experiences surveying the crowd

reflections on methods, participation, and reliability

Cathy Marshall (Microsoft Research) & Frank Shipman (Texas A&M)

Websci 2013, Paris
2 May 2013
Recruiting study participants is a balancing act. You need diversity. You’d like convenience. You want to keep costs in check. So researchers in different disciplines have been conducting surveys by offering them as Mechanical Turk HITs. If you’ve logged onto Mechanical Turk as a worker, you might encounter 500 or so different surveys available for you to take for a small payment.
For years I used to be an avid survey taker for money... making anywhere from $200 to $300 a month...

...it's such a nice change of pace doing so many different types of surveys/studies on Mturk. I swear, if I did one more survey on my cell phone or on sanitary protection, I was going to pull my hair out.

--sandooch (a woman located in California)
US turkers take surveys for a variety of reasons

- for fun
  
  “...I do [surveys] quite often, largely for fun....”
  User rlrsk8str1, TurkerNation

- out of interest/to express opinions
  
  “…Since both surveys I did were on subjects where I have strong opinions anyway (Global Warming and Privatizing Social Security) they were very easy to write…”
  User JoeTurker, TurkerNation

- easy $ (e.g. while watching tv)
  
  “Took me 6 minutes, but I'm watching Judge Mathis. Also, one AC.”
  from HITsWorthturingfor subreddit

- also... desperation, community (c.f. Irani and Silberman), competition (c.f. von Ahn) and because they’re already in front of the screen!
Mechanical Turk as study venue

3 years, 7 media type-specific surveys offered as a HITS to >1000 qualified participants

We applied and extended best practices from literature (e.g. Kittur et. al., 2008)

In so doing, we learned about:

- methods (our focus here)
- Turkers who take surveys*
- reliability of the crowd*

*These topics are covered in the paper
using scenarios & hypotheticals to elicit social norms

step 1. characterize the phenomenon of interest
step 2. factor it into parts (actors, actions, concepts)
step 3. make up scenarios and hypotheticals
step 1: examples of real reuse phenomena

a convenient illustration

try to avoid social desirability bias
(Antin and Shaw, 2012)

satire/humor

commercial reuse
hypotheticals: from legal education and CBR

The pure hypothetical is a counterfactual variation on the fact pattern of an actual case. The hypothetical plays an important role in the Socratic style of law school teaching.

Solum, *Legal Theory Lexicon*, 2013

Hypotheticals may be used to ... *factor a complex situation into component parts (e.g., by exaggerating strengths, weaknesses or eliminating features).*

Rissland and Ashley, *Hypotheticals as Heuristic Device*, 1986
<Reuser> takes <action> as instance of <concept>

Dimension 1

Dimension 2

Dimension 3
Dimension 1: <reuser>
- Photographer (owner)
- Fishermen (subjects, possible rights holders)
- People of varying social distance
dimension 2: <action>

- **store**
  - e.g. save to USB key
  - e.g. store to the cloud

- **share**
  - e.g. post to Facebook
  - e.g. attach to email

- **(re)publish**
  - e.g. post to a public blog
  - e.g. upload to You Tube

- **remove**
  - e.g. untag a photo
  - e.g. change privacy settings
Types of reuse:

• Fishermen
• Photographer
• Other people of varying social distance from them

Dimension 3: <concept>
• Normal fair use
• Satirical reuse
• Commercial reuse
scenario + hypotheticals

On their vacation, Jules and Jim go fishing for catfish in the Mekong river. They make friends with Loan-anh. She snaps their picture after they catch what surely must be a record-breaking fish.

1. Loan-anh can post the photo to her Facebook account.
2. Jules can download the photo from Facebook and store it on his computer.
3. Jules can upload the photo to illustrate his public blog.
adding a twist

Carlos, a college student, is doing a talk on Chernobyl for class and wants to illustrate it with a funny picture. He finds the photo on Jules’ public blog and downloads it. He then adds cooling towers to the photo.

5. Carlos can use the photo for his presentation.
6. Carlos can post his presentation on a public slideshare site.

Carlos edits the photo to add a funny caption.

7. Carlos can submit the captioned photo to an e-postcard site.
8. Carlos can accept royalties from the e-postcard site.
participants judge hypotheticals using a Likert scale

- Strongly Agree
- Agree
- Agree somewhat
- Undecided
- Disagree somewhat
- Disagree
- Strongly disagree
this gives us the ability to compare across variables

- Carlos can use photo for pres.
- Carlos can post pres.
- Carlos can re-caption photo
- Carlos can accept royalties.
verdict

hypotheticals are an effective means of eliciting social norms provided:

– the story is real and grounded
– you don’t enumerate (3 actors x 4 concepts x 4 actions)
– the distinctions aren’t too nuanced

engagement matters! (e.g. Mason & Watts, 2010)
how to learn more about participants

• take an ethnographic approach to asking questions—Turkers answer them!

• read forums—Turkers talk among themselves
  – turkernation
  – hitsworthturkingfor and mturk subreddits
  – turkopticon (Irani and Silberman, 2013)
  – mturkforum

• read the crowdsourcing literature

_caveat: all crowdsourcing research seems to hit different segments of the worker community._
open-ended questions: a penny for your thoughts?

<table>
<thead>
<tr>
<th>Media</th>
<th>hypos</th>
<th>demo/practice</th>
<th>MC</th>
<th>open-ended</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>tweets</td>
<td>16</td>
<td>12</td>
<td>3</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>photos</td>
<td>18</td>
<td>14</td>
<td>3</td>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td>reviews</td>
<td>28</td>
<td>14</td>
<td>3</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>podcasts</td>
<td>22</td>
<td>14</td>
<td>2</td>
<td>4</td>
<td>42</td>
</tr>
<tr>
<td>videoconferences</td>
<td>20</td>
<td>13</td>
<td>2</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>ed. recordings</td>
<td>25</td>
<td>14</td>
<td>2</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>mmogs (games)</td>
<td>27</td>
<td>17</td>
<td>2</td>
<td>4</td>
<td>50</td>
</tr>
</tbody>
</table>

relative structure of the 7 media type specific surveys that we’ve done so far
open-ended questions elicit self-absorbed and uninhibited responses
(see Kiesler and Sproull, 1986)

e.g. From Games survey: What do you talk about with other players?

“Anything. A clan can sometimes be like a second family. If someone had a horrible day at work, you might empathize or give suggestions. Or it might be fun, if your playing a first person shooter to provoke them with political conversation. I'm not a jerk. Really. I just realized it sounds like I am. But i'm not. It can be in good fun! :P”
e.g. online activities
(an open-ended question with open-coded results)

what else do Turkers do online?
<table>
<thead>
<tr>
<th>category</th>
<th>subcategories</th>
<th>total</th>
<th>example (excerpt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social media</td>
<td>social networking; Facebook; Twitter; Forums; Reddit; Myspace</td>
<td>1020</td>
<td>&quot;Facebook! I am addicted to letting everyone know what I am doing &amp; reading what they are doing!&quot; [ED117]</td>
</tr>
<tr>
<td>Consume</td>
<td>reading; watching videos/tv/movies; listening to music/podcasts/radio</td>
<td>972</td>
<td>&quot;I also like to view the various websites that relate to Asian ball jointed dolls.&quot; [ED067]</td>
</tr>
<tr>
<td>Communicate</td>
<td>email; Skyping; IM/chat</td>
<td>757</td>
<td>&quot;I Skype with so many people even though I could just call them. Lol.&quot; [RE112]</td>
</tr>
<tr>
<td>Research</td>
<td>researching specific topics; using specific resources</td>
<td>548</td>
<td>&quot;I do use the Internet at work to look up information about medications and diseases. (I am a nurse.)&quot; [RE130]</td>
</tr>
<tr>
<td>Work and school</td>
<td>work-related activities; learning or homework; looking for jobs</td>
<td>460</td>
<td>&quot;Mostly I use the internet while I am at work. I am a manager at a law firm so I frequently use Excel and Word.&quot; [VC154]</td>
</tr>
<tr>
<td>Shopping</td>
<td>shopping (specific stores); shopping for specific items; find coupons</td>
<td>432</td>
<td>&quot;I often shop online, but even when I purchase from a bricks and mortar store I research my purchases online beforehand.&quot; [PH159]</td>
</tr>
<tr>
<td>Publish media</td>
<td>photo or video sharing; art; blogging; website development; media aggregation</td>
<td>297</td>
<td>&quot;...I work as a freelance illustrator. I use sites such as Blogger, Etsy, and Flickr to share my artwork and look at others' artworks and photos.&quot; [PC095]</td>
</tr>
<tr>
<td>category</td>
<td>subcategories</td>
<td>total</td>
<td>example (excerpt)</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Gaming</td>
<td>gaming (casual, online, multiplayer); fantasy sports; specific games (e.g. WoW)</td>
<td>257</td>
<td>&quot;I also play in some social virtual character websites such as Gaia Online and TinierMe.&quot; [PH147]</td>
</tr>
<tr>
<td>&quot;Get Paid To...&quot;</td>
<td>Mechanical Turk/HITs; surveys; other; Etsy and eBay selling</td>
<td>185</td>
<td>&quot;I spend a lot of my internet time (about 2-5 hours a day) doing mTurk because I am a single Mom and I need the extra cash.&quot; [PH069]</td>
</tr>
<tr>
<td>Entertainment</td>
<td>entertainment/fun; killing time/leisure; hobbies and crafts; porn</td>
<td>147</td>
<td>&quot;I also use the internet to get rid of stress and take a break from life.&quot; [PH214]; “I spend most of my time looking up craft patterns” [PH039]</td>
</tr>
<tr>
<td>non-specific</td>
<td>&quot;everything&quot;; various unspecified (e.g. &quot;etc&quot;); &quot;I don't&quot;</td>
<td>83</td>
<td>&quot;Ever since I was like 13 (I'm 29 now) the internet has been an integral part of my life.&quot; [PC144]</td>
</tr>
<tr>
<td>Financial</td>
<td>banking; bill-paying; investing/stock trading</td>
<td>48</td>
<td>&quot;The most time consuming activity that I do on the internet is day trade. I actively trade five days per week for four to five hours per day.&quot; [TW152]</td>
</tr>
</tbody>
</table>
a word or two about privacy

We didn’t ask about privacy. Respondents talked about it anyway.

e.g. what do you publish online? “likely too much, hah!”

Turkers readily surrender private data—they tend to rely on privacy through obscurity.

Data curation and reporting must be done with this in mind.

Adversarial relationships are reciprocated!
wrap-up: why we wrote this paper

• Big companies and researchers now rely on human computation.

• We had gathered extensive data on a large number of participants.

• The WORK is crucial to determining worker reliability. (engagement, purpose, creativity)

• The changing shape of information work is intriguing (and sometimes alarming).
contact info:
cathymar@microsoft.com
shipman@cs.tamu.edu
http://www.csdll.tamu.edu/~marshall
http://www.csdll.tamu.edu/~shipman
http://twitter.com/ccmarshall