

# Getting Started with MTurk Data Collection

## Tutorial Slides

Nandana Sengupta, Postdoc @ Knowledge Lab

March 2, 2017

# Introduction

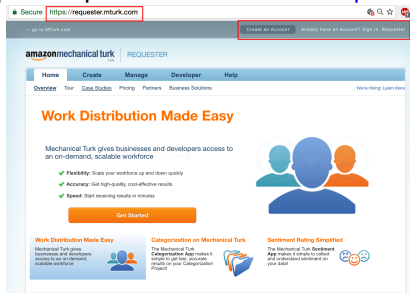
- ▶ What is M(echanical) Turk?
  - ▶ MTurk is a crowdsourcing internet marketplace. *Requesters* get access to an on-demand workforce (*workers*) to perform surveys and tasks which computers are unable to do.
- ▶ Is this tutorial for Workers or Requesters?
  - ▶ This tutorial is for MTurk Requesters who want to collect and analyze crowdsourced data.
- ▶ What is a HIT?
  - ▶ HIT stands for Human Intelligence Task – these are the tasks requesters design and workers complete.
- ▶ In this Tutorial:
  - ▶ Step by Step MTurk Requester Guide.
  - ▶ Some things to keep in mind while using MTurk.
  - ▶ Appendix: Simple HTML code for MTurk.

# Step by Step MTurk Tutorial

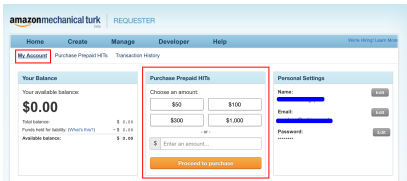
- ▶ **Step 1:** Setup a Requester Account
- ▶ **Step 2:** Purchase Prepaid HITS
- ▶ **Step 3:** Create a new survey
  - ▶ **Step 3a:** HIT properties
  - ▶ **Step 3b:** Design Layout
  - ▶ **Step 3c:** Preview and Finish
- ▶ **Step 4:** Publish HIT
- ▶ **Step 5:** Track Progress and Manage Results
- ▶ **Step 6:** Approve Workers and Download Results

# Step by Step MTurk Tutorial

- **Step 1:** Setup a Requester Account on [requester.mturk.com](https://requester.mturk.com).



- **Step 2:** Follow instructions on the 'My Account' tab to Purchase Prepaid HITS.



# Step by Step MTurk Tutorial

- ▶ **Step 3:** For creating a new HIT click on the 'Create' tab. MTurk provides three types of options:
  - ▶ **External Survey Link** (direct to an app or SurveyMonkey, Qualtrics survey).
  - ▶ Customizable **in-built** HIT options.
  - ▶ Customized HITS by changing **HTML source code**.

The screenshot shows the Amazon Mechanical Turk Requester interface. The 'Create' tab is highlighted with a red box. Below the navigation bar, the 'Start a New Project' section is visible. A red arrow points to the 'Survey Link' option in the left sidebar. Another red arrow points to the 'Other' option at the bottom of the sidebar. A green arrow points to the 'Example of Survey Link' section, which includes a 'Survey link' input field with the value 'http://example.com/survey345.html' and a 'Provide the survey code here:' input field with the value 'e.g. 123456'. A blue button labeled 'Create Project >' is visible next to the 'Example of Survey Link' section. A blue box with the text 'You must ACCEPT the HIT before you can submit the results.' is also present. Red text labels with arrows point to the 'Survey Link' option (labeled 'External survey link'), the 'Other' option (labeled 'In-built tasks'), and the 'Provide the survey code here:' input field (labeled 'Customized Survey').

# Step by Step MTurk Tutorial

- ▶ **Step 3a:** Once you choose the type of HIT to create, you will be directed to the 'Properties' tab to
  - ▶ provide instructions to workers, name the project etc;

The screenshot shows the 'Enter Properties' tab of the MTurk HIT creation process. It includes fields for 'Project Name' (Survey Link), 'Title' (Answer a survey about your opinions), 'Description' (Give us your opinion about our products), and 'Keywords' (survey, demographics). There are also instructions for each field, such as 'Describe the task to Workers' and 'Provide keywords that will help Workers search for your HITs'.

- ▶ assign rewards, select number of workers etc;

The screenshot shows the 'Setting up your HIT' tab. It includes fields for 'Reward per assignment' (\$), 'Number of assignments per HIT' (20), 'Time allotted per assignment' (1 hour), 'HIT expires in' (7 days), and 'Auto-approve and pay Workers in' (3 days). There are also instructions for each field, such as 'This is how much a Worker will be paid for completing an assignment' and 'Maximum time a Worker has to work on a single task'.

- ▶ specify qualifications (for an added cost).

The screenshot shows the 'Worker requirements' tab. It includes a section for 'Require that Workers be Masters to do your HITs' (Yes/No), a section for 'Specify any additional qualifications Workers must meet to work on your HITs' (Add another criterion), and a section for 'Project contains adult content' (Yes/No). There are also instructions for each field, such as 'Qualifying criteria, based on worker quality and/or demographic characteristics. Extra cost to requester'.

# Step by Step MTurk Tutorial

- ▶ **Step 3b:** The next tab is the **Design Layout**.
- ▶ For **external survey links** the main thing to keep in mind is that the external link provides the worker a survey code (preferably random) at the end of the task.
- ▶ With an external link MTurk cannot ensure that the worker completes the task. This is checked by the requester by matching the survey codes once the results are in.

The screenshot shows the MTurk Design Layout interface. At the top, there are three tabs: "Enter Properties", "Design Layout" (which is selected), and "Preview and Finish". Below the tabs, the "Project Name" is set to "Survey Link". A "Frame Height" of 450 is specified. The main content area contains a blue box with "Survey Link Instructions (Click to expand)". The instructions state: "We are conducting an academic survey about social networks. We need to understand your opinion about social networks. Select the link below to complete the survey. At the end of the survey, you will receive a code to paste into the box below to receive credit for taking our survey. Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box." Below this, a yellow box contains a "Template note for Requesters": "To verify that Workers actually complete your survey, require each Worker to enter a unique survey completion code to your HIT. Consult with your survey service provider on how to generate this code at the end of your survey." At the bottom, there are two input fields. The first is labeled "Survey link:" and contains the text "http://example.com/survey/345.html". The second is labeled "Provide the survey code here:" and contains the text "e.g. 123456". Red boxes highlight these two input fields. To the right of the first field, the text "Your survey link" is written. Below the second field, the text "Important. Needed to confirm that worker completed the task. Will be important when approving results." is written.

1 Enter Properties 2 Design Layout 3 Preview and Finish

Project Name: Survey Link This name is not displayed to Workers.

Frame Height: 450 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 Bold Italic Underline Link Image Video Audio Text Source

**Survey Link Instructions (Click to expand)**

We are conducting an academic survey about social networks. We need to understand your opinion about social networks. Select the link below to complete the survey. At the end of the survey, you will receive a code to paste into the box below to receive credit for taking our survey.

Make sure to leave this window open as you complete the survey. When you are finished, you will return to this page to paste the code into the box.

**Template note for Requesters** - To verify that Workers actually complete your survey, require each Worker to enter a unique survey completion code to your HIT. Consult with your survey service provider on how to generate this code at the end of your survey.

Survey link:  Your survey link

Provide the survey code here:

Important. Needed to confirm that worker completed the task. Will be important when approving results.

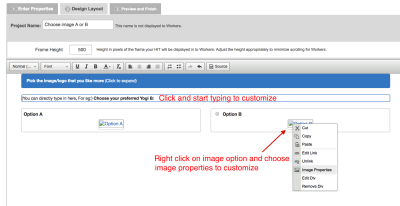
body

# Step by Step MTurk Tutorial

- ▶ **Step 3b: Inbuilt HITs** are customizable in the Design Layout tab without going into source code.
- ▶ These customizations are mostly straightforward. The screenshots below provide an example using the inbuilt HIT '**Choose Image A or B**'.

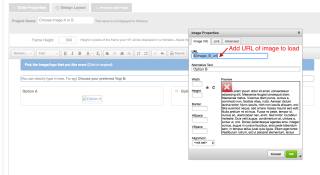
## Edit Project

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 (e.g., main menu).



## 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 (e.g., main menu).





# Step by Step MTurk Tutorial

- ▶ **Step 3b: Inbuilt HITs** are customizable in the Design Layout tab without going into source code.
- ▶ These customizations are mostly straightforward. Another example is **multiple choice** questions in the inbuilt HIT 'Survey' very easily customizable within the Design Layout tab, screenshot provided below.

[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](#)).

The screenshot shows the MTurk Design Layout editor interface. At the top, there are three tabs: '1 Enter Properties', '2 Design Layout' (which is active), and '3 Preview and Finish'. Below the tabs, the 'Project Name' is set to 'Survey'. A 'Frame Height' of 600 is specified. A toolbar with various formatting and editing tools is visible, with the 'Source' button highlighted by a red box. The main content area is divided into two sections. The first section, titled 'Survey Instructions (Click to expand)', contains a text input field for instructions and a list of three instructions: 'Instruction #1', 'Instruction #2', and 'Instruction #3'. A red box highlights the instruction list, and a red text annotation says 'Just click in this section and type in any custom instructions'. The second section, titled '1. What is your gender?', contains a radio button form with three options: 'Female', 'Male', and 'Prefer not to say'. A red box highlights the form, and a red text annotation says 'Just click in this section and type in custom questions and choices.' Below the form, there is additional text: 'To add more choices and change variable names, click the Source button highlighted above (details in Appendix on HTML code Page 19)'.

Project Name: Survey

Frame Height: 600

Normal (Font) [U] [I] [B] [A] [I] [List] [Table] [Link] [Source]

**Survey Instructions (Click to expand)**

Enter instructions for your survey:

- Instruction #1
- Instruction #2
- Instruction #3

Just click in this section and type in any custom instructions

**1. What is your gender?**

☐ Female

☐ Male

☐ Prefer not to say

Just click in this section and type in custom questions and choices.

To add more choices and change variable names, click the Source button highlighted above (details in Appendix on HTML code Page 19)

# Step by Step MTurk Tutorial

- **Step 3c:** Once you have finalized the HIT properties and Design Layout, the last tab requires you to '**Preview and Finish**' (indicating you are happy with the HIT you designed).

1 Enter Properties 2 Design Layout 3 Preview and Finish

Project Name: Choose image A or B This name is not displayed to Workers.

Choose between image A or B

Requester: [redacted] Reward: \$0.10 per HIT HITs available: 0 Duration: 1 Hours

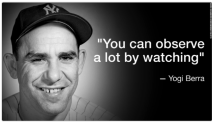
Qualifications Required: [redacted]

HIT Preview


Pick the image/logo that you like more (Click to expand)

Choose your preferred Yogi B:

Option A



Option B



Submit

Finish

# Step by Step MTurk Tutorial

- **Step 3c:** At this stage, some inbuilt HITs allow you to download a sample csv file. For bulk tasks (like comparing large number of images) you may prepare your input data based on the fields in the downloaded csv.

Your project was successfully saved.

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

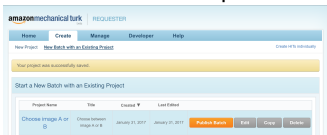
Project Name: Choose image A or B This name is not displayed to Workers

	A	B
1	image_A_url	image_B_url
2	Hit1_image_A_url_data	Hit1_image_B_url_data
3	Hit2_image_A_url_data	Hit2_image_B_url_data
4	Hit3_image_A_url_data	Hit3_image_B_url_data
5		
6	In this example you have a choice of adding N pairs of image	
7	urls for the image comparison task	
8		
9		
10		

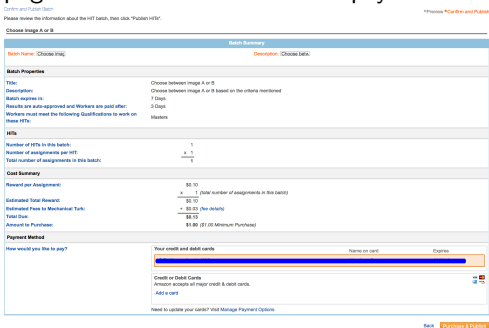
- To **change variable names**, you will need to go into the **source code** in the **Design Layout** tab. This is explained in detail in the Appendix on Page 19.

# Step by Step MTurk Tutorial

- **Step 4:** Once you have created the HIT it shows up under the 'Create' Tab with options to **Publish, Edit, Copy or Delete**.



- On selecting '**Publish Batch**', you will be directed to a payments page where the fee details and payment details will be confirmed.



- Click '**Purchase and Publish**' to make HIT available to workers.

# Step by Step MTurk Tutorial

- ▶ **Step 5:** Once the HIT is live you can track its progress under the 'Manage' Tab.
- ▶ This is also where completed batches of HITs (i.e. Batches for Review) can be accessed.

The screenshot shows the Amazon Mechanical Turk Requester interface. The top navigation bar includes 'Home', 'Create', 'Manage' (highlighted with a red circle), 'Developer', and 'Help'. Below this, there are tabs for 'Results', 'Workers', and 'Qualification Types'. The main content area is titled 'Manage Batches' and includes a link to 'Manage HITs individually'. A sub-header says 'Click on the name of the batch to see more details'. There are two sections: 'Batches in progress (0)' and 'Batches ready for review (5)'. A red arrow points from the text 'Go here to track progress, track average time, cancel a batch if required' to the 'Batches in progress (0)' section. Another red arrow points from the text 'Click here to download results and approve workers' to the 'Results' button in the batch details section. The batch details for 'Batch 4 Street Safety Perception Rollout Jan 2' are shown, including 'Created: January 18, 2017', 'Time Elapsed: 3 days', 'Assignments Completed: 65 / 65', and 'Estimated Completion Time: COMPLETE'. A progress bar shows '100% submitted' and '100% published'.

- ▶ **Step 6:** Once the HIT batch is completed it is ready for review. Click on 'Results' to Approve or Reject workers, Download CSV files etc. This can also be done before the batch is completed.

# Step by Step MTurk Tutorial

## Your crowdsourced data is now available for analysis!!!

[Manage Batches](#) > [Review Results](#)

### 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.

~~Protect Sensitive Information~~

[Customize View](#) [Filter Results](#) [Upload CSV](#) [Approve All](#) [Download CSV](#)

10 of 10 assignments (FILTER APPLIED: only show assignments that are in 'Approved' status)

<div>Approve</div> <div>Reject</div>						
		HIT ID ▲	Worker ID	Lifetime Approval Rate	Tag4	Surveycode
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (1/1)	yes	xWoshbOy1HAXrqQ0kC6PynQ2YR9VEF
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (2/2)	yes	gitAi0SWP2G1TqeJyZr46N4VKgi4ME
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (1/1)	yes	IG7ZH6t3rrOwo77LP8r5pq0m6jvsai
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (1/1)	yes	6eZq4ylCB2ladieVaTVoLJXVrQ0KOt
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (1/1)	yes	7H2fkknL54wnukvTdVSNsl7TchgD1y
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (3/3)	yes	to6W7NzHoFYWksIDHe4VuOieTqaN6Z
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (2/2)	yes	XQVDhQMLZN9cGglekNHv2Qt0xTvvwv
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (2/2)	yes	IT0kl943Pwl5ffura3IREoqwbWO68m
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (2/2)	yes	7FUD2mSWNZi7qBUNK7rx5TOknxqown
		3FSEU3P2NRKFR7SBHP2C5FRR2GPRRY	A186BMA578K99A	100% (1/1)	yes	YR.JjphLymfMPvIQnuX5Jx0Xms3BsuT
		HIT ID ▲	Worker ID	Lifetime Approval Rate	Tag4	Surveycode
<div>Approve</div> <div>Reject</div>						

# Some things to keep in mind when using MTurk

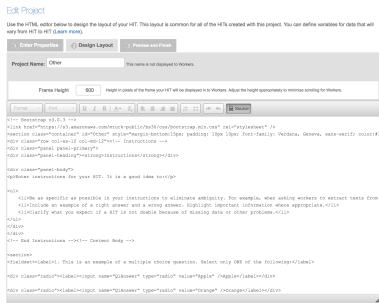
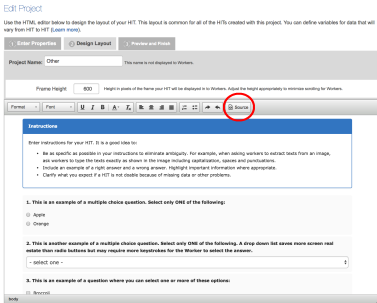
- ▶ For those interested in trying out different MTurk functionalities without conducting actual experiments, a useful site to checkout is [requestersandbox.mturk.com](http://requestersandbox.mturk.com).
- ▶ Make sure your HIT is well-designed and not buggy. Also make sure compensation is fair. There are a number of MTurk worker discussion boards where **you don't want a bad reputation as a requester!**
- ▶ While crowdsourcing is an efficient and cost-effective way of collecting data, it may not be suitable for all research questions. For instance, there is a large chance of selection bias in the data (i.e. the workers are not randomly distributed across geographies, ages, income etc.). **Make sure crowdsourced data collection is appropriate for your research question.**
- ▶ Finally, here is an interesting [Pew Center report](#) on Crowdsourcing and MTurk if you are interested in knowing more about the profile of MTurk Workers, Requesters and Tasks.

## Appendix: Simple HTML code for MTurk



## Simple HTML code for MTurk

- 



- ▶ In the next few slides we'll go over some simple HTML code to help you customize your HITs.

# Simple HTML code for MTurk

## HTML formatting basics

- ▶ Basic formatting commands include a prefix and a suffix, for eg:
  - ▶ `< p > ... < /p >`: paragraph,
  - ▶ `< br >`: line break,
  - ▶ `< b > ... < /b >`: bold font,
  - ▶ `< u > ... < /u >`: underlined font,
  - ▶ `< i > ... < /i >`: italic font.
- ▶ You can learn more basics [here](#).
- ▶ An example:

### Source Code

```
<!DOCTYPE html>
<html>
<body>

<h1>Use "h1" Tag For Headings</h1>

<p>Use "p" tag to separate paragraphs (i.e. add a line
at the end of the paragraph).</p>

<p> <u> The "u" tag is used to underline </u>, whereas
<b> the "b" tag is used for bold letters </b>, and <i>
the "i" tag is used for italics. </i> </p>

</body>
</html>
```

### Webpage Output

## Use "h1" Tag For Headings

Use "p" tag to separate paragraphs (i.e. add a line at the end of the paragraph).

The "u" tag is used to underline, whereas the "b" tag is used for bold letters, and the "i" tag is used for italics.

# Simple HTML code for MTurk

- ▶ The next step is setting up forms for user input
  - ▶ `< form > ... < /form >` : creates a form for user input.
  - ▶ The `input type` defines different types of input elements, eg: text fields, checkboxes, radio buttons etc.
  - ▶ The `name` attribute assigns the **variable name**, the `value` attribute assigns **pre-determined values** for multiple choice questions.
  - ▶ Some examples next. You can use these as a base and change the source code in the **Design Layout** (described in Step 3b on Pages 7-9) to make your own custom HITs.
- ▶ **Text, number and textbox input**

```
<!DOCTYPE html>
<html>
<body>

<p> Here are two input types, the first is text and the
second is a number. </p>

<form>
<b>Describe the weather in one word:</b>
<input type = "text" name = "weather">
<br>
<b>Guess the temperature in Celsius: </b>
<input type = "number" name = "temp">
</form>

<br>
<p> Next is an example of an open-ended question. Use
this when you need Workers to enter multi-line texts. Try
varying cols and rows to see what happens.</p>

<form>
<b>Can you describe the weather in some more detail? </b>
<br>

<textarea cols="80" name="MultiLineTextInput"
rows="5"></textarea>

</form>
```

Here are two input types, the first is text and the second is a number.

**Describe the weather in one word:**

**Guess the temperature in Celsius:**

Next is an example of an open-ended question. Use this when you need Workers to enter multi-line texts. Try varying cols and rows to see what happens.

**Can you describe the weather in some more detail?**

# Simple HTML code for MTurk

## ► Multiple choice with radio button (one pre-checked)

```
<!DOCTYPE html>
<html>
<body>
<p> <b> What's your favorite pastime? </b> </p>
<form>

  <input type="radio" name="gender" value="Netflix">
  Watching Netflix <br>

  <input type="radio" name="gender" value="Books">
  Reading books <br>

  <input type="radio" name="gender" value="MTurk">
  Making MTurk Surveys <br>

  <input type="radio" name="gender" value="Food"> FOOD
  <br>

  <input type="radio" name="gender" value="Other"
  checked> Other <br>
</form>

</body>
</html>
```

What's your favorite pastime?

- ☐ Watching Netflix
- ☐ Reading books
- ☐ Making MTurk Surveys
- ☐ FOOD
- ☒ Other

## ► Multiple choice with checkboxes

```
<!DOCTYPE html>
<html>
<body>

<p> Ready to start running your own customized Mturk
surveys? </p>
<form>

<input type="checkbox" name="ready" value="yes"> Yes
<br>
<input type="checkbox" name="ready" value="no"> No
<br>
<input type="checkbox" name="ready" value="maybe"> Maybe so
<br>
</form>

</body>
</html>
```

Ready to start running your own customized Mturk surveys?

- ☐ Yes
- ☐ No
- ☐ Maybe so

Thank you!!