in Bonferroni correction ( $\alpha=0.05$ ;  $\alpha=0.007$ ;  $\alpha=0.0024$ ), we find that we had sufficient power to detect economically meaningful differences (differences of at least 5 percentage points for most measures even using the more conservative level of Bonferroni correction) in all but three cases: PD and 3P for all the alphas and HO for the most conservative one. See Appendix Table A2 for details.

We also ask how variance (rather than mean values) differs by day and time. The only difference we find that survive Bonferroni correction is that variance in antisocial punishment is lowest on weekday days, followed by weekday nights, and then higher in the two weekend timeslots. See Appendix Figure A2 for details.

Taken together, these results do not provide evidence that incentivized behavior in economic decisions on MTurk varies meaningfully with time or day.

**3.1.2 Demographics and Personality traits**—To investigate demographic and personality variations across time and day, we perform an ANOVA on each of the eighteen variables shown in Figure 3, with a *night* dummy, a *weekend* dummy, and the interaction between the two.<sup>8</sup> Once we apply Bonferroni correction, we find no significant interactions and only eight results with significance at the 5% level. In particular, people who participated at night took longer to complete the study (day: 2.90 log(sec); night: 2.95 log(sec);  $p<0.001$ ), were less experienced (day: 2.56 log(studies); night: 2.32 log(studies);  $p<0.001$ ), were less likely to be participating during their usual MTurk work times (day: 85%; night: 75%;  $p<0.001$ ), were less conscientious (C, Likert scale between 1 [less conscientious] and 7 [more conscientious]; day: 5.45; night: 5.26;  $p<0.001$ ) and more neurotic (N, reverse-coded Likert scale between 1 [more neurotic] and 7 [less neurotic]; day: 5.03; night: 4.84;  $p<0.002$ ); whereas people who participated on weekends were less experienced (weekday: 2.49 log(studies); weekend: 2.32 log(studies);  $p<0.001$ ), less likely to be participating during their usual MTurk work times (weekday: 83%; weekend: 74%;  $p<0.001$ ), and were less reflective (CRT correct answers; weekday: 1.57; weekend: 1.36;  $p<0.001$ ).

**3.1.3 Robustness Checks**—To ensure the robustness of our results we perform the following three robustness checks:

1. *The direction of the results does not change when splitting the data into two:* We divide our dataset into two based on whether the participant's serial order is odd or even. The incentivized economic behaviors that were strongly null in the full dataset are similarly null in each half. For DG and CH, which were weakly significantly (i.e. did not survive Bonferroni correction) larger at night than during the day in the full dataset, these results were not consistently apparent in

<sup>7</sup>We also test those seven null results for robustness to demographic and personality controls in stepwise regressions (Appendix Table A3). We find that such controls have no effect on the non-significance of the time/day coefficients.

<sup>8</sup>There were no significant interactions for the demographics, Figure 3 shows means for each condition to allow readers to see the absolute levels (which may be of general interest).

both halves of the data, further indicating lack of robustness. For AP, which was weakly significantly higher on weekends than weekdays in the full dataset, we observe the same result in both halves, suggesting that this result might be more robust. Finally, considering the significant demographic/personality results that were significant in the full dataset, the results were similar in the two halves (see Appendix Figure A3).

2. *Finer-grained definitions of time and day of the week:* For participation time we focus on four 6-hour intervals: *morning*, between 8am and 2pm; *afternoon*, between 2pm and 8pm; *evening*, between 8pm and 2am; and *pre-dawn*, between 2am and 8am. For day of the week we classify each day of the seven-day week separately (Monday-Sunday). Using these new definitions, we perform an ANOVA (with Bonferroni corrections) on all of the behavioral, demographic, and personality items. Doing so recovers all of the results described above using the more coarse-grained measures of time and day, with the only exceptions that neuroticism did not vary with time of day. We also found two new results that were not significant using the more coarse-grained analysis: participants at pre-dawn gave more generous donations to charity compared to the other times of day (pre-dawn: \$13.38; not pre-dawn: \$10.90,  $p < 0.001$ ), and age varied with time of day such that participants during the evening were younger while participants during pre-dawn were older (evening: 32.95; not evening: 35.74;  $p = 0.002$ ; pre-dawn: 35.74; not pre-dawn: 33.71;  $p < 0.001$ ). We note that the result regarding charitable giving was also evident in the coarse-grained analysis (Figure 2), but was only significant at the 5% level in that analysis (and thus did not survive Bonferroni correction).<sup>9</sup>
3. *Alternative definitions of night and weekend:* When nights are defined as either 9am to 9pm or 7am to 7pm (“N9” and “N7”, respectively), or when weekend is instead defined as Saturday day through Sunday night (“WS”), we only note six minor changes in terms of the significance that nevertheless shift the value of the affected variables to (non-)significant Bonferroni-corrected values. Specifically, the significance of Conscientiousness and Neuroticism disappears if N9 is used (from  $p < 0.001$  to  $p = 0.009$  and from  $p < 0.001$  to  $p = 0.017$ , respectively), the significance of CRT also disappears if WS is used (from  $p < 0.001$  to  $p = 0.003$ ), the significance of usual time on weekends disappears if WS is used (from  $p < 0.001$  to  $p = 0.001$ ), and the relationship between passing comprehension checks and spending time on the task at weekends gains significance when WS is used (from  $p = 0.001$  to  $p < 0.001$  and from  $p = 0.007$  to  $p < 0.001$ , respectively). See Table A1 for a complementary analysis of all remaining tasks.

### 3.2 Participant order

Finally, we test whether participants who take part in a study early on differ from those who participate later in the course of the study (and thus how important it is to have full

---

<sup>9</sup>See Appendix Table A4 for a complete list of the significance levels, and Tables A5 and A6 for regression analyses with dummies for each of the day/time categories as independent variables.

randomization over all treatments of an experiment, versus running some treatments after others have been completed). We run regressions on each of the measures presented in the previous section using the chronological order in which participants accessed our study as the independent variable.<sup>10</sup>

After Bonferroni corrections, we find that this variable predicts significant changes in five measures. To give a sense of the magnitude of these changes, we report values predicted from the regression models for the first participant (participant 1) and for the last participant (participant 2,336). We find that later-participating individuals are less experienced ( $b = -0.0002$ ,  $p < 0.001$ ; from 2.72 to 2.17 log(studies) [525 studies to 148 studies]), work at more unusual times ( $b = -0.00005$ ,  $p < 0.001$ ; from 14% to 26%), give more donations to charity ( $b = 0.002$ ,  $p < 0.001$ ; from \$9.65 to \$13.21 given, a 37% increase), are less reflective ( $b = -0.0002$ ,  $p < 0.001$ ; from 1.70 to 1.32 correct CRT answers, a 22% decrease), and are less agreeable ( $b = -0.0002$ ,  $p < 0.001$ ; from 0.156 to  $-0.156$  z-scored response, a 0.3 standard deviation decrease). We also note that when controlling for experience, the only difference that remains significant is agreeableness ( $b = -0.0001$ ,  $p < 0.001$ ).

## 4 Discussion

We investigated whether participants' economic game behavior, as well as demographics and personality factors, varied based on time of day and day of the week. Our key results are nulls: there are no significant differences on any of the incentivized economic behaviors. With respect to the non-incentivized measures, we do find that people participating on weekends were less reflective and less experienced, and less experienced, conscientious, and more neurotic when participating at night. Our finer-grained analysis also revealed more charitable giving between 2am and 8am. In addition to exploring time of day and day of the week effects, we also compared subjects who participated earlier in the study with those who participated later. We found later-participating subjects to have less prior experience, less reflectiveness, more charitable giving and less agreeableness.

With respect to the non-incentivized measures, a comparison between our results and those of Casey et al. (2016) reveals substantial convergence: both papers find more experienced participants earlier in the day and earlier in the data collection process, that participants who scored lower on the Big-5 personality dimension of conscientiousness were more likely to complete HITs later in the day, and that participants tended to score higher in the Big-5 personality dimension of agreeableness earlier in the data collection process.

We also note that our null result regarding time of day and honesty is inconsistent with prior work suggesting that people are more honest in the mornings (Kouchaki and Smith 2014). It is possible that this inconsistency results from the use of somewhat different honesty measures, or from some feature of how MTurk workers self-select into time of day for participation (e.g. their chronotype, as argued by Gunia et al. 2014). A more general point regarding our null results is that our games used instructions which were much shorter than

<sup>10</sup>See Appendix Figure A4 for a visual representation of cumulative averages over the data collection process, and Tables A7 and A8 for regression results. Our findings are qualitatively similar when using either a dummy for week, session number or the total of hours passed since the first session as an independent variable.

is typical for experimental economics, which could have led to more noise; however, we did screen for comprehension of the game payoffs, and prior work with the same short instructions has successfully observed correlations between game play and various other factors (Peysakhovich et al., 2014). Finally, we note that there was some evidence of more giving in the DG and charitable donation in the night relative to the day, but these differences were only significant when not including Bonferroni correction. Future work could assess whether our null findings for these measures replicate.

Broadly, our results suggest that researchers using MTurk to explore economic behavior need not be especially concerned about running studies during the day versus the night, or on weekdays versus weekends, or even without full randomization across treatments. This frees researchers to make fuller use of MTurk's ease of recruitment, collecting participants around the clock and throughout the week – and potentially comparing treatments and studies conducted at different times (although we note that lack of full randomization always introduces the possibility of threats to causal inference and encourage researchers to randomize across all conditions). However, if participants' level of prior experience, charitable giving, reflectiveness, agreeableness, neuroticism or consciousness seem likely to impact task performance (or, more importantly, interact with treatment effects for a given study, e.g. as in Rand et al. (2014) and Chandler et al. (2015)), researchers should use full randomization across treatments and be mindful of when they launch online studies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

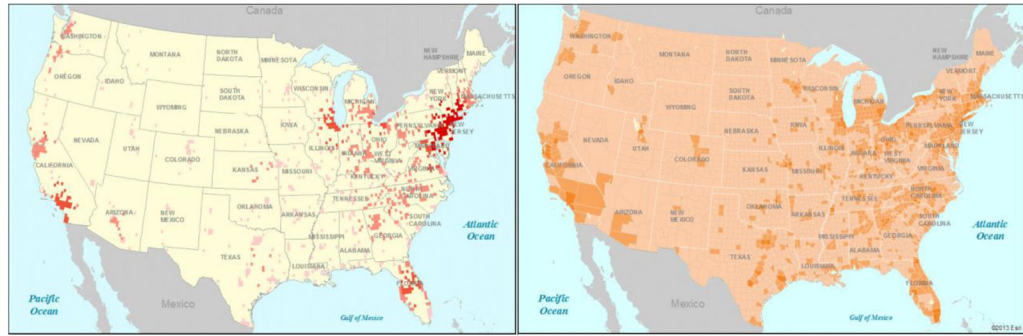
The authors gratefully acknowledge funding from the Templeton World Charity Foundation (grant no. TWCF0209), the Defense Advanced Research Projects Agency NGS2 program (grant no. D17AC00005), and the National Institutes of Health (grant no. P30-AG034420). We also thank Becky Fortgang, the Editor, and two anonymous reviewers for their valuable feedback, and SJ Language Services for copyediting.

## References

- Amir O, Rand DG, Gal YK. Economic Games on the Internet: The Effect of \$1 Stakes. *PLoS ONE*. 2012; 7(2):e31461. [PubMed: 22363651]
- Arechar, AA., Molleman, L., Gachter, S. Conducting interactive experiments online. *Social Science Research Network (SSRN)*; 2016. 2884409
- Aviv AL, Zelenski JM, Rallo L, Larsen RJ. Who comes when: personality differences in early and later participation in a university subject pool. *Personality and Individual Differences*. 2002; 33(3):487–496.
- Berinsky AJ, Huber GA, Lenz GS. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*. 2012; 20(3):351–368. DOI: 10.1093/pan/mpr057
- Capraro V, Jordan JJ, Rand DG. Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Scientific Reports*. 2014; 4:6790. [PubMed: 25348470]
- Casey LS, Chandler J, Levine AS, Proctor A, Strolovitch DZ. Intertemporal Differences Among MTurk Worker Demographics [Web]. *PsyArXiv*. 2016
- Chandler J, Paolacci G, Peer E, Mueller P, Ratliff KA. Using Nonnaive Participants Can Reduce Effect Sizes. *Psychological Science*. 2015; 26(7):1131–1139. [PubMed: 26063440]



- Deetlefs, J., Chylinski, M., Ortmann, A. MTurk 'Unscrubbed': Exploring the Good, the 'Super', and the Unreliable on Amazon's Mechanical Turk. 2015. Available at SSRN:<http://ssrn.com/abstract=2654056>
- Frederick S. Cognitive Reflection and Decision Making. *The Journal of Economic Perspectives*. 2005; 19(4):25–42.
- Gosling SD, Rentfrow PJ, Swann WB Jr. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*. 2003; 37(6):504–528.
- Gunia BC, Barnes CM, Sah S. The Morality of Larks and Owls: Unethical Behavior Depends on Chronotype as Well as Time of Day. *Psychological Science*. 2014; 25(12):2272–2274. [PubMed: 25287664]
- Horton JJ, Rand DG, Zeckhauser RJ. The Online Laboratory: Conducting Experiments in a Real Labor Market. *Experimental Economics*. 2011; 14(3):399–425. DOI: 10.1007/s10683-011-9273-9
- Kirby KN, Petry NM, Bickel WK. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology-General*. 1999; 128(1):78–87. [PubMed: 10100392]
- Kouchaki M, Smith IH. The Morning Morality Effect: The Influence of Time of Day on Unethical Behavior. *Psychological Science*. 2014; 25(1):95–102. DOI: 10.1177/0956797613498099 [PubMed: 24166855]
- Paolacci G, Chandler J, Ipeirotis PG. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*. 2010; 5(5):411–419.
- Rand DG. The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*. 2012; 299:172–179. [PubMed: 21402081]
- Rand DG, Peysakhovich A, Kraft-Todd GT, Newman GE, Wurzbacher O, Nowak MA, et al. Social Heuristics Shape Intuitive Cooperation. *Nature Communications*. 2014; 5:3677.
- Shenhav A, Rand DG, Greene JD. Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*. 2012; 141(3):423–428. [PubMed: 21928924]
- Stagnaro, MN., Arechar, AA., Rand, DG. From good institutions to generous citizens: Top-down incentives to cooperate promote subsequent prosociality but not norm enforcement. *Cognition*. 2017. <http://dx.doi.org/10.1016/j.cognition.2017.01.017>



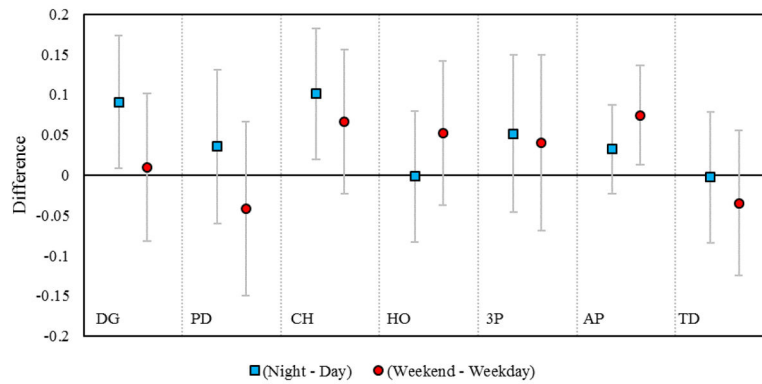
**Figure 1.** Location and population density of our sample (left) and the US (right). Darker points depict denser areas

Author Manuscript

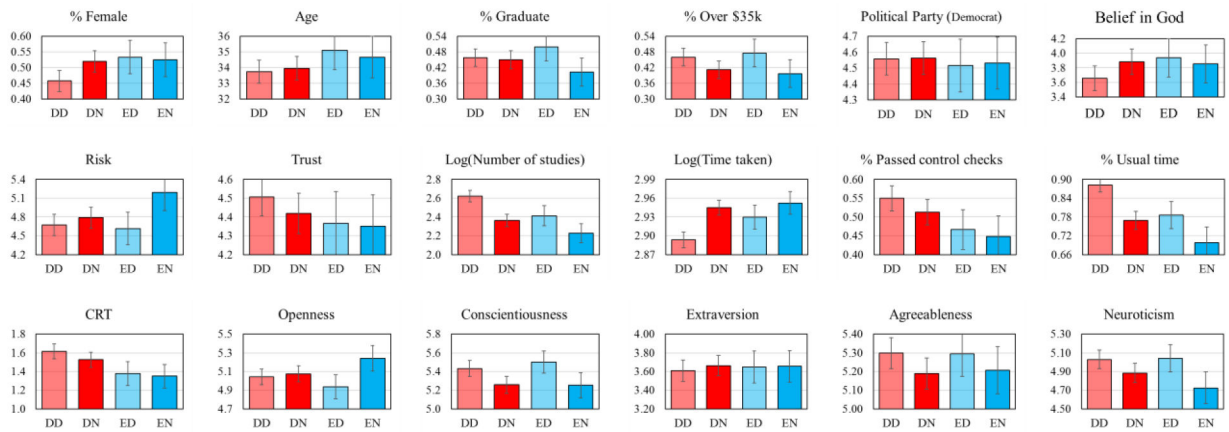
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2.** Differences in economic game behavior; 95% confidence intervals reported. We visualize *differences* in z-scored values to allow readers to more easily interpret the (lack of) main effects; positive values in the figure indicate higher values of the dependent variable during the day compared to the night, and weekday compared to weekend. DG: Dictator Game; PD: Prisoner’s dilemma game; CH: Charity task; HO: Honesty task; 3P: Prosocial third-party punishment; AP: Antisocial third-party punishment; TD: (log) Time discounting task.



**Figure 3.** Differences between nights and weekends on: gender, age, education, income, political party (1=Strongly Republican; 7=Strongly Democrat), belief in God (1=Very little; 7=Very much), willingness to take risks (0=Not at all willing to take risks; 10=Very willing to take risks), trust in others (1=Very little; 7=Very much), number of previous studies completed on MTurk, time taken to complete the current study, whether the comprehension questions were answered correctly, whether participants usually completed MTurk tasks at the time of day when they completed our study, number of correct answers in the Cognitive Reflective Task, Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. DD: Weekday day; DN: Weekday night; ED: Weekend day; EN: Weekend night. 95% confidence intervals reported. Mean values reported to allow readers to see the absolute levels (which may be of general interest).