

Crowdsourcing with MTurkR

Thomas J. Leeper

Department of Political Science,



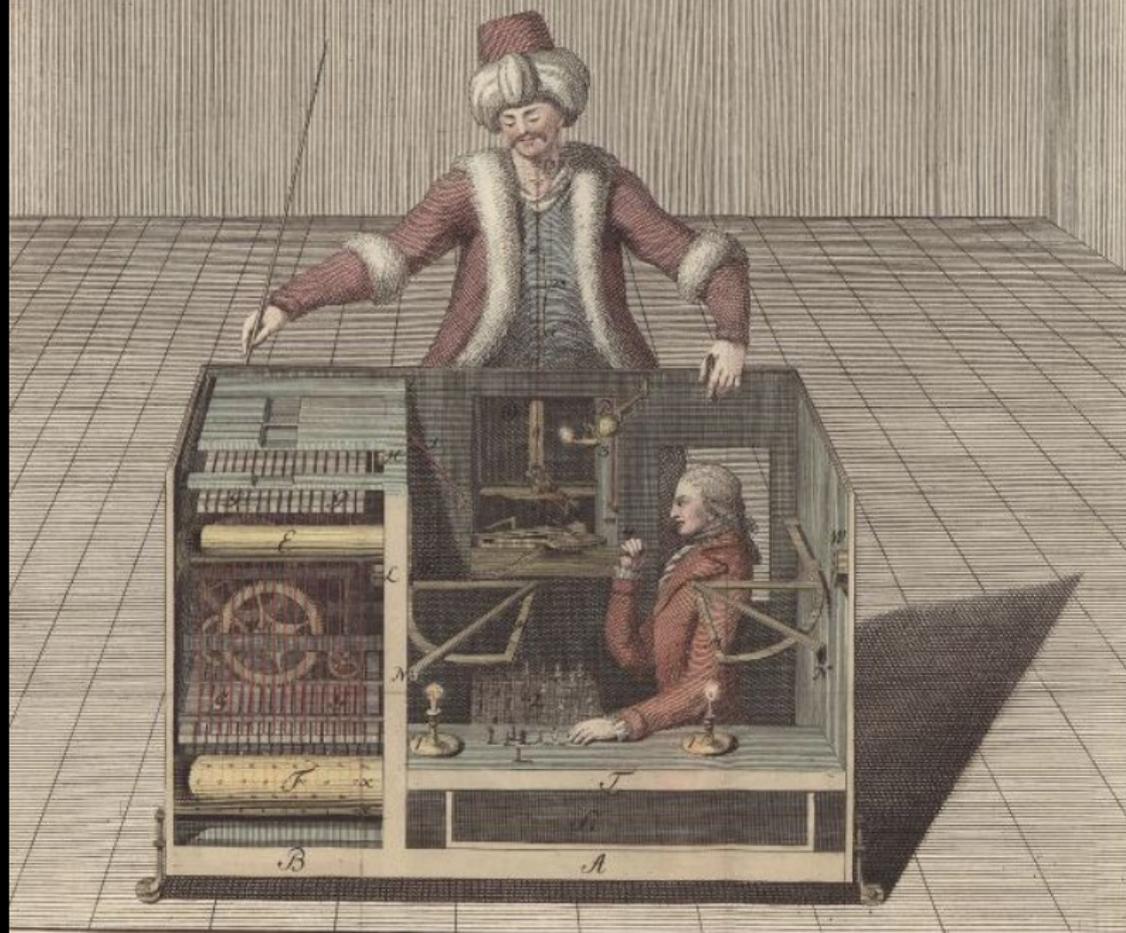
AARHUS UNIVERSITY

Twitter: @thosjleeper
GitHub: leeper
thosjleeper@gmail.com

Imagine we have some data...

	gender	var1	var2	first	last	image
1	female	0.5	1	sara	annala	img94.jpg
2	male	0.6	3	julius	haataja	img69.jpg
3	male	1.2	2	ross	meyer	img32.jpg
4	female	0.3	1	sarah	lahti	img96.jpg
5	female	1.1	5	ada	park	img24.jpg
6	female	0.9	2	joan	hernandez	img92.jpg
7	female	0.4	1	sofia	korhonen	img87.jpg
8	female	0.1	3	helle	kivela	img52.jpg
9	male	1.8	4	kasper	johnson	img17.jpg
10	male	0.6	2	dirk	luoma	img62.jpg

...but how do we analyze an *image* variable?



Coding

Data search/
retrieval/scraping

Categorization

Content
moderation

Manual
translation

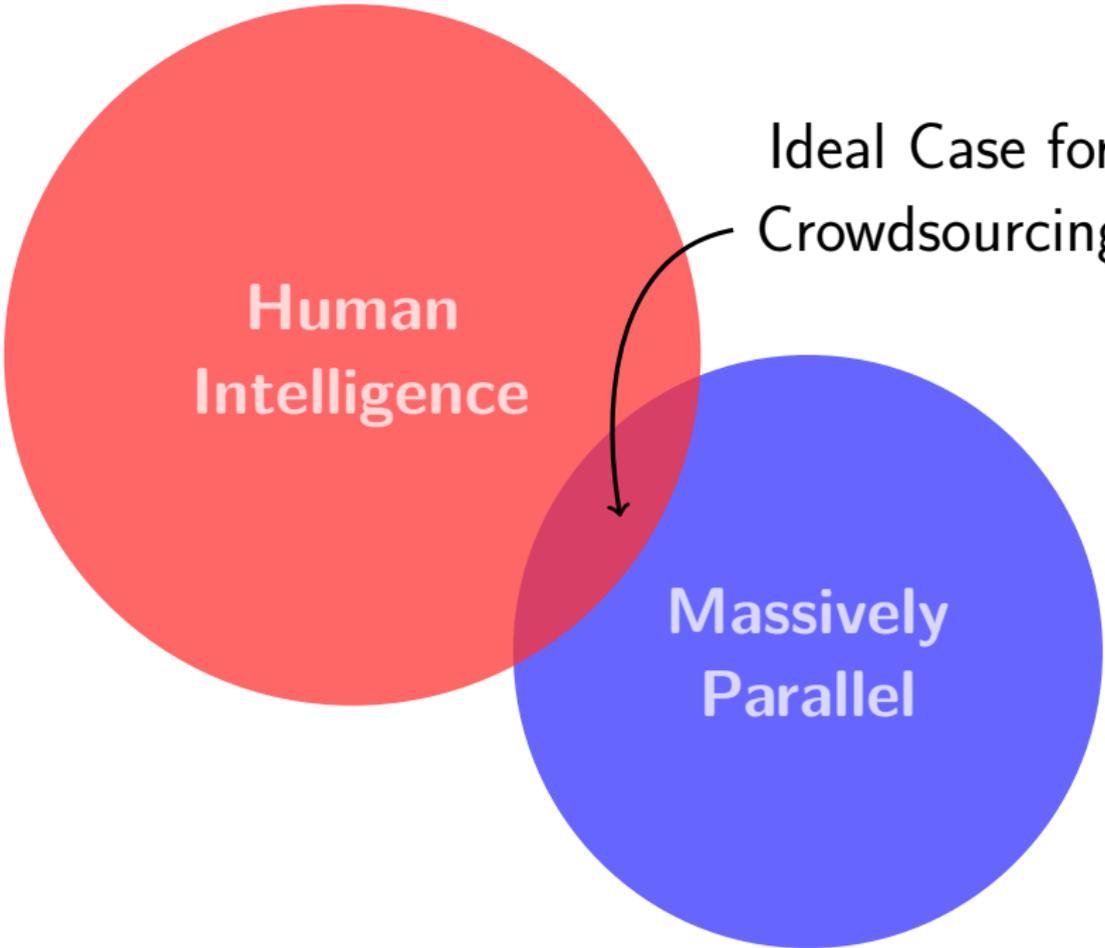
Audio/Video
Transcription

**Human
subjects
research**

Writing tasks

Building
training sets

UX testing

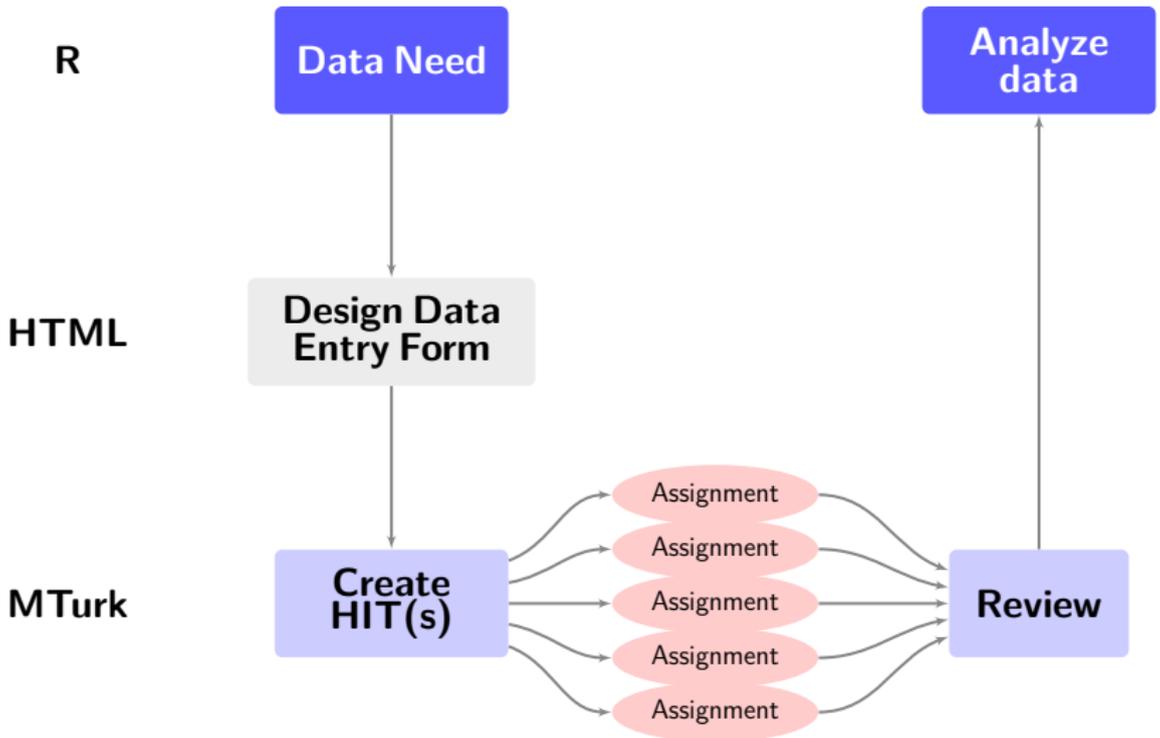


A Venn diagram consisting of two overlapping circles. The left circle is red and contains the text "Human Intelligence". The right circle is blue and contains the text "Massively Parallel". The overlapping area between the two circles is shaded a darker red. An arrow points from the text "Ideal Case for Crowdsourcing" to this overlapping area.

**Human
Intelligence**

Ideal Case for
Crowdsourcing

**Massively
Parallel**




```
# set API keys in environment variables

library("MTurkR")

BulkCreateFromURLs(
  url = paste0("https://example.com/", 1:10, ".html"),

  title = "Image Categorization",
  description = "Describe contents of an image",
  keywords = "categorization, image",
  reward = .01,
  duration = seconds(minutes = 5),

  annotation = "My Project",
  expiration = seconds(days = 4),
  auto.approval.delay = seconds(days = 1)
)
```

Get back a data.frame:

```
GetAssignments(annotation = "My Project")
```

The image coding task with **27,500 images**

took **225 workers**

about **75 minutes**

and cost **\$412.50**

Pay workers with:

```
ApproveAssignments(annotation = "My Project")
```

```
a = GenerateHTMLQuestion(file = "hit.html")

hit = CreateHIT(
  title = "Short Survey",
  description = "5 question survey",
  keywords = "survey, questionnaire",
  duration = seconds(hours = 1)
  reward = .10,

  assignments = 5000,
  expiration = seconds(days = 4),
  question = a$string,
)
```

```
GetHIT(hit$HITId)
```

```
ExtendHIT(hit$HITId,  
  add.assignments = 500)  
  add.seconds = seconds(days = 1)  
)
```

```
ExpireHIT(hit$HITId)
```

```
ChangeHITType(hit$HITId,  
  title = "New, better title",  
  reward = 5.00  
)
```

Advanced Features

Choose who works
for you



Qualifications
and tests

Monitor HITs



Notifications

Sanction and reward
workers



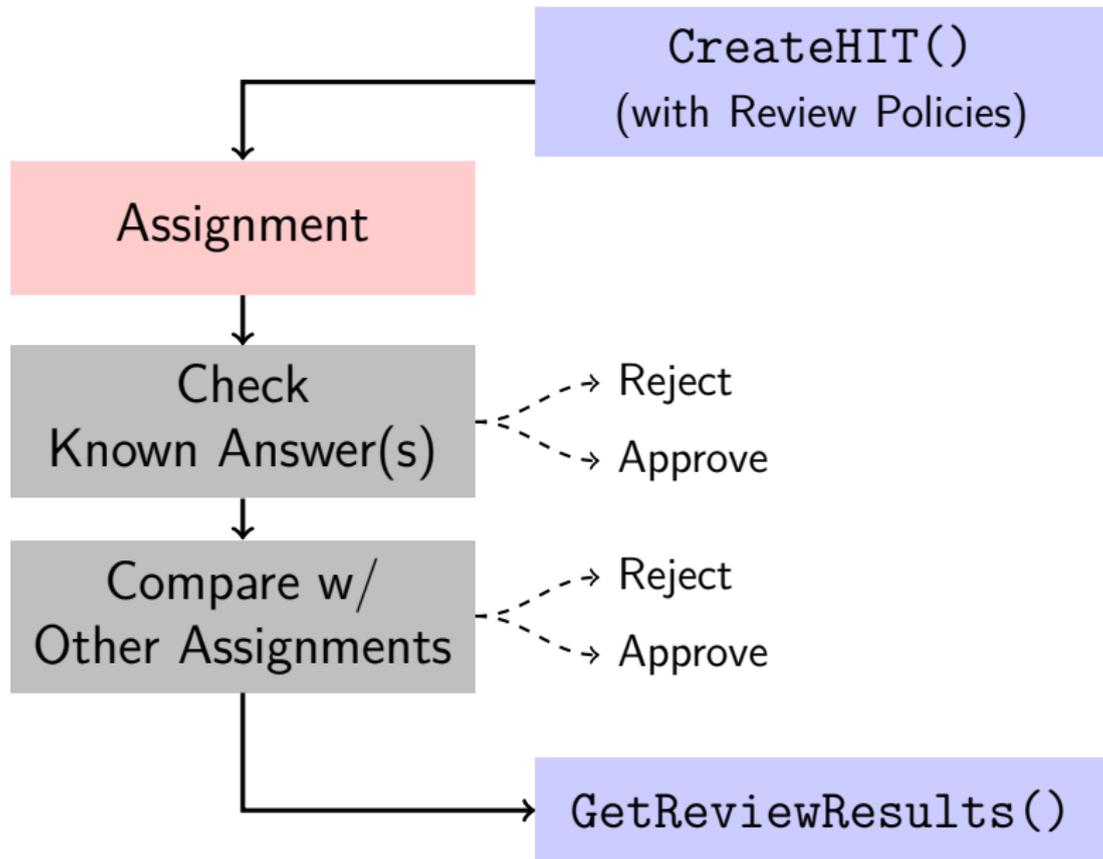
Qualifications,
bonuses, and blocks

Automatic review



Review Policies

Anatomy of an MTurkR App



What's next?

1 Packages for more crowdsourcing platforms

- Common interface?



2 HIT templates



3 Performance improvements



```
# Start Crowdsourcing
```

```
# CRAN
```

```
install.packages("MTurkR")
```

```
# GitHub
```

```
install_github("leeper/MTurkR")
```

```
# Questions?
```

```
# thosjleeper@gmail.com
```

```
# https://github.com/leeper/MTurkR/wiki
```