

# **Turkshop – week 1**

Experimental Syntax-Semantics lab meeting  
February 28, 2013

Hadas Kotek  
Martin Hackl

# Today..

- Empirical methods in linguistic research.
  - Using Amazon's mechanical Turk
  - Experiments run on Turk in the ESSL
  - Designing a simple Turk experiment
- 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 process of uploading an experiment to Turk; do it yourself.

# Materials

- Readings and materials on ESSL website:

<http://web.mit.edu/hackl/www/lab/turkshop/>

- Download and unzip “week1-blocking” in examples.

# Getting started

- Why/when AMT?
- Using experimental methods in linguistics
- Using Amazon's Mechanical Turk to in language studies.
  - Previous studies done in the ESSL.
    - Acceptability ratings
    - Sentence completion
    - Verification,
    - Sentence picture matching

# Why/when AMT?

Q: Why is it better to gather data from hundreds potentially very noisy informants in a potentially very noisy environment than from a handful of trusted and cooperative informants in your office?

- A1: Representativeness across speakers – but why?
  - Observer biases
  - Informants are Non-representative sample
- A2: Representativeness across items
- Noise
  - The experimental task involves engaging noisy cognitive systems (memory, etc.)
  - The phenomenon of interest itself is noisy

# Single informant vs. AMT crowd

- What makes a traditional, single informant data gathering session successful?
- Good informant
  - Cooperative
  - Teachable
  - Constructive feedback, insights
- Good items
  - Appropriate context
  - Representative items
- Suitable task
  - Training
- Conducive environment

# Getting started

- Design considerations
- An example experiment
- Structure of files
- Uploading experiment to Turk

## Simple Designs: 2×1

- In case we are simply interested in the presence or absence of one property of a sentence on its grammaticality.
  - Effect of animacy on grammaticality of English ‘make’-causative.
- (1) a. The coach made the ball bounce on the floor.  
b. The coach made the gymnast bounce on the floor.



## Simple Designs: 2×1

- Presence or absence of Complex NP island:
  - (2) a. What did you claim that John bought?
  - b. \*What did you make the claim that John bought?

## 2×2 Designs

- Example: “week1-blocking” (based on Kotek&Erlewine, in prep.; for more on the theory come to our Ling Lunch on 3/14!).
  - Effect of animacy of the causee in two kinds of causative constructions: Animacy × Type.
    - Animate vs. inanimate
    - Lexical causative vs. ‘make’ causative (*v* vs. *make-v*)
- (3) a. That’s the ball that the coach *bounced* on the floor.  
b. That’s the gymnast that the coach *bounced* on the floor.  
c. That’s the ball that the coach *made bounce* on the floor.  
d. That’s the gymnast that the coach *made bounce* on the floor.

## 2×2 Designs and Latin Square

- (3) a. That's the ball that the coach *bounced* on the floor.  
b. That's the gymnast that the coach *bounced* on the floor.  
c. That's the ball that the coach *made bounce* on the floor.  
d. That's the gymnast that the coach *made bounce* on the floor.
- We want each participant to only rate *one* of these conditions for a given item.
  - **Latin square design:** cycles through our items to create lists.
    - List 1: (1a), (2b), (3c), (4d), (5a), ...
    - List 2: (1b), (2c), (3d), (4a), (5b), ...
    - ... (for 4 conditions, we need a multiple of 4 lists)

A	B	C
C	A	B
B	C	A

## 2×2 Designs

- Designs more complicated than 2×2 are not recommended because they are hard to analyze and interpret.
- To check for an effect, we create multiple items that manipulate the factors of interest.
  - Normally: 10-12.
  - (but sometimes more; the blocking study had 120 items, but the example today only has 5).

# Fillers

- Normally, at least as many as the targets
- Distract from the true purpose of the experiment
- Even out skewness in items
- Serve as an exclusion criterion
  - Overall accuracy
  - “catch items”

# The structure of an items file

# experiment item-number condition

Fillers all have the same condition name

Item #1 in blocking experiment, has four conditions.

Item #2 repeats same four conditions.

```
1 # filler 1 filler
2 That's the girl that the teacher played with the doll.
3
4 # filler 2 filler
5 That's the student that the councilor made skip the road.
6
7 # blocking 1 inanimate-make-v
8 That's the book that the girl made balance on her head.
9
10 # blocking 1 inanimate-v
11 That's the book that the girl balanced on her head.
12
13 # blocking 1 animate-make-v
14 That's the gymnast that the coach made balance on the beam.
15
16 # blocking 1 animate-v
17 That's the gymnast that the coach balanced on the beam.
18
19 # blocking 2 inanimate-make-v
20 That's the ladder that the worker made lean against the door.
21
22 # blocking 2 inanimate-v
23 That's the ladder that the worker leaned against the wall.
24
25 # blocking 2 animate-make-v
26 That's the gang that the officer made lean against the wall.
27
28 # blocking 2 animate-v
29 That's the gang that the officer leaned against the wall.
```

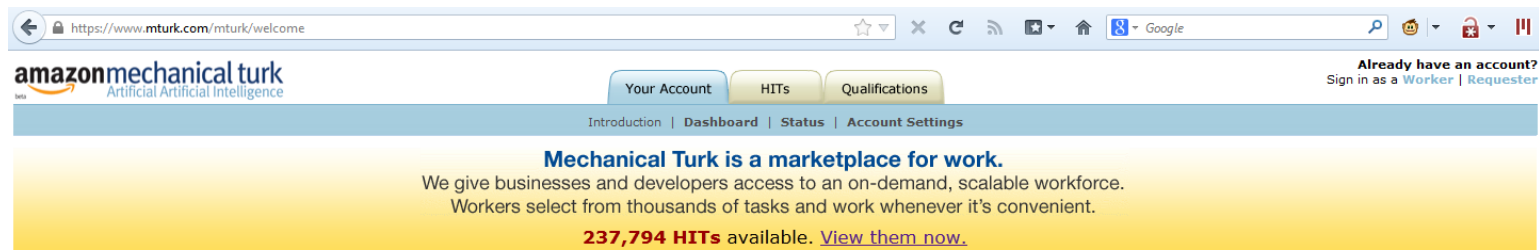
# The magic of the Turkolizer

- The Turkolizer takes the items file
  - Creates a file that can be uploaded to Turk
  - Randomizes the items
  - Creates lists
  - Credit for the Turkolizer: Gibson et al. (2011)
  - mitcho's tweak creates the reverse of each list, fixes several issues (more in two weeks).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	list	trial_1	trial_2	trial_3	trial_4	trial_5	trial_6	trial_7	trial_8	trial_9	trial_10					
2	0	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the gang that the officer leaned against the wall.					
3	1	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the rabbit that the magician made vanish into thin air.					
4	2	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the boy that the instructor floated in the pool.					
5	3	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the rabbit that the magician made vanish into thin air.					
6	4	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the book that the girl made balance on her head.					
7	5	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the student that the councilor made skip the road.					
8	6	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the book that the student returned to the library.					
9	7	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the	That's the film that the assistant developed in the darkroom.					

# Amazon's Mechanical Turk

- [www.mturk.com](http://www.mturk.com)
- Sign in as requester
  - [mit.turkshop@gmail.com](mailto:mit.turkshop@gmail.com)
  - Building20



## 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



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:

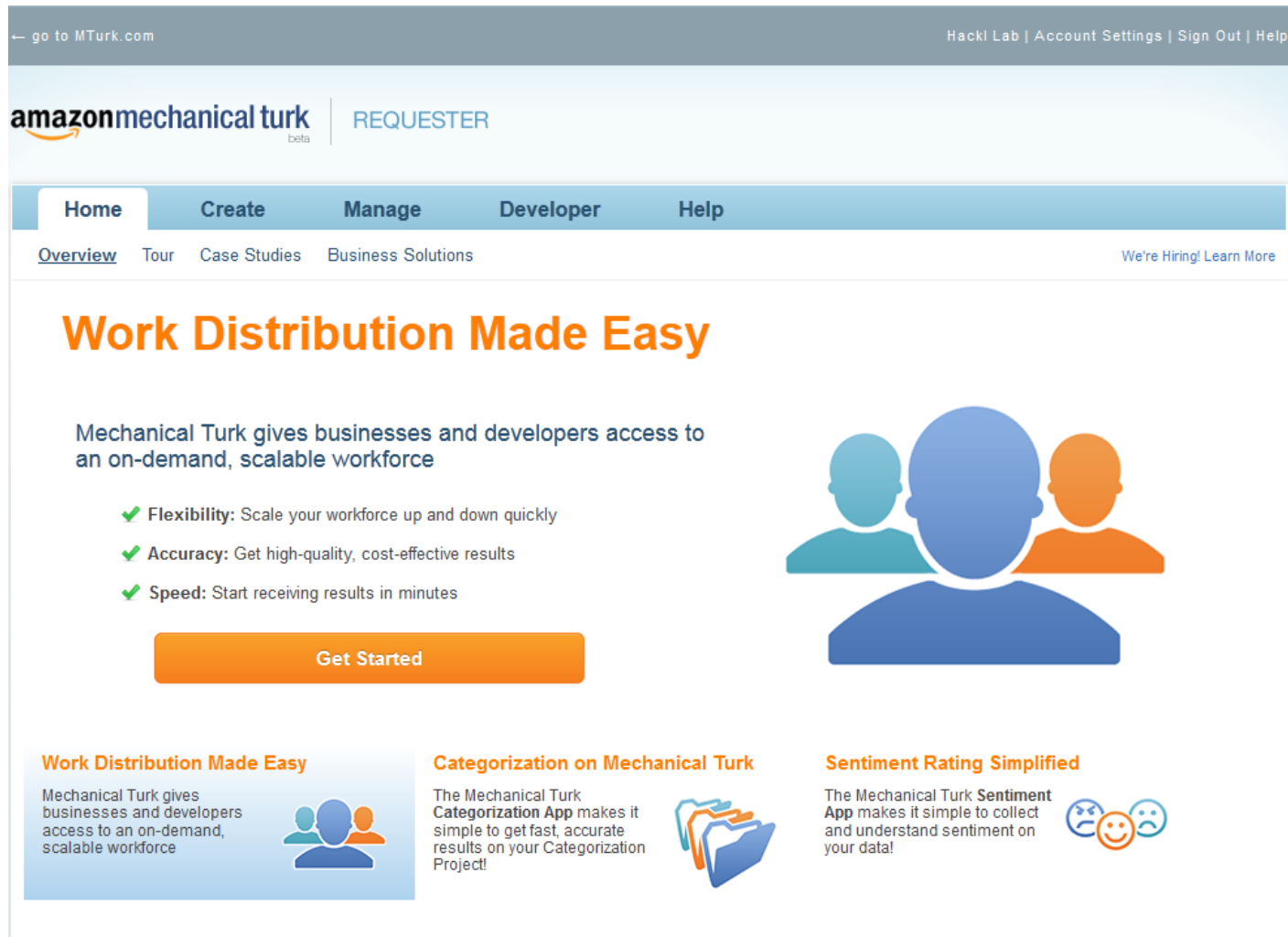
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results





# The home screen

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



# The *Create* tab

- Create a new project here.

← go to MTurk.com HackI Lab | Account Settings | Sign Out | Help

amazonmechanicalturk beta REQUESTER

Home Create Manage Developer Help

New Project [New Batch with an Existing Project](#) Create HITs individually

Start a New Batch with an Existing Project

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

# The *Create* tab

- Each experiment that is set up requires two steps:
  - Set up a template for your experiment:
    - Name and description
    - Worker (=participant) qualifications
    - Payment
    - Number of workers
    - HTML template to present items, including instructions and practice items.
  - Upload a turkolized items file and publish a “batch.”
- For Turkshop purposes: put your name as part of the project name!

# Creating a new project: Specifications

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

amazonmechanical turk beta 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: Turkshop blocking example 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

- We normally pay \$.10-.15 for a short survey, \$.2-.3 for longer surveys.
- Calculate how many observations per item you would like to have.
  - We've 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
  - We have 5 items in our 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 \$6.00 hourly wage.

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

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

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

Results are automatically approved in    
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)

Advanced

Worker requirements «

Worker requirements:  
Customize Worker Requirements...

Specify ALL the qualifications Workers must meet to work on your HITs:

HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 95 remove

Location is UNITED STATES remove

(+) Add another criterion (up to 5)

Only Workers who qualify to do my HITs can preview my HITs.

☒ Yes ☐ No

# Creating a new project: Template

# Creating a new project: Template

- Enter HTML template here.
  - HTML template is created automatically via mitcho's script.
  - Open template in text editor, copy and paste in the window below.

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

U

I

B

T

↶

↷

↺

↻

↵

↶

↷

↺

↻

↵

Edit HTML Source

```
<p>Survey Code: <font color="blue">BIR</font></p>
<p><strong><font color="red">PLEASE COMPLETE AT MOST ONE </font><font color="blue" class="Apple-style-span">BIR</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
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:</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 carpenter opened a hole</td>
```



# Creating a new project: Template

- Contents of the template that you do need to edit:
  - Instructions
  - Practice items
  - Your contact information
- In addition: don't forget the consent statement!

# Creating a new project: Finish

## 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: Turkshop sample

This name is not displayed to Workers.

Sentence judgement task

Requester: Hackl Lab

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: Finish

- Re-read your instructions.
  - Make sure your examples are clear for someone who is not reading the text carefully (i.e., most users).
  - Verify consent statement is there.
- Put an email address that you use for communication purposes.
- Check the consent statement.
- Check that there are many items as the text says there are (and you think there are).
- Verify that the counter at the bottom of the page works properly.

# Creating a new project: Finish

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](mailto: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.

$\S\{\text{trial\_1}\}$

NATURAL

☐

UNNATURAL

☐

2.

$\S\{\text{trial\_2}\}$

NATURAL

☐

UNNATURAL

☐

3.

$\S\{\text{trial\_3}\}$

NATURAL

☐

UNNATURAL

☐

# Creating a new project: Finish

HIT Preview

---

\$\{trial\\_9\}\$

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

---

10.

\$\{trial\\_10\}\$

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

---

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

# Posting your items on Turk: New batch

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

amazonmechanical turk beta | REQUESTER

Home Create Manage Developer Help

New Project New Batch with an Existing Project Create HITs individually

Start a New Batch with an Existing Project

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

Leave feedback for this page.

# Posting your items on Turk: New batch

The screenshot shows the Amazon Mechanical Turk requester interface. At the top, there's a navigation bar with 'go to MTurk.com' and links for 'MIT Turkshop | Account Settings | Sign Out | Help'. Below this, the 'amazonmechanical turk' logo and 'REQUESTER' label are visible. The main navigation has 'Home' and 'Create' tabs. Under 'Create', there are links for 'New Project' and 'New Batch with an Existing Project'. A modal dialog titled 'New Batch' is open, prompting the user to 'Choose a .csv file with the variables you specified in your project.' It features a 'Choose File' button, a text input field containing 'blocking items.turk.csv', and an 'Upload' button. In the background, a table titled 'Start a New Batch with an Existing Project' lists projects. The first project is 'Turkshop sample: 5-item blocking' with the title 'Sentence judgement task' and creation date 'February 25, 2013'. It has buttons for 'New Batch', 'Edit', 'Copy', and 'Delete'. A link 'Create HITs individually' is also present. At the bottom right, there is a link to 'Leave feedback for this page.'

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

amazonmechanical turk beta REQUESTER

Home Create

New Project New Batch with an Existing Project

**New Batch** [X]

Choose a .csv file with the variables you specified in your project.

Choose File blocking items.turk.csv Upload

Create HITs individually

Start a New Batch with an Existing Project

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

Leave feedback for this page.

# Posting your items on Turk: Preview HITs

[Home](#) [Create](#) [Manage](#) [Developer](#) [Help](#)

[New Project](#) [New Batch with an Existing Project](#) [Create HITs individually](#)

## Preview HITs

1 Select HIT Template 2 Upload Input Data 3 Preview 4 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

Grammaticality Judgement Task

Requester: Hackl Lab

Reward: \$0.1 per HIT

HITs available: 8

Duration: 1 Hours

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: Preview HITs

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: Turkshop blocking example 3

Description: 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	(total number of assignments in this batch)
Estimated Total Reward:	\$2.400	
Estimated Fees to Mechanical Turk:	+ \$0.240	(fees paid to Mechanical Turk) (fee details)
Estimated Total Cost:	\$2.640	(this is the amount that will be deducted from your Available Balance when you click "Publish HITs")
<hr/>		
Your Available Balance:	\$300.000	(before clicking "Publish HITs")
Your Projected Balance:	\$297.360	(after clicking "Publish HITs")

Back

Publish HITs

# Posting your items on Turk

- 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.

# The *Manage* tab

[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% submitted100% published

▼ Batches ready for review (25)

[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% submitted100% published

« Previous 1 2 3 Next »

# 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"/>	24ZMVHVKCJ019YNOJUY9PQW7HJ475Z	A1DKEJ4G4QQD10	0% (0/0)	4	That's the student that the councilor 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 i...	That's the girl that the teacher played with th...	That's the book that the girl made balance on h...
<input type="checkbox"/>	2BC00OQKKA4IQ6T9H5T6D0EXGNGYY6	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 girl that the teacher played with th...	That's the idea that the team made float at the...	That's the wall that the carpenter broke into s...	That's the engine that the mechanic ran idle fo...	That's the film that the assistant developed in...	That's the book that the student made return to...	That's the student that the councilor made skip...	That's the gang that the officer leaned against...
<input type="checkbox"/>	2BC00OQKKA4IQ6T9H5T6D0EXGNGYY6	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 girl that the teacher played with th...	That's the idea that the team made float at the...	That's the wall that the carpenter broke into s...	That's the engine that the mechanic ran idle fo...	That's the film that the assistant developed in...	That's the book that the student made return to...	That's the student that the councilor made skip...	That's the gang that the officer leaned against...

# The *Manage* tab

- We reject subjects who did not comply with instructions.
  - Completed more than one survey, if we asked for just one (often a requirement for language studies!).
  - Did not complete experiment.
- Possibly also:
  - Failed on ‘catch’ items.
  - Exhibit guessing behavior.
- But **NOT**: Non-native speakers!
  - That may bias participants into saying they are native speakers when in fact they are not.

# What it looks like from the workers' end

amazonmechanical turk Artificial Intelligence

Your Account | **HITS** | Qualifications | 240,066 HITS available now

Hadas Kotek | Account Settings | Sign Out | Help

All HITS | HITS Available To You | HITS Assigned To You

Find  containing  that pay at least \$  ☐ for which you are qualified ☐ require Master Qualification

## All HITS

1-10 of 3069 Results

Sort by:

[Show all details](#) | [Hide all details](#)

1 2 3 4 5 > [Next](#) >> [Last](#)

Inv. B. 2 <a href="#">Request Qualification (Why?)</a>   <a href="#">View a HIT in this group</a>			
Requester: <a href="#">rohzi0d</a>	HIT Expiration Date: Mar 16, 2013 (2 weeks 4 days)	Reward: \$0.00	
	Time Allotted: 48 minutes	HITS Available: 15308	
Search: Keywords on Google.com (US) <a href="#">View a HIT in this group</a>			
Requester: <a href="#">CrowdSource</a>	HIT Expiration Date: Feb 25, 2014 (52 weeks)	Reward: \$0.08	
	Time Allotted: 16 minutes	HITS Available: 12455	
Copy Text from Business Card <a href="#">View a HIT in this group</a>			
Requester: <a href="#">Oscar Smith</a>	HIT Expiration Date: Feb 25, 2013 (1 hour 59 minutes)	Reward: \$0.02	
	Time Allotted: 10 minutes	HITS Available: 12109	
Collect pricing information for parts <a href="#">Request Qualification (Why?)</a>   <a href="#">View a HIT in this group</a>			
Requester: <a href="#">Jeffrey Werby</a>	HIT Expiration Date: Feb 26, 2013 (15 hours 15 minutes)	Reward: \$0.05	
	Time Allotted: 10 minutes	HITS Available: 7630	
Check If a Tweet Violates Twitter's Ads Policy (CWR ID: 1053) <a href="#">Request Qualification (Why?)</a>   <a href="#">View a HIT in this group</a>			
Requester: <a href="#">Jack Stone</a>	HIT Expiration Date: Mar 1, 2013 (3 days 11 hours)	Reward: \$0.15	
	Time Allotted: 15 minutes	HITS Available: 7527	
Copy Text from Business Card <a href="#">View a HIT in this group</a>			
Requester: <a href="#">Oscar Smith</a>	HIT Expiration Date: Feb 25, 2013 (1 hour 59 minutes)	Reward: \$0.02	
	Time Allotted: 10 minutes	HITS Available: 7073	
Find company contact information Not Qualified to work on this HIT <a href="#">(Why?)</a>   <a href="#">View a HIT in this group</a>			
Requester: <a href="#">Readabl</a>	HIT Expiration Date: Mar 4, 2013 (6 days 23 hours)	Reward: \$0.05	

# More on AMT

- 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/>



# Demographics, Economics, Ethics, etc.

- **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 differently 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, Ethics, etc.

- 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
  - They train their workforce
  - They pay minimum wage
- Implications of labor costs for research/funding expectation etc.

## Naïveté, Cheating, Satisficing, Task Complexity, etc.

- **Naïveté:** It is often said that gathering data from naïve participants is important to ensure getting unbiased data (e.g. Gibson et al. 2010, etc.)
- 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

# A worker on an AMT forum

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.




# AMT Crosstalk

- Mturk forums:
  - turkernation.com
  - mturkforums.com
  - turkopticon.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

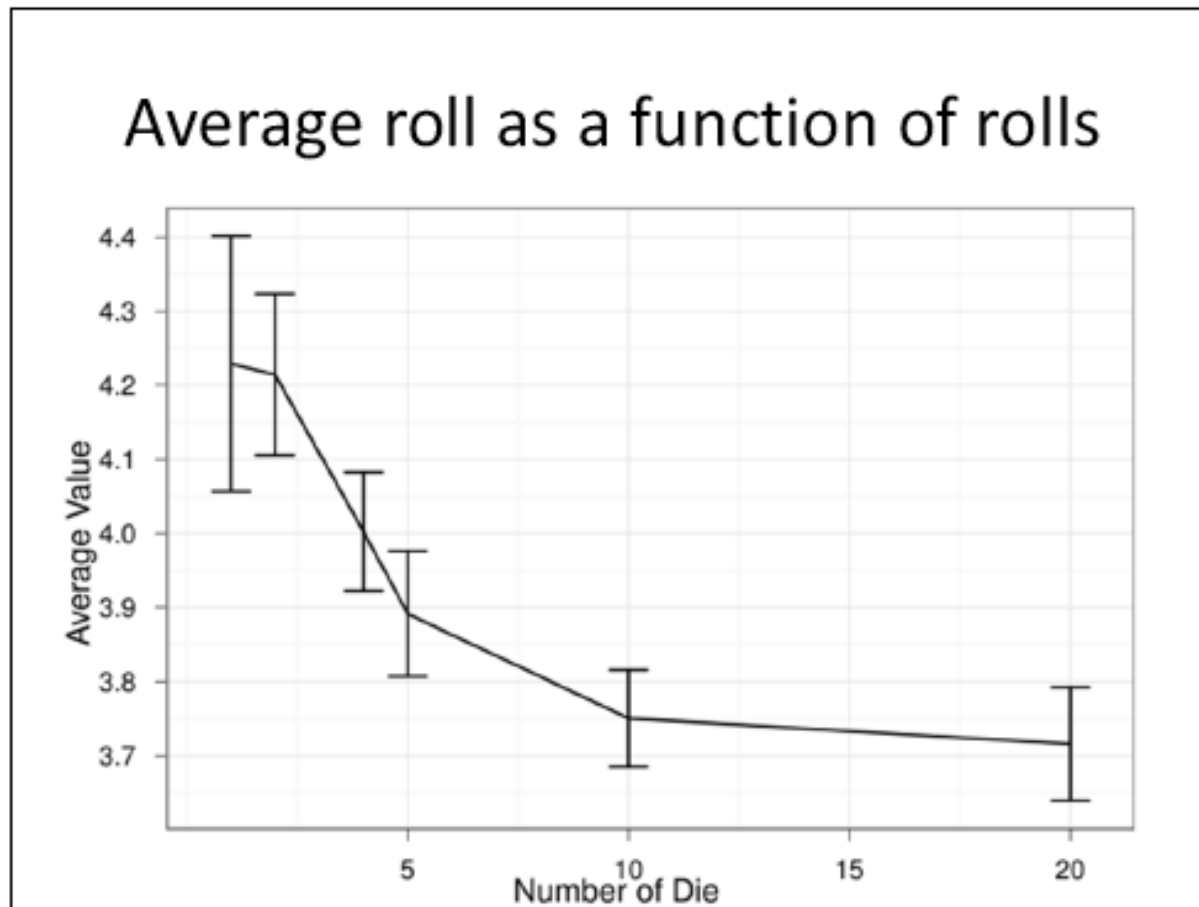
Not really funny but a good heads-up. A new thing I have seen in a couple of surveys today is special instructions about only copying part of the "code" at the end. For example, the last one I did said this:

*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.*

# Naïveté, Cheating, Satisficing, Task Complexity, etc.

- **Cheating**



# Naïveté, Satisficing, Task Complexity Ceiling, etc.

- 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 are as good as trained annotators;
    - Bardwaj et al. (2010): Words had, on average, 9.5 sense AMT workers are 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

Standard:        Who said my brother was kept tabs on by the FBI? 100 Item:  
                      What did Lisa meet the man that bought? \_\_\_\_\_

- Analyzed: Whether-Islands, CNP-islands, Subject Islands, Adjunct Island, Center Embedding Illusion, Comparative Illusion, Agreement Attraction Illusion
- Sprouse (2011) finds 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 (cont.)

- Munro et al. (2010)
  - Compared data from 7 different linguistic tasks to published data and find 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
    - Corpus frequency of Agent embedding interpretations
  - (1) The investigation of the police took 3 weeks to complete
  - (2) It took 3 weeks to complete the investigation of the police
    - Metaphorical frequency

# Things to consider when using AMT

- What is the theoretical question you are trying to address?
- Why is it important to get data from many speakers?
- What is the task?
- Do your theories make different predictions when you consider possible confounds coming 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

