

Using Amazon Mechanical Turk

Hadas Kotek

(based on materials created in collaboration with
Michael Y. Erlewine and Martin Hackl)


Experimental semantics and pragmatics
NYU, October 2017

In these slides

- **Last time: *turktools* 101**

- Creating a randomized items file (.csv) using `lister.py`
- Creating an html template from a skeleton using `templeter.py`
- Simulating the experiment using `simulator.py`

Today:

- The basics: how to use Turk.
 - The *Create* tab: creating a new project
 - The *Manage* tab: paying workers, retrieving a results file
 - Requesters and workers, background on Turk users.
-  Goal: understand the benefits, challenges, and basic process of uploading an experiment to Turk.

§1 Getting started

§2 The *Create* tab

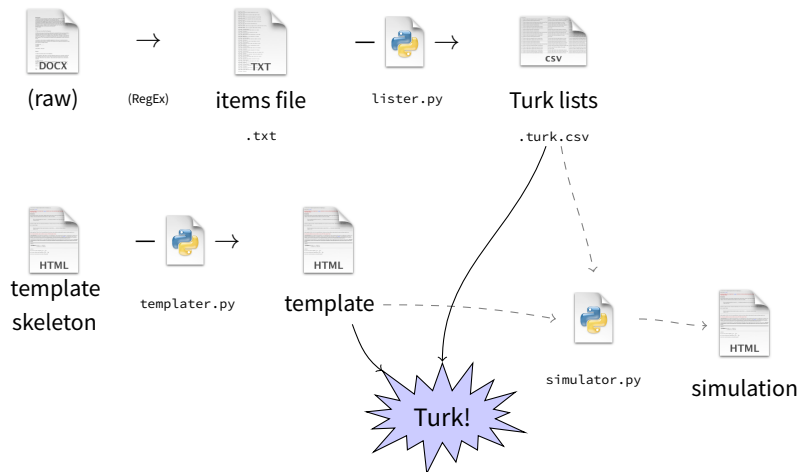
§3 The *Manage* tab

§4 The decoder .py script

§5 More on AMT and its workers

Last time: Turk technical workflow overview

aka "Turkflow"



Last time: Your template and items

What you give Turk:



Template file
.html



Turk items file
.turk.csv

Amazon's Mechanical Turk

- www.mturk.com
- Sign in as a *Requester*

The screenshot shows the Amazon Mechanical Turk homepage. At the top, there's a navigation bar with "Your Account", "HITS", and "Qualifications" buttons. Below that, a yellow banner reads "Mechanical Turk is a marketplace for work." and "237,794 HITS available. View them now." The main content is split into two columns: "Make Money by working on HITS" for workers and "Get Results from Mechanical Turk Workers" for requesters. Both sections include a three-step process diagram and a "Get Started" button.

amazon mechanical turk
Artificial Intelligence

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it's convenient.
237,794 HITS available. [View them now.](#)

Make Money by working on HITS

HITS - Human Intelligence Tasks - are individual tasks that you work on. [Find HITS now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → **Work** → **Earn money**

[Find HITS now](#)

[or learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITS - Human Intelligence Tasks - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITS completed in minutes
- Pay only when you're satisfied with the results

Fund your account → **Load your tasks** → **Get results**

[Get Started](#)

FAQ | Contact Us | Careers at Amazon | Developers | Press | Policies | Blog
©2005-2013 Amazon.com, Inc. or its Affiliates

An **amazon.com** company

Some terminology

- NB:** Our goal for today: learn to create and post your experiment on Amazon's Mechanical Turk.
- Participants are called **Workers**
 - Experimenters (i.e., we) are called **Requesters**
 - Tasks posted on AMT are called Human Intelligence Tasks, or **HITs**.
 - You should spend some time looking at AMT as a participant...
 - Most HITs are short and simple minded, and don't pay much; Participants can usually do more than one HIT per batch
 - ...so academic studies are unusual in (a) paying more, (b) taking longer, and (c) often, not allowing participants to do multiple tasks.
 - That's worth keeping in mind when writing instructions and thinking of the design.

The home screen

- Two main tabs we will discuss: *Create* and *Manage*

The screenshot shows the Amazon Mechanical Turk Requester interface. At the top, there is a navigation bar with the Amazon Mechanical Turk logo and the word "REQUESTER". Below this is a secondary navigation bar with tabs for "Home", "Create", "Manage", "Developer", and "Help". Under the "Home" tab, there are links for "Overview", "Tour", "Case Studies", "Pricing", "Partners", and "Business Solutions". On the right side of the navigation bar, there are links for "We're Hiring!" and "Learn More".

The main content area features a large orange heading: "Categorization on Mechanical Turk". Below this heading is a paragraph: "The Mechanical Turk **Categorization App** makes it simple to get fast, accurate results on your Categorization project!". To the right of this text is a graphic of three overlapping folders in blue, orange, and light blue.

Below the paragraph is a list of three bullet points, each with a green checkmark:

- ✓ Quick and easy HIT design
- ✓ Pre-qualified workers
- ✓ Start receiving results in minutes

Below the list is an orange button with the text "Create a Categorization Project".

At the bottom of the page, there are three sections:

- Work Distribution Made Easy**: Mechanical Turk gives businesses and developers access to an on-demand, scalable workforce. Accompanied by an icon of three stylized human figures.
- Categorization on Mechanical Turk**: The Mechanical Turk **Categorization App** makes it simple to get fast, accurate results on your Categorization Project. Accompanied by the folder icon.
- Sentiment Rating Simplified**: The Mechanical Turk **Sentiment App** makes it simple to collect and understand sentiment on your data! Accompanied by an icon of three sentiment faces (sad, neutral, happy).

§1 Getting started

§2 **The *Create* tab**

§3 The *Manage* tab

§4 The decoder .py script

§5 More on AMT and its workers

Setting up an experiment on AMT

Two steps to setting up an experiment:

- Create a template for your experiment:
 - `html` template, ← `templater.py` does this for you including instructions and practice items.
 - HIT specifications: ← do this on the spot
 - Name and description of task
 - Worker qualifications
 - Payment
 - Number of Workers
- Create an item lists file ← `lister.py` does this for you

...upload the items into the template and publish a “batch.”

The *Create* tab

- Create a new project here.

The screenshot shows the Amazon Mechanical Turk requester interface. At the top, there is a navigation bar with the Amazon Mechanical Turk logo and the word 'REQUESTER'. Below this is a secondary navigation bar with tabs for 'Home', 'Create', 'Manage', 'Developer', and 'Help'. The 'Create' tab is currently selected. Underneath the 'Create' tab, there are two links: 'New Project' and 'New Batch with an Existing Project', with the latter being highlighted. To the right of these links is the text 'Create HITs individually'. The main content area is titled 'Start a New Batch with an Existing Project' and contains a table of existing projects. Each row in the table represents a project and includes columns for 'Project Name', 'Title', 'Creation Date', and a set of action buttons: 'New Batch', 'Edit', 'Copy', and 'Delete'.

Project Name	Title	Creation Date ▼				
Completions the-every #2	Sentence Completion Task (COMP5)	February 19, 2013	New Batch	Edit	Copy	Delete
Completions same-every #2	Sentence Completion Task (COMP4)	February 11, 2013	New Batch	Edit	Copy	Delete
Completions same-every	Sentence Completion Task (COMP3)	February 4, 2013	New Batch	Edit	Copy	Delete
Completions the-every	Sentence Completion Task (COMP2)	February 4, 2013	New Batch	Edit	Copy	Delete
Completions pilot 2	Sentence Completion Task (COMP)	January 26, 2013	New Batch	Edit	Copy	Delete
Completions pilot	Sentence Completion Task (COMP)	January 21, 2013	New Batch	Edit	Copy	Delete
Presupp Proj: Inference Also	YES/NO Inference Task	January 5, 2013	New Batch	Edit	Copy	Delete
Presupp Proj: Inference 2nd	YES/NO Inference Task	November 13, 2012	New Batch	Edit	Copy	Delete
Presupp Proj: Inference 1	YES/NO Inference Task	October 17, 2012	New Batch	Edit	Copy	Delete

Creating a new project: Specifications

go to MTurk.com MIT Turkshop | Account Settings | Sign Out | Help

amazonmechanicalturk | REQUESTER

Home | **Create** | Manage | Developer | Help

New Project | New Batch with an Existing Project Create HITs individually

Edit Project

Specify the properties that are common for all of the HITs created using this project.

1 Enter Properties | 2 Design Layout | 3 Preview and Finish

Project Name: This name is not displayed to Workers.

Describe your HIT to Workers

Title:
Describe the task to Workers. Be as specific as possible, e.g. "answer a survey about movies", instead of "short survey", so Workers know what to expect.

Description:
Give more detail about this task. This gives Workers a bit more information before they decide to view your HIT.

Keywords:
Provide keywords that will help Workers search for your HITs.

This project may contain potentially explicit or offensive content, for example, nudity. [\(See details\)](#)

Creating a new project: Specifications

- Payment decisions: discuss with your instructors!
- Calculate how many observations per item you would like to have.
 - Recall: we created 8 lists (4 lists and their reverse) → 8 HITs
 - Each condition in each item appears in 2 lists
 - 3 assignments per HIT → 6 observations per condition/item
 - 5-item experiment → 30 observations per condition (across items)

Setting up your HIT

Reward per assignment
Tip: Consider how long it will take a Worker to complete each task. A 30 second task that pays \$0.05 is a \$0.00 hourly wage.

Number of assignments per HIT
How many unique Workers do you want to work on each HIT?

Time allotted per assignment Hours
Maximum time a Worker has to work on a single task. Be generous so that Workers are not rushed.

HIT expires in Days
Maximum time your HIT will be available to Workers on Mechanical Turk.

Results are automatically approved in Days
After this time, all unreviewed work is approved and Workers are paid.

[Advanced >](#)

Creating a new project: Specifications

- Advanced tab allows for some control over Workers.
 - Require high approval rate
 - Limit IP address to within the US (or other country)
(but remember: VPNs! This isn't fool-proof)

The screenshot shows the 'Advanced' tab of a project creation interface. At the top right, there is a link 'Worker requirements <'. Below this, the 'Worker requirements:' section contains a dropdown menu labeled 'Customize Worker Requirements...'. A sub-section titled 'Specify ALL the qualifications Workers must meet to work on your HITs:' contains two criteria: 'HIT Approval Rate (%) for all Requesters' HITs' set to 'greater than or equal to' 95, and 'Location' set to 'is' UNITED STATES. Each criterion has a 'remove' link. Below these is a button '(+) Add another criterion (up to 5)'. At the bottom left, there is a section 'Only Workers who qualify to do my HITs can preview my HITs.' with radio buttons for 'Yes' (selected) and 'No'. At the bottom right, there are 'Save' and 'Design Layout' buttons.

- AMT offers the use of “Masters.”
- I haven't found that they are any better at performing linguistics tasks than other workers.

Creating a new project: The template

The screenshot shows a web interface for creating a new project. At the top, there is a navigation bar with tabs for 'Home', 'Create', 'Manage', 'Developer', and 'Help'. Below this, there are links for 'New Project' and 'New Batch with an Existing Project', and a link for 'Create HITs Individually'. A yellow banner at the top of the main content area states 'Your project was successfully saved.' Below this is the 'Edit Project' section, which includes three steps: '1. Enter Properties', '2. Design Layout', and '3. Preview and Finish'. The 'Design Layout' step is currently active. In this step, the 'Project Name' is set to 'Turkshop sample' with a note that this name is not displayed to workers. The 'Frame Height' is set to 600, with a note that this height is in pixels and will be displayed to workers. A toolbar at the bottom of the design area includes an 'Edit HTML Source' button, which is circled in red. Below the toolbar, the 'Survey Code' is 'BLR', and there is a warning: 'PLEASE COMPLETE AT MOST ONE BLR SURVEY. YOU WILL NOT BE PAID FOR COMPLETING MORE THAN ONE SURVEY WITH THIS CODE.' The 'Instructions' section describes a task where the user must decide if 10 English sentences sound natural or not, and provides an example sentence: 'That's the play that the producer turned into a film.'

Home Create Manage Developer Help

New Project New Batch with an Existing Project Create HITs Individually

Your project was successfully saved.

Edit Project

Use the HTML editor below to design the layout of your HIT. This layout is common for all of the HITs created with this project. You can define variables for data that will vary from HIT to HIT. ([Learn more](#)).

1 Enter Properties 2 Design Layout 3 Preview and Finish

Project Name: Turkshop sample This name is not displayed to Workers.

Frame Height: 600 Height in pixels of the frame your HIT will be displayed in to Workers. Adjust the height appropriately to minimize scrolling for Workers.

Format Font Edit HTML Source

Survey Code: BLR

PLEASE COMPLETE AT MOST ONE BLR SURVEY. YOU WILL NOT BE PAID FOR COMPLETING MORE THAN ONE SURVEY WITH THIS CODE.

Instructions

This questionnaire presents 10 English sentences. Your task is to decide whether each sentence sounds natural or not. You might not be uttering the sentences yourself, but you might nevertheless have an intuition as to how natural they are. Please indicate whether the sentence sounds natural or not by selecting one of the buttons below the sentence. Here is an example:

That's the play that the producer turned into a film.

Creating a new project: The template

- Enter html template here
 - html template is created automatically via `templater.py`.
 - Open template in text editor, copy all, and paste into the window below.
 - Click on “Edit HTML Source” again to see how the template looks.

The screenshot shows a web development tool interface with three tabs: "Enter Properties", "Design Layout", and "Preview and Finish". The "Design Layout" tab is active, showing a project name "Turkshop sample" and a frame height of 600. Below this is a text editor with a toolbar and a "Save" button. The text editor contains the following HTML code:

```
<pre><p>Survey Code: <font color="blue">BLB</font></p>
<p><strong><font color="red">PLEASE COMPLETE AT MOST ONE </font><font color="blue" class="Apple-style-span">BLB</font><font color="red" class="Apple-style-span">SURVEY. YOU WILL NOT BE PAID FOR COMPLETING MORE THAN ONE SURVEY WITH THIS CODE.</font></strong></p>
<h3>Instructions</h3>
<p>This questionnaire presents 10 English sentences. Your task is to decide whether each sentence sounds natural or not. You might not be uttering the sentence yourself, but you might nevertheless have an intuition as to how natural they are. Please indicate whether the sentence sounds natural or not by selecting one of the buttons below the sentence. Here is an example:</p>
<hr />
<table>
  <tbody>
    <tr>
      <td style="width:600px" colspan="2">That's the play that the producer turned into a film.</td>
    </tr>
    <tr height="30">
      <td width="50%" align="center" style="">NATURAL</td>
      <td width="50%" align="center" style="">UNNATURAL</td>
    </tr>
    <tr>
      <td align="center" style=""><input type="radio" value="Y" name="RatingSample1" /></td>
      <td align="center" style=""><input type="radio" value="N" name="RatingSample1" /></td>
    </tr>
  </tbody>
</table>
<hr />
<p>In the above example, it is very clear that the sentence is natural, so please select the button on the left, labeled "Natural"</p>
<p>Here is another example:</p>
<hr />
<table>
  <tbody>
    <tr>
      <td style="width:600px" colspan="2">That's the door that the architect opened a hole.</td>
    </tr>
  </tbody>
</table>
</pre>
```


Creating a new project: The template

If you haven't done this already (when editing your skeleton), now is the time to edit your:

- Instructions
- Practice items
- Contact information
- Consent statement

(i.e., all the things that the template can't do for you.)

Creating a new project: The template

Edit Project

This is how your HIT will look to Mechanical Turk Workers. Before you publish these HITs, any variables in the HIT will be replaced with the input data that you provide when you publish the HIT. You can download a sample of the input file for this project or learn more about [acceptable file formats](#) [Download sample](#)

1. Enter Properties

2. Design Layout

3. Preview and Finish

Project Name:

This name is not displayed to Workers.

Sentence judgement task

Requester: HackLab

Reward: \$0.15 per HIT

HITs available: 0

Duration: 1 Hours

Qualifications Required: HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 95 , Location is US

HIT Preview

Survey Code: [BLR](#)

PLEASE COMPLETE AT MOST ONE BLR SURVEY. YOU WILL NOT BE PAID FOR COMPLETING MORE THAN ONE SURVEY WITH THIS CODE.

Instructions

This questionnaire presents 10 English sentences. Your task is to decide whether each sentence sounds natural or not. You might not be uttering the sentences yourself, but you might nevertheless have an intuition as to how natural they are. Please indicate whether the sentence sounds natural or not by selecting one of the buttons below the sentence. Here is an example:

That's the play that the producer turned into a film.

NATURAL


UNNATURAL



In the above example, it is very clear that the sentence is natural, so please select the button on the left, labeled "Natural."

Here is another example:

Creating a new project: Finishing touches

- Re-read your instructions.
 - Make sure your examples are clear for someone who is not reading the text carefully (i.e., most users).
 - Pick an email address that you check regularly.
 - Check the consent statement.
 - Check that there are many items as the instructions say there are (and as you expect there to be!).
 - Verify that the counter at the bottom of the page works properly.
-  All of this can (and should!) be done with the simulator, but go through everything again to verify that nothing went wrong.

Creating a new project: The template

HIT Preview

In order to get paid, please make sure that you answer all 10 items.

Consent Statement: By answering the following questions, you are participating in a study being performed by linguists in the Department of Linguistics and Philosophy at Massachusetts Institute of Technology. If you have questions about this research, please contact hackl.lab@gmail.com. Your participation in this research is voluntary. You may decline to answer any or all of the following questions. You may decline further participation at any time without adverse consequences. Your anonymity is assured, the researchers who have requested your participation will not receive any personal information about you.

If your browser has JavaScript turned on, a counter will be displayed at the bottom of the page indicating how many questions have been answered. It is **highly recommended that you turn on JavaScript and use this tool before submitting to ensure that all questions have been answered and you can receive payment.**

1.

\$(trial_1)

NATURAL	UNNATURAL
<input type="radio"/>	<input type="radio"/>

2.

\$(trial_2)

NATURAL	UNNATURAL
<input type="radio"/>	<input type="radio"/>

3.

\$(trial_3)

NATURAL	UNNATURAL
<input type="radio"/>	<input type="radio"/>

Finish

Creating a new project: The template

HIT Preview

\$(trial_9)

NATURAL UNNATURAL

10.

\$(trial_10)

NATURAL UNNATURAL

Are you a native speaker of English? Yes No

Do you speak a language other than English? Yes No

(Your answers to these questions do not affect the payment)

0 questions (out of 10 total) have been answered. If you submit now, you will not be paid.

HIGHLIGHT UNANSWERED QUESTIONS

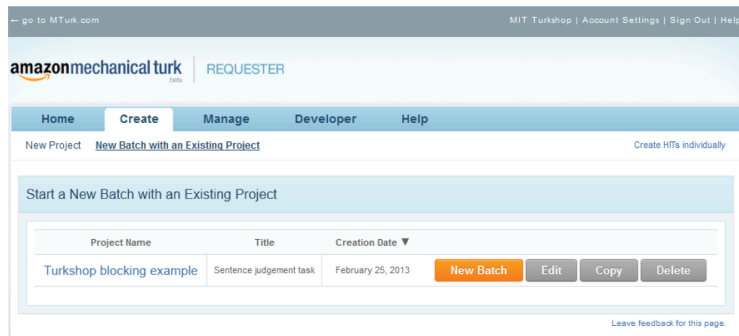
Submit

Finish

Posting your items on Turk: New batch

At this point a new Project should appear in your Create tab.

To post it to AMT, click on **New Batch**.



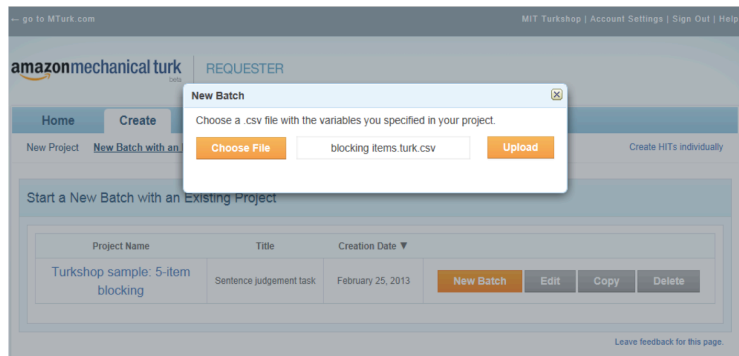
The screenshot shows the Amazon Mechanical Turk requester interface. At the top, there is a navigation bar with "Home", "Create", "Manage", "Developer", and "Help". The "Create" tab is selected. Below the navigation bar, there are links for "New Project" and "New Batch with an Existing Project", and a link for "Create HfFs individually". The main content area is titled "Start a New Batch with an Existing Project". It contains a table with the following data:

Project Name	Title	Creation Date ▼				
Turkshop blocking example	Sentence judgement task	February 25, 2013	New Batch	Edit	Copy	Delete

At the bottom right of the table area, there is a link: "Leave feedback for this page."

Posting your items on Turk: New batch

Choose the `.turk.csv` items file that you created using `lister.py`.



The screenshot shows the Amazon Mechanical Turk requester interface. A modal dialog box titled "New Batch" is open, prompting the user to "Choose a .csv file with the variables you specified in your project." The dialog contains a "Choose File" button, a text input field containing "blocking items.turk.csv", and an "Upload" button. The background interface shows the "REQUESTER" dashboard with a table of existing projects. The table has columns for "Project Name", "Title", and "Creation Date". One project is listed: "Turkshop sample: 5-item blocking" with the title "Sentence judgement task" and creation date "February 25, 2013". Below the table are buttons for "New Batch", "Edit", "Copy", and "Delete".

Project Name	Title	Creation Date	
Turkshop sample: 5-item blocking	Sentence judgement task	February 25, 2013	New Batch Edit Copy Delete

Posting your items on Turk: Previewing HITs

Go through all of the HITs you've created to make sure they all look good.

Home Create Manage Developer Help

New Project New Batch with an Existing Project Create HITs Individually

Preview HITs

Select HIT Template Upload Input Data **Preview** Confirm and Publish

This is how your HIT will look to Workers. Make sure that any variables in the HIT are correctly replaced by your input data, then click "Next".

Turkshop blocking example

Grammatically Judgement Task

Requester: HackLab Reward: \$0.1 per HIT HITs available: 8 Duration: 1 Hour

Qualifications Required: HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 95 .
Location is UNITED STATES

HIT Preview

Survey Code: BLR

PLEASE COMPLETE AT MOST ONE BLR SURVEY. YOU WILL NOT BE PAID FOR COMPLETING MORE THAN ONE SURVEY WITH THIS CODE.

Instructions

This questionnaire presents 10 English sentences. Your task is to decide whether each sentence sounds natural or not. You might not be uttering the sentences yourself, but you might nevertheless have an intuition as to how natural they are. Please indicate whether the sentence sounds natural or not by selecting one of the buttons below the sentence. Here is an example:

Posting your items on Turk: Previewing HITs

Go through all of the HITs you've created to make sure they all look good.

7.

That's the film that the assistant developed in the darkroom.

NATURAL UNNATURAL

8.

That's the book that the student made return to the library.

NATURAL UNNATURAL

9.

That's the student that the councilor made skip the road.

NATURAL UNNATURAL

10.

That's the gang that the officer leaned against the wall.

NATURAL UNNATURAL

Showing HIT 1 of 8 [Next HIT](#)

[Cancel](#) [Next](#)

Last step: Confirm payment and publish!

Batch Summary	
Batch Name: <input type="text" value="Turkhop blocking example 3"/>	Description: <input type="text" value="Decide whether 10 sentences sound natural or unnatural"/>
Batch Properties	
Title:	Sentence judgement task
Description:	Decide whether 10 sentences sound natural or unnatural (Survey for native English speakers only)
Batch expires in:	7 Days
Results are automatically approved after:	3 Days
Workers must meet the following Qualifications to work on these HITs:	HIT Approval Rate (%) for all Requesters' HITs score greater than or equal to 95 Location score is UNITED STATES
HITs	
Number of HITs in this batch:	8
Number of assignments per HIT:	x 3
Total number of assignments in this batch:	24
Cost	
Reward per Assignment:	\$0.100
	x 24 <small>(total number of assignments in this batch)</small>
Estimated Total Reward:	\$2.400
Estimated Fees to Mechanical Turk:	+ \$0.240 <small>(fee paid to Mechanical Turk) (see details)</small>
Estimated Total Cost:	\$2.640 <small>(this is the amount that will be deducted from your Available Balance when you click "Publish HITs")</small>
Your Available Balance:	\$300.000 <small>(before clicking "Publish HITs")</small>
Your Projected Balance:	\$297.360 <small>(after clicking "Publish HITs")</small>

[Back](#)[Publish HITs](#)

Last step: Confirm payment and publish!

If you decide to change anything at this stage:

- Payment
- Number of assignments per HIT
- Description
- Instructions, training items, etc.

Go back to template. Edit template, save, reload batch...

What it looks like from the Workers' end

The screenshot shows the Amazon Mechanical Turk worker interface. At the top, there's a navigation bar with 'Your Account', 'HITS', and 'Qualifications' tabs. A notification indicates '240,066 HITS available now'. Below this is a search bar with 'HITS' selected and a filter for 'that pay at least \$ 0.00'. The main content area is titled 'All HITS' and shows '1-10 of 3069 Results'. The results are sorted by 'HITS Available (most first)'. A list of HITs is displayed, each with a title, requester name, expiration date, time allotted, reward, and number of available HITs. Each entry includes links for 'Request Qualification', 'Why?', and 'View a HIT in this group'.

HIT Title	Requester	HIT Expiration Date	Time Allotted	Reward	HITs Available
Inv. B. 2	rahz@id	Mar 16, 2013 (2 weeks 4 days)	48 minutes	\$0.00	15308
Search: Keywords on Google.com [US]	CrowdSource	Feb 25, 2014 (52 weeks)	16 minutes	\$0.08	12455
Copy Text from Business Card	Oscar Smith	Feb 25, 2013 (1 hour 59 minutes)	10 minutes	\$0.02	12109
Collect pricing information for parts	Jeffrey Werby	Feb 26, 2013 (15 hours 15 minutes)	10 minutes	\$0.05	7630
Check if a Tweet Violates Twitter's Ads Policy (CWR ID: 1052)	Jack Stone	Mar 1, 2013 (3 days 11 hours)	15 minutes	\$0.15	7527
Copy Text from Business Card	Oscar Smith	Feb 25, 2013 (1 hour 59 minutes)	10 minutes	\$0.02	7073
Find company contact information	Readab	Mar 4, 2013 (6 days 23 hours)		\$0.05	

- §1 Getting started
- §2 The *Create* tab
- §3 **The *Manage* tab**
- §4 The decoder .py script
- §5 More on AMT and its workers

The *Manage* tab

(Time permitting: exercise — get your experiment onto AMT!)

Okay, so you've posted your experiment.. now what.

The *Manage* tab

(Time permitting: exercise — get your experiment onto AMT!)

Okay, so you've posted your experiment.. now what.

The *Manage* tab

- Monitor the progress of your experiment
- Get your data!
- Approve/reject subjects

Home Create **Manage** Developer Help

Results Workers Qualification Types Manage HITs individually

Manage Batches

Click on the name of the batch to see more details

▼ Batches in progress (1)

Turkshop blocking example **Results** Cancel this batch

Created:	February 25, 2013	Assignments Completed:	1 / 48
Time Elapsed:	8 minutes	Estimated Completion Time:	February 26, 2013 4:11 AM PST (Tuesday)
Average Time per Assignment:	1 minute 46 seconds	Effective Hourly Rate:	\$3.396

Batch Progress:

2% submitted 100% published

▼ Batches ready for review (25) < Previous 1 2 3 Next >

Completions the every #2 **Results** Delete

Created:	February 19, 2013	Assignments Completed:	144 / 144
Time Elapsed:	6 days	Estimated Completion Time:	COMPLETE
Average Time per Assignment:	11 minutes 30 seconds	Effective Hourly Rate:	\$1.304

Batch Progress:

100% submitted 100% published

The Manage tab

Review Results

Select the check boxes on the left to approve or reject results. You only pay for approved results. To evaluate results offline, select Download CSV.

For additional batch information, view batch details.

Turkshop blocking example

Customize View

Filter Results

Upload CSV

Download CSV

20 of 20 assignments (FILTER APPLIED: only show assignments that are in "Submitted" status)

Approve

Reject

<input type="checkbox"/>	HIT ID ▲	Worker ID	Lifetime Approval Rate	Input List	Input Trial 1	Input Trial 2	Input Trial 3	Input Trial 4	Input Trial 5	Input Trial 6	Input Trial 7	Input Trial 8	Input Trial 9	Input Trial 10
<input type="checkbox"/>	24ZMV1HVKCJ019YNOJU9PQW7HJ475Z	A1DKEJ4G4QQD10	0% (0/0)	4	That's the student that the counselor made skip...	That's the ladder that the worker leaned against...	That's the wall that the carpenter broke into s...	That's the intern that the supervisor made retu...	That's the film that the assistant developed in...	That's the sprinter that the coach ran for 2 ho...	That's the rabbit that the magician made vanish...	That's the boy that the instructor made float with th...	That's the girl that the teacher played with...	That's the book that the girl made balance on h...
<input type="checkbox"/>	2BC00QQKKA4Q6T9H5T6D0EXGNGYY6	A1MLJ69XSGEYB8	0% (0/0)	0	That's the rabbit that the magician made vanish...	That's the gymnast that the coach made balance...	That's the idea that the team played with th...	That's the wall that the carpenter made float at the...	That's the film that the mechanic ran idle fo...	That's the engine that the assistant developed in...	That's the book that the student made return to...	That's the student that the counselor made skip...	That's the officer that the counselor leaned against...	That's the gang that the officer learned against...
<input type="checkbox"/>	2BC00QQKKA4Q6T9H5T6D0EXGNGYY6	A1A03XN33VQGA2	0% (0/0)	0	That's the rabbit that the magician made vanish...	That's the gymnast that the coach made balance...	That's the idea that the team played with th...	That's the wall that the carpenter made float at the...	That's the film that the mechanic ran idle fo...	That's the engine that the assistant developed in...	That's the book that the student made return to...	That's the student that the counselor made skip...	That's the officer that the counselor leaned against...	That's the gang that the officer learned against...

The *Manage* tab

The *Manage* tab is also where you approve or reject submissions.

- Reject subjects who didn't comply with instructions in obvious ways.
 - Did not complete experiment.
 - Exhibit guessing behavior.
 - Failed miserably on 'catch' items.
- But do **NOT** reject Non-native speakers!
- I'd also recommend paying those who completed more than one survey (you'll be able to ignore that data later).

In general, you should aim to pay anyone who's done honest work for you!

👉 Approval rates matter a lot to Turk Workers.

The *Manage* tab

The *Manage* tab is also where you approve or reject submissions.

- Reject subjects who didn't comply with instructions in obvious ways.
 - Did not complete experiment.
 - Exhibit guessing behavior.
 - Failed miserably on 'catch' items.
- But do **NOT** reject Non-native speakers!
- I'd also recommend paying those who completed more than one survey (you'll be able to ignore that data later).

In general, you should aim to pay anyone who's done honest work for you!

 Approval rates matter a lot to Turk Workers.

The *Manage* tab

The *Manage* tab is also where you approve or reject submissions.

- Reject subjects who didn't comply with instructions in obvious ways.
 - Did not complete experiment.
 - Exhibit guessing behavior.
 - Failed miserably on 'catch' items.
- But do **NOT** reject Non-native speakers!
- I'd also recommend paying those who completed more than one survey (you'll be able to ignore that data later).

In general, you should aim to pay anyone who's done honest work for you!

 Approval rates matter **a lot** to Turk Workers.

- §1 Getting started
- §2 The *Create* tab
- §3 The *Manage* tab
- §4 **The `decoder.py` script**
- §5 More on AMT and its workers

Decoding MTurk's results file



Raw results file
.CSV



decoder.py



Decoded output
.CSV

👉 Take care of some basic analysis steps automatically; makes spreadsheet easier to work with.

- Open the raw `example4-results.csv` results file in Excel.
 - What information does it give us?
- Decode using `decoder.py`
- Now open the decoded file in Excel.
 - What information does it give us?

Fields in the decoded results file

- **WorkerId**: Amazon ID for the worker
- **AssignmentId**: unique to that submission
- **AssignmentStatus**: Submitted, Approved, Rejected
- **WorkTimeInSeconds**
- **ListNumber, PresentationOrder**
- **Section, Item, Condition**
 - Based on the input to Lister
 - Factor settings: if condition name included “-”
- **field_N**: your actual stimuli
- From the template skeletons:
 - **Choice**: forced choice value
 - **Extras**: english, foreignlang, numanswered, useragent

Roadmap

- §1 Getting started
- §2 The *Create* tab
- §3 The *Manage* tab
- §4 The decoder .py script
- §5 **More on AMT and its workers**

- Demographics
- Locations, languages
- Naïveté, Worker Training, Satisficing
- Accuracy/Reliability for Linguistic tasks

<https://sites.google.com/site/amazonturkshop/link>

<http://experimentalturk.wordpress.com/2012/10/09/slides-from-acr-2012/>

More than 500,000 registered Workers

- 46% US
- 34% India (50% as of May 2010)
- 19.20% Other (64 countries)

<http://techlist.com/mturk/global-mturk-worker-map.php>

<http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html>

More than 700,000 HITs per week

- Conjectured to be actually about 5,950,000 HITs per week
- 80% of HITs are performed by 20% most active workers (est. to be between 3,011 and 8,582 workers)
- They spend more than 15 hours a week on AMT

(Unless stated otherwise the information is from Fort et. al (2011): Amazon Mechanical Turk: Gold Mine or Coal Mine? *Computational Linguistics* 37,2. 413-420.)

Where is the AMT workforce?



50,000 MTurk workers who have performed techlist HITs. All data copyright © 2011 techlist.com. (<http://techlist.com/mturk/global-mturk-worker-map.php>)

Demographics, Economics, and Ethics

- Workers earn less than US\$ 2.- per hour (below minimum wage)
 - For 20% of workers, their AMT earnings represent primary source of income,
 - For 50% it is their secondary income
 - Unregulated workplace (no collective bargaining)
 - No channels of redress of employer wrongdoing (e.g. no payment)
- There are alternative crowd-sourced work environments
 - Train their workforce
 - Pay minimum wage
- Implications of labor costs for research/funding expectation etc.
- The NPR Planet Money podcast recently (January 2015) did a story interviewing Turk workers: <http://www.npr.org/blogs/money/2015/01/30/382657657/episode-600-the-people-inside-your-machine>

Demographics, Economics, and Ethics

- Workers earn less than US\$ 2.- per hour (below minimum wage)
 - For 20% of workers, their AMT earnings represent primary source of income,
 - For 50% it is their secondary income
 - Unregulated workplace (no collective bargaining)
 - No channels of redress of employer wrongdoing (e.g. no payment)
- There are alternative crowd-sourced work environments
 - Train their workforce
 - Pay minimum wage
- Implications of labor costs for research/funding expectation etc.
- The NPR Planet Money podcast recently (January 2015) did a story interviewing Turk workers: <http://www.npr.org/blogs/money/2015/01/30/382657657/episode-600-the-people-inside-your-machine>

Demographics, Economics, and Ethics

- Workers earn less than US\$ 2.- per hour (below minimum wage)
 - For 20% of workers, their AMT earnings represent primary source of income,
 - For 50% it is their secondary income
 - Unregulated workplace (no collective bargaining)
 - No channels of redress of employer wrongdoing (e.g. no payment)
- There are alternative crowd-sourced work environments
 - Train their workforce
 - Pay minimum wage
- Implications of labor costs for research/funding expectation etc.
- The NPR Planet Money podcast recently (January 2015) did a story interviewing Turk workers: <http://www.npr.org/blogs/money/2015/01/30/382657657/episode-600-the-people-inside-your-machine>

Demographics, Economics, and Ethics

- Workers earn less than US\$ 2.- per hour (below minimum wage)
 - For 20% of workers, their AMT earnings represent primary source of income,
 - For 50% it is their secondary income
 - Unregulated workplace (no collective bargaining)
 - No channels of redress of employer wrongdoing (e.g. no payment)
- There are alternative crowd-sourced work environments
 - Train their workforce
 - Pay minimum wage
- Implications of labor costs for research/funding expectation etc.
- The NPR Planet Money podcast recently (January 2015) did a story interviewing Turk workers: <http://www.npr.org/blogs/money/2015/01/30/382657657/episode-600-the-people-inside-your-machine>

Naïveté, Cheating, Satisficing, Task Complexity

- **Naïveté:** It is often said that gathering data from naïve participants is important to ensure getting unbiased data (e.g. Gibson et al. 2010)
- How naïve are the AMT workers?
 - Between 3,011 and 8,582 workers do 80% of the HITs
 - Repeat participants
 - Cross-talk
- When/what type of Naïveté is desirable?
- Informant training
 - Metalinguistic tasks
 - Co-reference, Variable Binding

Naïveté, Cheating, Satisficing, Task Complexity

- **Naïveté:** It is often said that gathering data from naïve participants is important to ensure getting unbiased data (e.g. Gibson et al. 2010)
- How naïve are the AMT workers?
 - Between 3,011 and 8,582 workers do 80% of the HITs
 - Repeat participants
 - Cross-talk
- When/what type of Naïveté is desirable?
- Informant training
 - Metalinguistic tasks
 - Co-reference, Variable Binding

Naïveté, Cheating, Satisficing, Task Complexity

- **Naïveté:** It is often said that gathering data from naïve participants is important to ensure getting unbiased data (e.g. Gibson et al. 2010)
- How naïve are the AMT workers?
 - Between 3,011 and 8,582 workers do 80% of the HITs
 - Repeat participants
 - Cross-talk
- When/what type of Naïveté is desirable?
- Informant training
 - Metalinguistic tasks
 - Co-reference, Variable Binding

It makes me wonder if some of the folks posting surveys are somehow unaware that turkers take hundreds of surveys. If they're unaware, they don't know what they're doing.

If they *are* aware, why do I see so many surveys cutting-and-pasting questions like, as an example, these:

A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

These questions can only work to measure what they're supposed to measure the very first time a survey-taker sees them. After that, the cat's out of the bag. Why would a serious researcher put that sort of content into a survey that goes to a pool of professional survey takers? It's a tainted pool. Do they not notice that MTurkers get those questions right way more than the general population? Again, I suspect many of these researchers simply don't care. They're either students, or they're just ripping off whomever gave them the money for the research.


<http://experimentalturk.files.wordpress.com/2012/10/paolacci-acr2012.pdf>

AMT cross-talk

- Mturk forums:
 - turkernation.com
 - mturkforums.com
 - turkoption.com
- On these forums turkers:
 - talk about HITs
 - talk about Requesters
 - link to HITs

09-06-2012 01:25 AM # 10

wendi5000 ◉
Member
★



Join Date: May 2012
Gender: ♀
Location: Houston, TX
Posts: 259

Not really funny but a good heads-up. A new thing I have seen in a couple of surveys today is special instructions about only copying part of the "code" at the end. For example, the last one I did said this:

On the next page, you will see a complete code which will consist of three letters, two numbers, and then the letter "x" followed by three additional numbers. In order to receive credit on Mechanical Turk, copy only the three letters and two numbers that precede the "x". If you copy the three numbers that follow the "x", we will assume that you did not follow our directions and you will not be paid.

*First they came for the Communists, and I didn't speak up, because I wasn't a Communist.
Then they came for the Jews, and I didn't speak up, because I wasn't a Jew.
Then they came for the Catholics, and I didn't speak up, because I was a Protestant.
Then they came for me, and by that time there was no one left to speak up for me.*

<http://experimentalturk.files.wordpress.com/2012/10/paolacci-acr2012.pdf>

Satisficing and Task Complexity

- **Satisficing:**
 - Satisficers perform tasks strategically so as to minimize effort without running the risk of being caught cheating
 - Advice: Do your own survey before posting!
- How high(low) is the **task complexity** ceiling?
 - Word sense disambiguation:
 - Snow et al. (2008): Words had on average 3 senses. AMT workers were as good as trained annotators;
 - Bardwaj et al. (2010): Words had on average 9.5 sense. AMT workers were worse than trained annotators.
 - How long can successful experiments be?
 - Ipeirotis (2010): submitting large jobs in MTurk results in low quality and unpredictable completion time

Validation of AMT for Linguistic Tasks

Sprouse (2011):

- Compared 176 laboratory participants with 176 AMT workers
- 24 sentence types (in 16 lexicalizations; Latin Square)
- Magnitude Estimation Task Judging acceptability of
 - (1) Who said my brother was kept tabs on by the FBI? baseline
 - (2) What did Lisa meet the man that bought? target
- Whether-Islands, CNP-islands, Subject Islands, Adjunct Island, Center Embedding Illusion, Comparative Illusion, Agreement Attraction Illusion
- Results: comparable data across the two populations (no meaningful difference in participant rejection, statistical power (except for comparative illusion?), shape of distribution)

Validation of AMT for Linguistic Tasks

Munro et al. (2010)

- Compared data from 7 different linguistic tasks to published data
- Found no loss of quality in AMT generated data
 - Semantic Transparency rating (cool down – cool, give up – give)
 - Segmentation of auditory speech (artificial language learning) into words
 - CLOZE tasks
 - Acceptability of ‘that’ omission in complement clauses and relative clauses
 - Metaphorical frequency
 - Corpus frequency of Agent embedding interpretations
 - (3) The investigation of the police took 3 weeks to complete
 - (4) It took 3 weeks to complete the investigation of the police

Things to consider when using AMT

- What is the theoretical question that you are trying to address?
- Why is it important to get data from many speakers?
- What is the task?
- Does your theory make different predictions when you consider potential confounds stemming from the task?
 - Spell out predictions quantitatively
- Do your informants need to be trained on the task?
 - Practice Items
 - Task description
- What are possible satisficing strategies?
 - Filler Items
 - Catch items
- Ethical considerations