



# Participant Crosstalk: Issues When Using the Mechanical Turk

John E. Edlund<sup>a</sup>, Kathleene M. Lange<sup>a</sup>, Andrea M. Sevene<sup>a</sup>, Jonathan Umansky<sup>a</sup>, Cassandra D. Beck<sup>a</sup> & Daniel J. Bell<sup>a</sup>

<sup>a</sup>Rochester Institute of Technology

**Abstract** ■ Past participants often talk about studies to those who have not yet participated (a problem termed participant crosstalk). Despite research exploring this issue in traditional settings, no research has explored crosstalk that occurs in online forums (in this case, related to mTurk). In study one, this research explores the kinds and prevalence of crosstalk in three online forums used by workers from mTurk. In study two, we assessed researchers' knowledge and attitudes of this crosstalk. Study three queried mTurk users about their experiences with crosstalk. Study four tested a simple method to attempt to reduce crosstalk on mTurk; this method eliminated crosstalk for this sample. We conclude by discussing the nature of crosstalk in online forums related to mTurk.

**Keywords** ■ mTurk; Participant Crosstalk.

✉ jeegsh@rit.edu

**iD** *JEE*: 0000-0003-3868-1844; *KML*: 0000-0002-8675-0739; *AMS*: 0000-0003-0765-806X; *JU*: 0000-0001-5033-6782; *CDB*: 0000-0003-1377-8818; *DJB*: 0000-0002-4044-914X

**doi** 10.20982/tqmp.13.3.p174

**Acting Editor** ■ Denis Cousineau (Université d'Ottawa)

**Reviewers**

■ One anonymous reviewer.

## Introduction

Foreknowledge of experimental methods and protocols has been identified as a key issue impacting the validity of psychological research (Nichols & Edlund, 2015). This foreknowledge has been termed Participant Crosstalk (Edlund, Sagarin, Skowronski, Johnson, & Kutter, 2009). For many years, researchers have examined the issues that occur when participants have foreknowledge of the study. For instance, Glinski, Glinski, and Slatin (1970) demonstrated that participants who had foreknowledge of a study protocol were meaningfully different in responses than participants who did not have the same foreknowledge. Specifically, the hypothesized effect only resulted when participants had foreknowledge of the research – calling into question the existence of the effect. More recently, researchers have demonstrated that most participants will attempt to confirm a researcher's hypothesis if they are able to ascertain the hypothesis (Nichols & Maner, 2008).

Estimates on the rates of participants coming in with foreknowledge are quite variable. Some researchers have

demonstrated that the rates can be as high as 78% (Lichtenstein, 1970), whereas other more contemporary multi-institutional studies have suggested lower rates in the single digits (Edlund et al., 2014). Importantly, any crosstalk that reveals key experimental information is critically important to identify and prevent as it has the potential to alter participant responses and thereby, the validity of the research (Nichols & Edlund, 2015).

Relative to the question of prevalence, the prevention of crosstalk has received less attention. Walsh and Stillman (1974) found mixed evidence for the effectiveness of an admonishment to avoid crosstalk. Edlund et al. (2009) found that a combined classroom and laboratory treatment effectively reduced crosstalk to a barely detectable level (<1%). In addition, Edlund et al. (2014) found evidence that a solely laboratory based treatment could be effective in reducing crosstalk, but only in institutions that demanded more out of their subject pools. However, to date, little research has investigated the prevention (or prevalence) of crosstalk outside of the typical introduction to psychology subject pool.



Perhaps the newest form of subject pool is Amazon's Mechanical Turk (mTurk: mturk.com). mTurk has been extensively used by researchers since its inception, resulting in hundreds of publications. However, mTurk presents its own set of challenges for researchers (for a comprehensive review of the challenges facing users of mTurk see Chandler & Shaprio, 2016). For instance, researchers have noted that there are differences in the quality of work provided by mTurk workers, such that high reputation workers actually provide more attentive responses to the research (Peer, Vosgerau, & Acquisti, 2014). Still, other factors can impact the results obtained. Siegel, Navarro, and Thomson (2015) explored how listing eligibility requirements impacts the participants recruited. They found that listing many eligibility requirements led to a skewed and less diverse sample of participants. Other researchers (Downs, Holbrook, & Peel, 2012) have explored various screening methods and metrics for mTurk and have found that these techniques are more likely to bias the data collection as opposed to improving data quality. Other labs (Chandler, Mueller, & Paolacci, 2014) have explored the frequency of which mTurk workers sign up for related studies. These issues (and others) have led some researchers to question the utility of mTurk and the conclusions reached, whereas others have argued that similar effects emerge when comparing samples collected through mTurk, social media, and traditional avenues (cbh15) as well as comparing mTurk to traditional campus and community samples (Necka, Cacioppo, Norman, & Cacioppo, 2016).

Researchers commonly need a certain kind of participant from the mTurk sample (e.g., only women, racial minorities, color-blind, etc). The commonly recommended method for obtaining these participants is to set up a screening questionnaire where a handful of demographic questions are assessed and used to invite mTurk workers to a different and more comprehensive study. Importantly, the validity of the research would be significantly impacted if mTurk workers were faking demographic factors for the purposes of being able to do the larger study (e.g., a male saying they were a naturally cycling female).

To date, few studies have explored whether mTurk workers are engaging in this form of crosstalk. However, one might ask, how would these workers engage in crosstalk as they are unlikely to be living in close proximity to other workers (as in the case of introduction to psychology subject pool crosstalk)? One potential avenue is a third party website. Numerous websites exist that are meeting spots for mTurk workers to discuss mTurk, "Turking", and individual experiences. For instance, Reddit has a subthread dedicated to mTurk and there are multiple websites that exist solely for users of mTurk (e.g., TurkOpti-

con). Still, other options exist where mTurk workers can discuss studies (such as twitter, Sun, 2016). As such, the purpose of this research was to explore whether crosstalk occurs on these third party forums, the kinds and severity of the crosstalk, and what other forums of generalized communication occurs.

## Study One

### Method

**Selection Protocol.** Three websites were identified due to their prominence in the mTurk user communities: The Hits Worth Turking For Reddit subthread (HWTF: [www.reddit.com/r/HITsWorthTurkingFor/](http://www.reddit.com/r/HITsWorthTurkingFor/)), Turkoption (TO: [turkooption.ucsd.edu/](http://turkooption.ucsd.edu/)), and the MTurk Forum (MTF: [www.mturkforum.com/](http://www.mturkforum.com/)). The research team coded the previous week's posts for each website.

**Coding protocol.** Numerous codes were developed for this protocol. The first code was termed: Key Crosstalk. This was crosstalk that the coding team deemed likely represented damaging and/or important crosstalk – crosstalk that appears to include information about manipulations in the research, key hypotheses or the like (e.g., "The study says it is about gut reactions, but they are really looking at people who might be racists"). The next code was termed: Qualifications. This code was indicated when the posting talked about qualifications necessary for qualifying for the mTurk study or what kinds of responses would open additional research (e.g., "They are looking for women not on birth control"). The third code was termed: Low Crosstalk. This was indicated when the participants were talking about the study, but in a way that likely did not reveal key experimental hypotheses (e.g., "You will be rating a number of pictures for how boring they are"). The fourth code was complaints about the hit. These codes were indicated when forum users were complaining about the research or the researcher (e.g., "The researcher will find any way to not credit your study. DO NOT PARTICIPATE"). The fifth code was praise about the research. This code was indicated when the forum posters gave praise to either the study or the researcher (e.g., "This was a very interesting study and they gave credits within 15 minutes"). The sixth code was discussions of the pay rate. This code occurred when posters comment on the rate of pay relative to the effort involved (e.g., "This study pays a rate of .50 per hour USD. We need to stop these researchers exploiting us"). The seventh code was mentions of violations of the mTurk Terms of Service (TOS). These codes were indicated when posters mentioned that something about the mTurk hit violated Amazon's TOS (e.g. "This is all to get you to like a facebook page. This clearly violates TOS. It has been reported to Amazon"). The eighth code was a comment on



**Table 1** ■ Breakdown of the presence of a type of communication by study discussed

	Key Crosstalk	Low Crosstalk	Qualifications	Complain About Hit	Praise Hit	Pay Rate	Violations of TOS	Dead Hit	Uncategorizable
HWTF	61/177	167/177	88/177	110/177	125/177	122/177	3/177	120/177	175/177
TO	14/50	50/50	18/50	49/50	45/50	49/50	8/50	41/50	0/50
MTF	12/51	50/51	36/51	36/51	28/51	51/51	0/51	49/51	37/51

the study representing a “dead hit”. Posters would make this comment after the research had concluded and was no longer available (e.g., “Dead hit”). The ninth and final code was reserved for any other comments that could not be otherwise categorized.

The two coders for this study conducted independent codings of all comments. The inter-rater reliability was .94. Discrepancies were resolved by a third coder who determined which codes were correct.

**Results**

During the week of January 26<sup>th</sup>, 2015, workers discussed 177 individual studies on HWTF, 50 studies on TO, and 51 studies on MTF, resulting in 4,338 individual comments.

**Level of Analysis: Study.** Given our interest in how each study was being talked about on mTurk, our first analysis approach focused on the study level (each discussed study was analyzed for the presence of each code). As such, regardless of how many individual comments were present in a particular thread, this analysis simply analyzed the presence of a particular kind of communication in the thread. See Table 1 for a full breakdown. Most importantly, across all three platforms, we saw approximately a third of all study discussions included Key Crosstalk information (87/278 studies), and nearly all postings involved either a discussion of key crosstalk, low crosstalk, or qualifications (270/278 studies).

**Level of Analysis: Proportion of Comments.** For this analysis, we took the raw numbers of codes from each thread and evaluated the percentage of each kind of comment. See Table 2 for a full breakdown of the number of comments. Most importantly, we saw that roughly 9% of comments involved Key Instances of Crosstalk and 39% involved Key Crosstalk, Low Crosstalk, or Qualifications.

**Discussion**

Overall, we found a disturbing level of crosstalk on all three investigated platforms (rates much higher than typically found in introduction to psychology labs). Key Crosstalk occurred at very high rates (33% of all studies discussed on the forums featured Key Crosstalk), and other concerning forms of communication occurred frequently as well (totaling 1679 instances in our sample). It is certainly likely that *some* of the analyzed comments represent

information that a participant may have learned in the description of the HIT or the consent form (such as qualification information) and in those cases, the level of concern that an individual researcher would have about that comment would be nil. However, in other cases, that information would be devastating (for instance, if a researcher was prescreening for participants with a physical disability and participants lied about that demographic information to qualify for a study). As such, qualification information being posted should be evaluated with caution.

Certainly, given that studies (e. g., Nichols & Maner, 2008) have suggested that at least in some cases, foreknowledge of the research question can change participant responses, we believe that any of the crosstalk (either key, low, or specifically on non-consent or description-based qualifications) is problematic. However, besides anecdotes, no study has explored how researchers who are experiencing participant crosstalk feel about such behaviors. Given the setup of the HitsWorthTurkingFor reddit subforum (the requestor’s ID was included), we had the unique opportunity to explore how the researchers who experienced crosstalk felt about such activities.

**Study Two**

**Method**

**Participants.** In the HWTF forum, mTurk requestor identity was provided. In many cases, the requestor ID provided meaningful real world clues to the identity. In some cases, the ID mentioned a specific individual’s name (e.g, John Doe), in some cases it alluded to a research laboratory (JDM lab at University of Nowhere), and in some cases the requestor ID was not attributable to a real world identity (e.g., MischiefManaged4). When names or university labs were mentioned, a google search was done attempting to link the requestor name to a real person’s email address (as you cannot contact a mTurk requestor outside of a hit). Of the 167 unique identifiers (some mTurk requestors in study one had multiple different studies discussed), 147 possible real individuals or labs were identified and sent an email (detailed below). Fifty individuals provided a response. Two individuals replied, but did not consent to their questions being included in this report and twelve individuals replied with a response that suggested they were

**Table 2** ■ Breakdown of a percentage of comments

Type of Comment	Percentage of Total Comments
Key Crosstalk	8.5% (370/4338)
Low Crosstalk	22.7% (985/4338)
Qualifications	7.5% (324/4338)
Complain about HIT	11.9% (518/4338)
Praise HIT	10.6% (461/4338)
Pay Rate	12.7% (553/4338)
Violations of TOS	0.06% (28/4338)
Dead HIT	9.5% (413/4338)
Uncategorizable	21.0% (912/4338)

the wrong person. As a result, we obtained data from 36 individual researchers.

**Procedure.** After a potential researcher’s email contact was ascertained, they were sent an email detailing some background for the email, the specific reason for the email, along with asking three research questions. The questions were: 1) Do you know your studies are being talked about in forums? Yes or No? 2) In your opinion is this sort of behavior of talking about studies harmful to your research? (Please elaborate) 3) Have you taken any steps to prevent participants from engaging in this outside of study communication? Yes or No and please explain

### Results

**Awareness.** 19 of the researchers (53%) expressed an awareness of their studies being discussed in the forums and 4 of the researchers (11%) noted that they were generally aware of the forums and the discussions but they were unaware any of their studies had been discussed. Multiple researchers who were previously unaware of the forums requested that the team provide them with the forum information (and it was provided as requested).

**Potential Harm.** The majority of respondents (61%) thought that the potential harm would depend based on the particular study they were running (variations on “it depends” or “maybe”). Some respondents (22%) thought the crosstalk would be harmful to their study and some respondents (17%) did not think it would be harmful. Preventative Steps. None of the surveyed respondents had taken any preventative steps to cut down on participant crosstalk.

### Discussion

This study sent targeted questionnaires to mTurk requestors whose studies on the reddit subforum featured potentially identifiable individuals. We found a great deal of variability in the requestor’s familiarity with the mTurk forums. Additionally, we found a great deal of variability in the concern that researchers have with this sort of

crosstalk, with a plurality of respondents thinking that it could (at least in some circumstances) pose an issue. Finally, none of our surveyed respondents had taken any steps to prevent crosstalk. Next we sought to evaluate the prevalence of crosstalk from the perspective of the mTurk workers.

### Study Three

In this study, we wanted to investigate the occurrence, prevalence, and nature of crosstalk from the perspective of the people potentially engaging in the crosstalk (which, to our knowledge, has never been pursued before). As such, in this study, we set up an mTurk study ostensibly about personality, where we assessed mTurk workers about their attitudes and experiences with crosstalk in mTurk.

### Methods

**Participants.** We collected data from 253 mTurk workers. There were 137 men and 113 women in the sample (along with 3 workers who did not answer the question), with an average age of 35.2 years ( $SD_{age} = 10.66$ ).

**Procedure.** In this study, we recruited mTurk workers to complete a survey on “Attitudes”. Of course, one might ask why we advertised our study as concerning attitudes (and even went so far as to include personality measures). Given our chief research question (the frequency and attitudes toward crosstalk in mTurk), we wanted our study to appear as similar as possible to other studies in mTurk and to not recruit a non-typical population (as we feared advertising our study as looking at attitudes towards mTurk or mTurk crosstalk could recruit a non-typical population).

After providing informed consent, participants initially completed three personality measures (the Big Five Inventory: John, Donahue, and Kentle, 1991; the Satisfaction With Life Scale: Diener, Emmons, Larsen, and Griffin, 1985; and the Need for Cognition Scale: Cacioppo, Petty, and Kao, 1984). After completing the personality measures, mTurk workers were asked about their personal behaviors regarding mTurk research and discussion boards,



**Table 3** ■ mTurk workers attitudes about crosstalk

	Mean (SD)
How often do you investigate a mTurk requestor prior to study <sup>a</sup>	3.39 (1.15)
Would negative (or positive) comments about a study influence your participation in a study <sup>a</sup>	3.58 (1.10)
Would negative (or positive) comments about a requestor influence your participation in a study <sup>a</sup>	3.70 (1.08)
How often do you talk about a study on a forum after participating in the study? <sup>a</sup>	2.13 (1.11)
What is your attitude towards investigating a requestor through mTurk forums? <sup>b</sup>	4.22 (0.84)
What is your attitude toward talking about a study after you participate in one through mTurk <sup>b</sup>	3.15 (1.16)

Note. <sup>a</sup> denotes a response scale of (1 = never, 2 = rarely, 3 = occasionally, 4 = often, 5 = always).

<sup>b</sup> denotes a response scale of (1 = strongly unsupportive, 2 = unsupportive, 3 = neither supportive nor unsupportive, 4 = supportive, 5 = strongly supportive).

what they have observed in discussion boards related to mTurk, along with basic demographics and questions related to their frequency of participation in mTurk.

### Results

Our first question of interest was investigating mTurk workers' attitudes towards various activities that span the range of potential crosstalk activities (see Table 3). Of particular note, mTurk workers were very supportive of investigating mTurk requestors prior to participation, whereas there was distinctly less support of discussing studies on the forums.

We also asked participants a number of questions regarding the behaviors they have engaged in and witnessed regarding crosstalk (see Table 4). Of particular note, we found that the majority of participants have seen key crosstalk in the forums (despite a majority of participants saying that they had not engaged in similar behaviors themselves) and that the majority of participants are aware of and attend to the forums.

Next (as part of a question), we revealed that this study was investigating crosstalk in the mTurk forums and asked participants directly about their experiences with crosstalk. The majority of participants (81%:  $N = 205$ ) had encountered studies that asked them not to talk about the studies outside of the research. However, it appears that in the mTurk workers' experience, this is a minority of studies (43%). Finally, it appears that these requests to avoid crosstalk are largely honored (88% of the mTurk workers report honoring the requests).

Finally, we also explored several questions related to the frequency of participation of the mTurk workers in studies and their belief regarding the appropriate pay rate of mTurk workers. We found tremendous variability in the number of studies per month that workers complete, ranging from 1 to over 40,000 ( $M_{studies} = 668.6, SD_{studies} = 3081.8$ ), which span a range of reasonable answers to a small number of absurd answers (40,000 studies in a month would suggest a rate of doing one study every

minute). After excluding likely outliers (>10,000 studies a month), the remaining 247 participants still report a large number of studies completed in a month ( $M_{studies} = 382.6, SD_{studies} = 700.9$ ). In terms of the appropriate pay rate for mTurk workers, we also found a great deal of variability ranging from a low of 50 cents an hour (USD) to 50 USD per hour ( $M_{pay} = 8.56USD, SD_{pay} 4.06$ ).

### Discussion

There are numerous findings that are both reassuring and concerning in the data from the mTurk workers. The majority of mTurk workers are clearly being exposed to key crosstalk information. Furthermore, in some cases, this exposure may "ruin" a participant for the study were they to participate in it. It also appears that some researchers are aware of the possibility of crosstalk and are taking steps to reduce it and these steps are largely being honored by the mTurk workers (at least, according to their self-reports).

### Study Four

Studies two and three provide multiple pieces of evidence which relate to crosstalk on mTurk. In study two, none of the identifiable researchers had taken any steps to reduce crosstalk. In study three, the data from the mTurk workers suggest that some mTurk requestors try to prevent crosstalk (whereas others do not). Furthermore, in study three, it appears that requesting that mTurk workers not engage in crosstalk is a potentially successful way to reduce crosstalk (as is suggested by Edlund et al., 2009; Edlund et al., 2014). However, these can best be described as converging, but non-definitive pieces of evidence suggesting that crosstalk on mTurk can be reduced by requesting that mTurk workers not engage in crosstalk. However, experimental evidence to this effect is lacking. As such, we decided to experimentally explore whether asking mTurk workers to refrain from engaging in crosstalk would be successful in decreasing the prevalence of crosstalk. In this study, we wanted to test a simple survey that would be unlikely to be impacted directly by the crosstalk (rather

**Table 4** ■ mTurk workers experiences with crosstalk on mTurk forums

	Percent of participants endorsing “Yes”
Read about study qualifications	70.8%
Read about study hypotheses	35.6%
Read about study goals	35.6%
Read about complaints related to a HIT	73.1%
Read about praise related to a HIT	68.8%
Read a discussion about pay rate	66.8%
Read about a study violating terms of service	59.7%
Read about a study being a dead HIT	51.0%

than feature a study that could be meaningfully impacted by crosstalk). In many ways, this study represents a lower bound of crosstalk rates (as suggested by Edlund et al., 2014).

### Method

**Participants.** We ran 400 mTurk workers through unrelated studies (200 in each condition). There were 187 men and 213 women in the sample, with an average age of 30.99 years ( $SD_{age} = 12.4$ ). **Procedure.** For this experiment, as part of an unrelated mTurk survey, we decided to test whether including a debriefing statement related to crosstalk at the end of the study decreased the prevalence of crosstalk. This debriefing statement followed the guidance of Edlund et al. (2009) in asking the mTurk worker to not post about the research for one month after the completion of the study on any mTurk forums due to the researchers desire to avoid future participants knowing about the goals/questions/hypotheses of the study. This debriefing was added to the end of the existing debriefing requested by the IRB detailing contact information and basic information about the study. Please see appendix 1 for the verbatim debriefing statement.

To avoid the influence of previous research conducted by our lab impacting participants (for instance, the requestor rating on Turkopticon), we set up two new mTurk requestor accounts for each condition (debriefing / no-debriefing) that were only used for this research. Administration of the study was standardized (in terms of crediting efficiency). Additionally, we temporarily retained mTurk ids to remove any overlapping participants from the study.

### Results

Given our focus on the discussion boards, we were only able to tally instances of the crosstalk (and were unable to link the occurrences to specific individuals). We actively tracked discussion boards for one week after the study concluded. We had 8 instances of crosstalk (4%) in the no debriefing condition and no instances of crosstalk (0%) in the

debriefing condition,  $\chi^2(1, N = 400) = 8.16, p = .004$ .

### Discussion

In line with the results found by Edlund et al. (2009), Edlund et al. (2014), we found that the request to mTurk workers to avoid crosstalk is a successful way to decrease participant crosstalk. Of note, in our relatively small sample (compared to the sample sizes used in the Edlund studies), we completely eliminated crosstalk in our sample where we requested participants avoid talking about the study for a predetermined length of time. It is also important to note that we are unable to determine whether the 8 instances of crosstalk were driven by a small number of participants posting repeatedly about the study, or whether this represents 8 discrete events (we will return to this issue in the general discussion).

### General Discussion

Overall, the results of this research suggest that the level of crosstalk among mTurk workers is quite concerning. A full third of the studies discussed in the forums featured what was likely key crosstalk information (involving things such as hypotheses, manipulations, and the like). Beyond that, a majority of studies discussed on the forums discussed ways of qualifying for a study, key crosstalk, or low crosstalk. Furthermore, it appears that many researchers are unaware of this method of extra-experimental communication. However, mTurk workers seem to be well informed on the ability to access this sort of information, and importantly, the workers themselves suggest that they will pay attention to requests to avoid discussing key experimental details. Furthermore, our final study suggests that by making a request to avoid crosstalk, mTurk requestors can significantly decrease the prevalence of crosstalk.

One important issue for researchers to consider when doing mTurk research is the impact that extra-experimental communication might have on their study. Certain forms of this communication may be beneficial to researchers. Comments such as “pays quickly and fairly”



and “interesting study” could potentially drive additional motivated mTurk workers to the research. Importantly, motivation level and the quality of the workers have already been independently noted as impacting the quality of the data collected (e. g., Peer et al., 2014). Certainly, this effect is not limited to mTurk and can be seen in the traditional research pool, but this is likely to be especially important in the mTurk environment.

However, other forms of communication are likely to be extremely harmful to the research. For instance, qualification information based on pretests being discussed on the forums could easily ruin a study and impact the validity of the conclusions reached in the research either by preventing a significant effect from emerging or creating a false positive). This is especially important as researchers have noted that the pool of mTurk workers might be much smaller than many researchers think. Stewart et al. (2015) has suggested that while there are tens of thousands of unique workers using mTurk in a six month window, there are only 7300 unique mTurk workers that may be encountered for a laboratory investigating similar issues).

One might rightfully ask, how prevalent crosstalk is on mTurk and could it meaningfully damage my research (indeed, as we suggest above, some kinds of crosstalk may even be beneficial)? After all, not every mTurk worker is likely to consult any (or all) of the forums. Furthermore, not every participant may be contaminated (by the very nature of crosstalk – it is impossible for the first person to be contaminated) and many participants may complete the study before the first crosstalk event occurs. Indeed, colleagues have suggested that the maximal possible adverse impact would be near 1% which would simply be part of the error variance in our analyses. We disagree with this evaluation for several reasons. One, study one suggests that key (and likely damaging) crosstalk occurs in one third of the crosstalk instances on the forums. Two, study four suggests in one of our ongoing mTurk studies (a study that would be unlikely to evoke crosstalk) that the observed crosstalk rate was 4% when we did not take steps to prevent crosstalk (but 0% when we included a simple treatment to reduce crosstalk). Finally, why wouldn't any researcher take a simple step to reduce crosstalk and increase the precision of their research?

Of course, this also indirectly points to one of the limitations of all of the recent research on crosstalk – we don't know what percentage of participants are engaging in the crosstalk (although the recent literature points to the num-

ber of participants likely contaminated by crosstalk). It is possible that crosstalk may be driven by a tiny number of former participants (perhaps one person crowing loudly to future participants). Conversely, crosstalk may be more widespread (a larger number of participants [although still a minority of participants] telling future participants about research). We believe that this question is one of the most important remaining unanswered questions in crosstalk research.

Based on this research, a number of suggestions for mTurk requestors can be offered. First, we encourage all mTurk requestors to familiarize themselves with the forums and to pay attention to what is being discussed about their individual studies and their researcher identification. The frequency of the attention paid to the forums should vary based on the research, but we argue all researchers should at a bare minimum familiarize themselves with the forums.

Second, we would encourage requestors to consider the

*We found a disturbingly high level of crosstalk in mTurk forums; ranging from key experimental details to discussions of how to qualify for research.*

impact crosstalk may have on their studies. In some cases, any form of crosstalk may have no meaningful impact (if a study does not engage in deception, such as the study we implemented in study four). In other cases, key crosstalk, qualification information, or low crosstalk could seriously impact the validity of the results obtained (for instance, Nichols and Maner, 2008 demonstrated that participants will attempt to confirm a re-

searcher's hypothesis; additionally, we have had studies in our laboratory that have been ruined by the goal of a pre-screening study being discussed on the forums).

Third, if any form of foreknowledge is problematic, we highly encourage researchers to have some form of a suspicion probe. Suspicion probes can vary in form, ranging from one short question, to a series of leading questions. We would point readers to Blackhart, Brown, Clark, Pierce, and Shell (2012) and Nichols and Edlund (2015) for a full explanation of some options associated with suspicion probes and the challenges that might be especially present in an mTurk sample.

Fourth, we highly suggest that researchers attempt to decrease the amount of crosstalk experienced. As demonstrated in this paper, this could be done through a fairly short and simple post-experiment debriefing. This form of debriefing has now been shown to significantly decrease crosstalk in mTurk and it has also been shown to virtually eliminate crosstalk in an undergraduate subject pool (suggesting significant utility).

Finally, we recommend that researchers build in some



manipulation checks or failsafes to detect when participants might not be giving truthful responses. This (in conjunction with a suspicion probe) could allow for the detection of crosstalk in samples before data analysis is commenced.

In summary, we found a disturbingly high level of crosstalk in mTurk forums; ranging from key experimental details to discussions of how to qualify for research. We also found many instances of less concerning extra-experimental communication. We found that many researchers are unaware that crosstalk occurs on the forums. However, we did not see a similar blindness to the forums by the mTurk workers. Finally, we experimentally demonstrated that crosstalk can be significantly reduced by the use of a simple debriefing statement related to crosstalk. Given the potential for crosstalk to significantly change the pattern of results generated by participants (e. g., Glinski et al., 1970; Nichols & Maner, 2008), we highly encourage researchers to pay attention to the forums looking for comments about themselves as requestors, their studies (especially when crosstalk would be particularly damaging), and to implement a debriefing to decrease crosstalk.

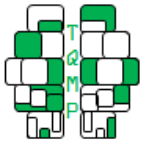
#### Authors' note

The authors wish to thank Austin Lee Nichols and Brad J. Sagarin for their helpful comments on earlier drafts. Portions of this research were supported by a Faculty Research Fund Grant through the College of Liberal Arts at the Rochester Institute of Technology.

#### References

- Blackhart, G., Brown, K. E., Clark, T., Pierce, D. L., & Shell, K. (2012). Assessing the adequacy of postexperimental inquiries in deception research and the factors that promote participant honesty. *Behavior Research Methods, 44*(1), 24–40. doi:10.3758/s13428-011-0132-6
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among amazon mechanical turk workers: consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112–130. doi:10.3758/s13428-013-0365-7
- Chandler, J. & Shaprio, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*, 53–81. doi:10.1146/annurev-clinpsy-021815-093623
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71–75.
- Downs, J. S., Holbrook, M. B., & Peel, E. (2012). *Screening participants on mechanical turk: techniques and justifications*. ACR North American Advances.
- Edlund, J. E., Nichols, A. L., Okdie, B. M., Guadagno, R. E., Eno, C. A., Heider, J. D., ... Wilcox, K. T. (2014). The prevalence and prevention of crosstalk: a multi-institutional study. *Journal of Social Psychology, 154*(3), 181–185. doi:10.1080/00224545.2013.872596
- Edlund, J. E., Sagarin, B. J., Skowronski, J. J., Johnson, S. J., & Kutter, J. (2009). Whatever happens in the laboratory stays in the laboratory: the prevalence and prevention of participant crosstalk. *Personality and Social Psychology Bulletin, 35*(5), 635–642. doi:10.1177/0146167208331255
- Glinski, R. J., Glinski, B. C., & Slatin, G. T. (1970). Nonnaivety contamination in conformity experiments: sources, effects, and implications for control. *Journal of Personality and Social Psychology, 16*(3), 478–485. doi:10.1037/h0030073
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory-versions 4a and 54*. Berkley, CA: University of California, Berkley, Institute of Personality and Social Research.
- Lichtenstein, E. (1970). "please don't talk to anyone about this experiment": disclosure of deception by debriefed subjects. *Psychological Reports, 26*(2), 485–486. doi:10.2466/pr0.1970.26.2.485
- Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the prevalence of problematic respondent behaviors among mturk, campus, and community participants. *PLoS ONE, 11*(6), 19.
- Nichols, A. L. & Edlund, J. E. (2015). Practicing what we preach (and sometimes study): methodological issues in laboratory research. *Review of General Psychology, 19*(2), 191–202.
- Nichols, A. L. & Maner, J. K. (2008). The good-subject effect: investigating participant demand characteristics. *The Journal of General Psychology, 135*(2), 151–165. doi:10.3200/GENP.135.2.151-166
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior Research Methods, 46*(4), 1023–1031. doi:10.3758/s13428-013-0434-y
- Siegel, J. T., Navarro, M. A., & Thomson, A. L. (2015). The impact of overtly listing eligibility requirements on mturk: an investigation involving organ donation, recruitment scripts, and feelings of elevation. *Social Science & Medicine, 1*, 42256–260. doi:10.1016/j.socscimed.2015.08.020
- Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300





amazon mechanical turk workers. *Judgment and Decision Making*, 10(5), 479–491. Retrieved from <http://wrap.warwick.ac.uk/71040/>

Sun, Y. (2016). *Listening to crowd workers and requesters on social media* (Master's thesis, Master's Thesis).

Walsh, W. B. & Stillman, S. M. (1974). Disclosure of deception by debriefed subjects. *Journal of Counseling Psychology*, 21(4), 315–319. doi:[10.1037/h0036683](https://doi.org/10.1037/h0036683)

#### Appendix: Debriefing statement of Study 4

*As you have surmised, one of the research questions in this study is investigating what sorts of communication occurs between workers surrounding various HITs. We are specifically investigating the kinds of information exchanged and whether various personality traits are related to how likely someone is to share this information.*

*We also ask that you not reveal this information to other workers (at least, not until we finish collecting data at the end of October). If you are interested in learning the results of this study, or you have any questions about the study, please contact the lead researcher (Dr. John Edlund at [john.edlund@rit.edu](mailto:john.edlund@rit.edu)).*

*Thank you for your participation!*

#### Citation

Edlund, J. E., Lange, K. M., Sevene, A. M., Umansky, J., Beck, C. D., & Bell, D. J. (2017). Participant crosstalk: Issues when using the mechanical turk. *The Quantitative Methods for Psychology*, 13(3), 174–182. doi:[10.20982/tqmp.13.3.p174](https://doi.org/10.20982/tqmp.13.3.p174)

Copyright © 2017, Edlund et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 18/08/2017 ~ Accepted: 07/09/2017