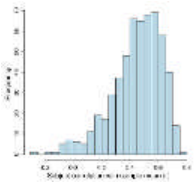
Akira O'Connor - Yesterday 12:02 - Limited

I am once again considering the use of Amazon's Mechanical Turk for some simple psychology experiments. I have spoken about this previously with +Tal Yarkoni who has, as usual, written an excellent blogpost on the topic, see below. I'm interested in gauging other folks' thoughts (and of course Tal's considered opinion once again) on the pros and cons of this experiment model.

[citation needed] » mechanical turk



some people are irritable, but everyone likes to visit museums: what personality inventories tell us about how we're all just like one another. Saturday, February 19th, 2011. I've recently sta…

- Comment - Share

Tom Hartley - I've thought about it. Seemed like there might be quite an overhead in getting it started. Keen to find out if it works for you and others.
Yesterday 12:07

Tal Yarkoni - We've had excellent results with MTurk. The main technical hurdle isn't really with MTurk, it's in getting your experiment online, period. You don't have to present your study using the MTurk tools (and I wouldn't recommend using them); you can use anything you like (e.g., SurveyMonkey if you're just talking about simple questionnaires). The way we use it is strictly as a recruiting and payment tool: workers sign up for HITs; they're directed to a separate site, where they do our (Ruby on Rails-based) experiments; once they're done they get a randomly generated code, which they then enter on the MTurk site to get their money. Works great.
Yesterday 16:00

Tom Hartley - Good to know - thanks for passing this on.
Yesterday 16:11

Yana Weinstein - I use something a little less depersonalized than MTurk but essentially the same idea - a company called Maximiles who send my experiment links out to their database of participants (they have 1+ million UK-based participants). You can collect up to 1,000 participants per day (if you have the money). I've never had results that differed from those collected in the lab.
Yesterday 16:18

Tom Hartley - How much does it cost, though Yana?
Yesterday 16:25

Yana Weinstein - The cost is 1 pound (roughly $1.60) per 10 mins per participant.
Yesterday 16:26

Tom Hartley - Thanks for responding. It sounds worth thinking about if the data is clean - especially useful for establishing norms etc. Would be a bit concerned it is the same one guy sitting in his basement or something - is there any way to verify the participants demographics etc.
Yesterday 16:32

Yana Weinstein - Well, you could certainly record their IP address, but for me the fact that each participant has a unique e-mail address (that's how they get paid by the company) is enough evidence. Also, roughly 100 people might be doing the experiment at any one time, so it can't physically be the same person.
Yesterday 16:34

Tal Yarkoni - FWIW, the going rate on MTurk is about $1 / hour (you can go cheaper or even free, you'll just get a slower response). At that rate you can basically get as much data as you like. I also haven't noticed any difference between MTurk and undergrad participants in quality control analyses (if anything, the MTurk participants give more reliable results). There are some nice features as far as worker management goes, e.g., (a) you don't have to pay workers who don't do the task properly (e.g., if you have reason to believe they're responding randomly); (b) you can see (and require as a qualification) a certain prior approval rate; and (c) Amazon has a pretty robust worker management system, so that you can restrict the sample to certain regions/countries (e.g., US only), and each worker has a unique ID so they can't participate in the same study twice. Overall I've been very pleased.
Yesterday 16:41

Yana Weinstein - It seems like MTurk has more going for it than the system I use (in particular, the rate is much lower). I guess I've just been sticking with what I know because I don't have to worry about learning my around the infrastructure.
Yesterday 16:43

Akira O'Connor - Thanks Yana, Tal and Tom - I can't tell you how much I appreciate your responses. For my next experiment I think I'll try MTurk in conjunction with some real participants, see how they compare and go from there.
Yesterday 16:45 (edited Yesterday 16:48) - Edit

Tom Hartley - Google+ working out really well, I note. Thanks for your advice all.
Yesterday 16:47

Tal Yarkoni - Michael Burhmester has a nice paper demonstrating the reliability of MTurk data as well as a how-to guide (haven't used it, but seems good) here: http://homepage.psy.utexas.edu/homepage/students/buhrmester/MTurk%20Guide.htm

Yesterday 16:48



Yana Weinstein - That has some useful links - I especially like the blog, http://experimentalturk.wordpress.com/. Tal, how easy is it to restrict the sample to certain demographics, e.g., age? Also, my one stumbling block with MTurk is that if your experiment is programmed on an outside website, you have to somehow get the user to open that website, complete the hit, and then go back to MTurk to say they've completed the hit. Then I guess you can use a unique code to match the two - but it seems a bit clunky. Does it work pretty seamlessly?
Yesterday 18:34



Tal Yarkoni - You can't (AFAIK) restrict the sample based on demographics, except for nationality. You could request that only some people participate, but they're not under any obligation to respect that, and you have no way of verifying it. But since the data are so cheap to collect, unless you have a very specialized group you need (in which case MTurk is probably not your best bet), you can always just analyze the subset you need (assuming you collecting demographic info).

Getting people to open another website is not a problem, it's very common. You just put the URL in the description of the task and have it open a new window when clicked (MTurk makes this easy). Beneath that is a form where the worker enters the code they get at the end. I'm sure the occasional worker might have problems, but I doubt it happens often, and since you don't pay unless the job is completed and you're satisfied with the work, it's not an issue for you.

The code doesn't have to be unique; you could use the same one for everyone. But that seems like a bad idea because then people could pass it along to their friends, who can claim they did your experiment without actually doing it. In practice, if you're going to be doing very large-scale studies (hundreds or thousands of participants), I think you have two viable options: (a) just approve everyone automatically and live with the fact that some small proportion (< 5%, I'd guess) is probably taking you for a ride (this isn't an issue in terms of data quality since you can clean it before analysis, it just means you're paying some people to cheat); or (b) use the MTurk API. The API allows you to programmatically access and download the worker results, and approve or reject their work. So with a bit of scripting, you could implement a system that automatically checks the quality of the data (e.g., my code compares the pattern of responding on personality questionnaires to the population average and rejects massive outliers), matches the data to each worker based on the code you gave them upon completion, and then informs MTurk which to approve and which to reject. This would be a fair amount of work up front, but you'd only have to do it once.

Personally I employ something of an intermediate approach: I run automated quality control analyses, flag the subject IDs that gave bad data, go into MTurk and reject those by hand (usually a small enough number that it doesn't take long), and then automatically approve everyone else. I'm a bit obsessive about data quality though, and even if you don't do any QC, you'll get data of comparable quality to undergrad samples. The one qualifier to that, un-PC as it may be, is that some Asian countries, most notably India, have a high proportion of people trying to game the system (at least in my studies). Once you exclude them, you're left with very few cheaters. I suspect the reason for this is that India has a highly educated and technically-savvy workforce where earnings are low, so $1/hour is actually quite a reasonable wage. If you recruit North American and European samples, you'll generally be dealing with people who aren't really doing it for the money, so it isn't an issue.
Yesterday 18:54 (edited Yesterday 18:55)



Tal Yarkoni - Oh, one other thing that can be problematic depending on what you're doing: the MTurk TOS forbid you from collecting personally identifying information. There's a way to email your workers through the system (haven't used it), but you can't actually ask them to give you their email address, name, etc. So that could be a problem if you need to track people over time.
Yesterday 18:56

Yana Weinstein - Thanks Tal for all this useful info. I guess I'm just clutching at straws as to the benefits of the system I'm using. I can ask the company to restrict the sample in terms of age/gender/education or whatever other variables I'm interested in that they have on file, if I so desire. One useful one could be "Native English Speaker". Another advantage of the company I use is that you do collect e-mail addresses, and so can do experiments with delays. Aside from that, the pricing structure of MTurk definitely trumps the small issues it has (people gaming the system, etc.) since you can afford to lose so many participants.
Yesterday 20:23

Pam Blundell - Are there any ethical issues though? Is it right to pay participants so little, etc?
Yesterday 23:24

Tal Yarkoni - Well, you're free to pay MTurk participants whatever you want... $1/hour is just the going rate, but if you're not comfortable paying them that, you can pay them as much as you want.

That said, I don't see any ethical issue at all. If it's okay to ask people to volunteer to participate in studies for **no** pay (which I've never heard anyone object to), I don't see the problem in asking them to participate in studies for **almost** no pay. Personally I think there's actually much more of an issue in requiring undergraduates to complete experiments for course credit, as you can make a very reasonable case that that's coercive (most of those students wouldn't voluntarily participate if grades weren't on the line!). So I see moving online as a win-win for everyone. It's easier to get data online, no one's forcing anyone to do anything they don't want to, and perhaps eventually the use of resources like MTurk will reduce the pressure on undergraduates to act as guinea pigs.
Yesterday 23:36

Add a comment...

©2011 Google - Terms - Content Policy - Privacy

Send feedback