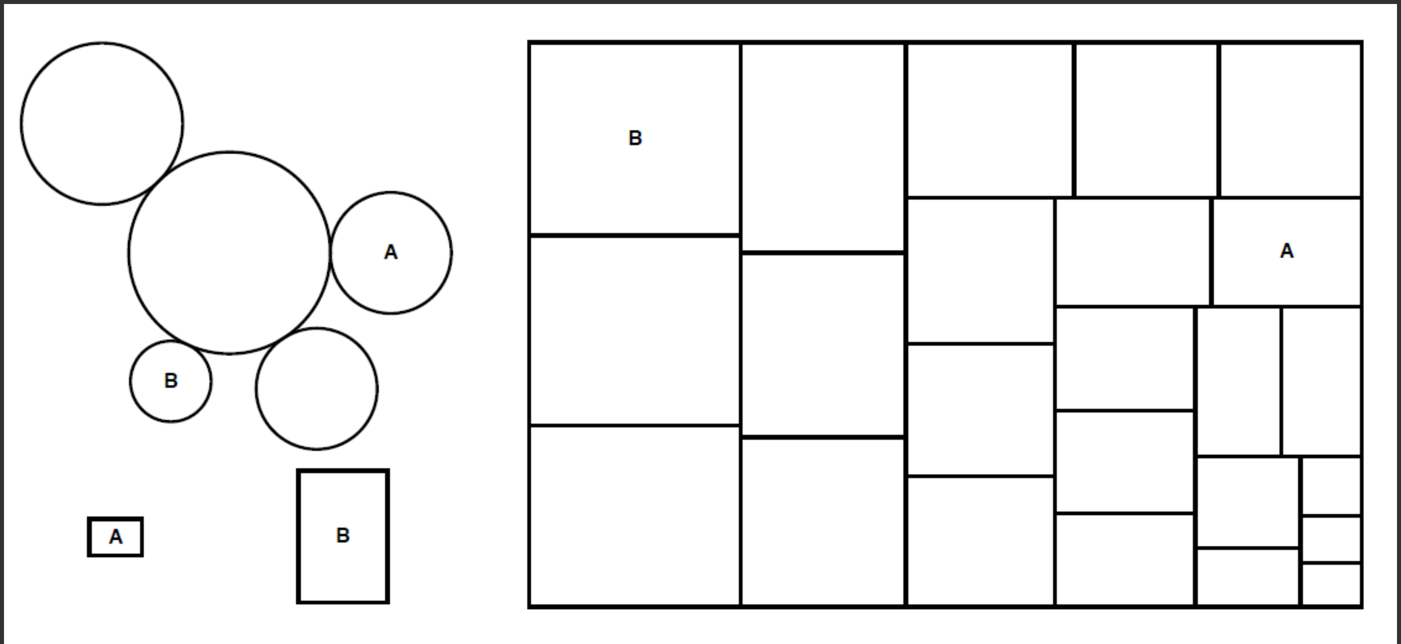
Crowdsourcing Graphical Perception

Using Mechanical Turk to Assess Visualization Design



Jeffrey Heer & Michael Bostock
Stanford University

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

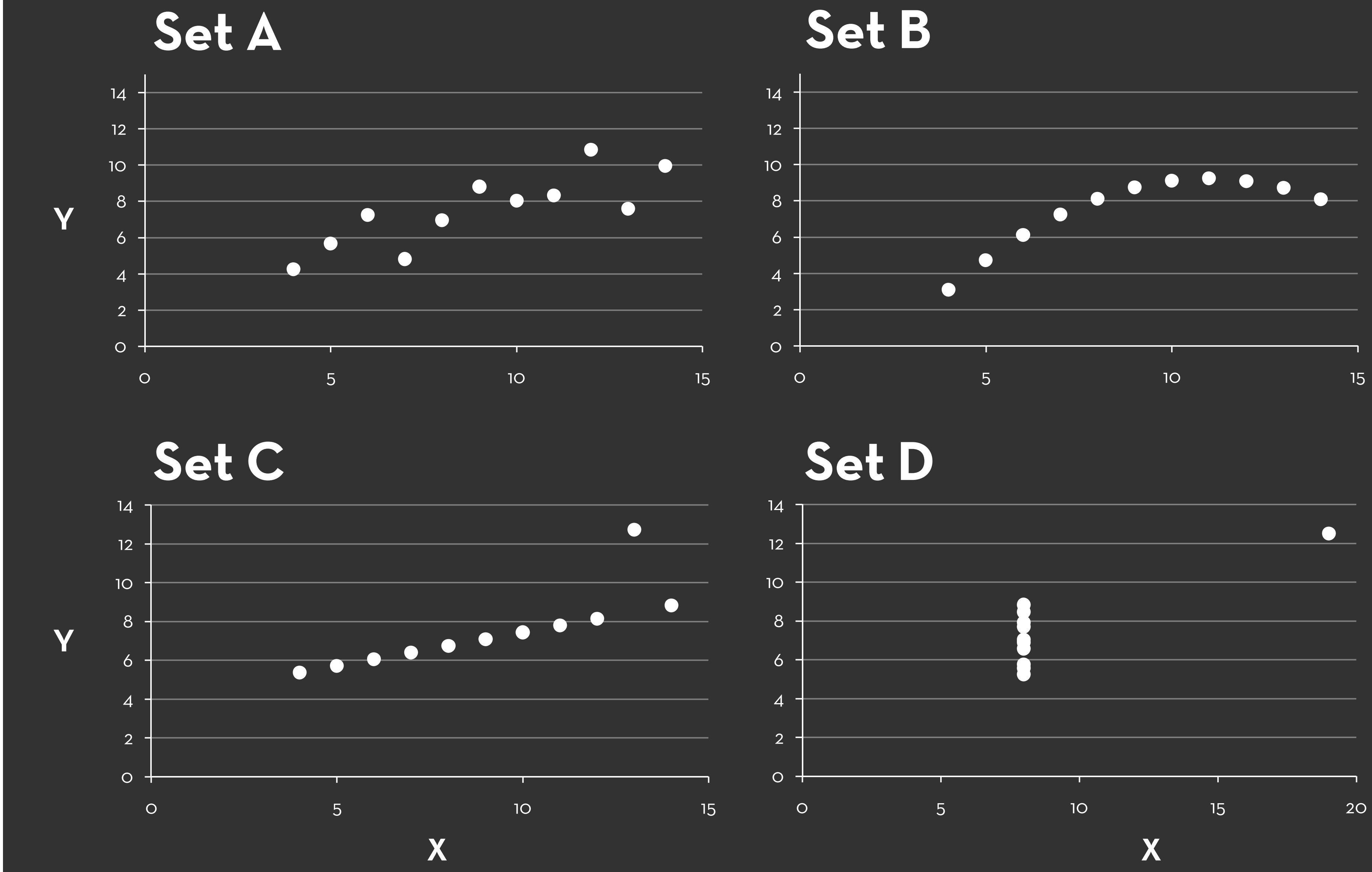
$u_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression

$Y^2 = 3 + 0.5\,X$

$R^2 = 0.67$

[Anscombe 73]

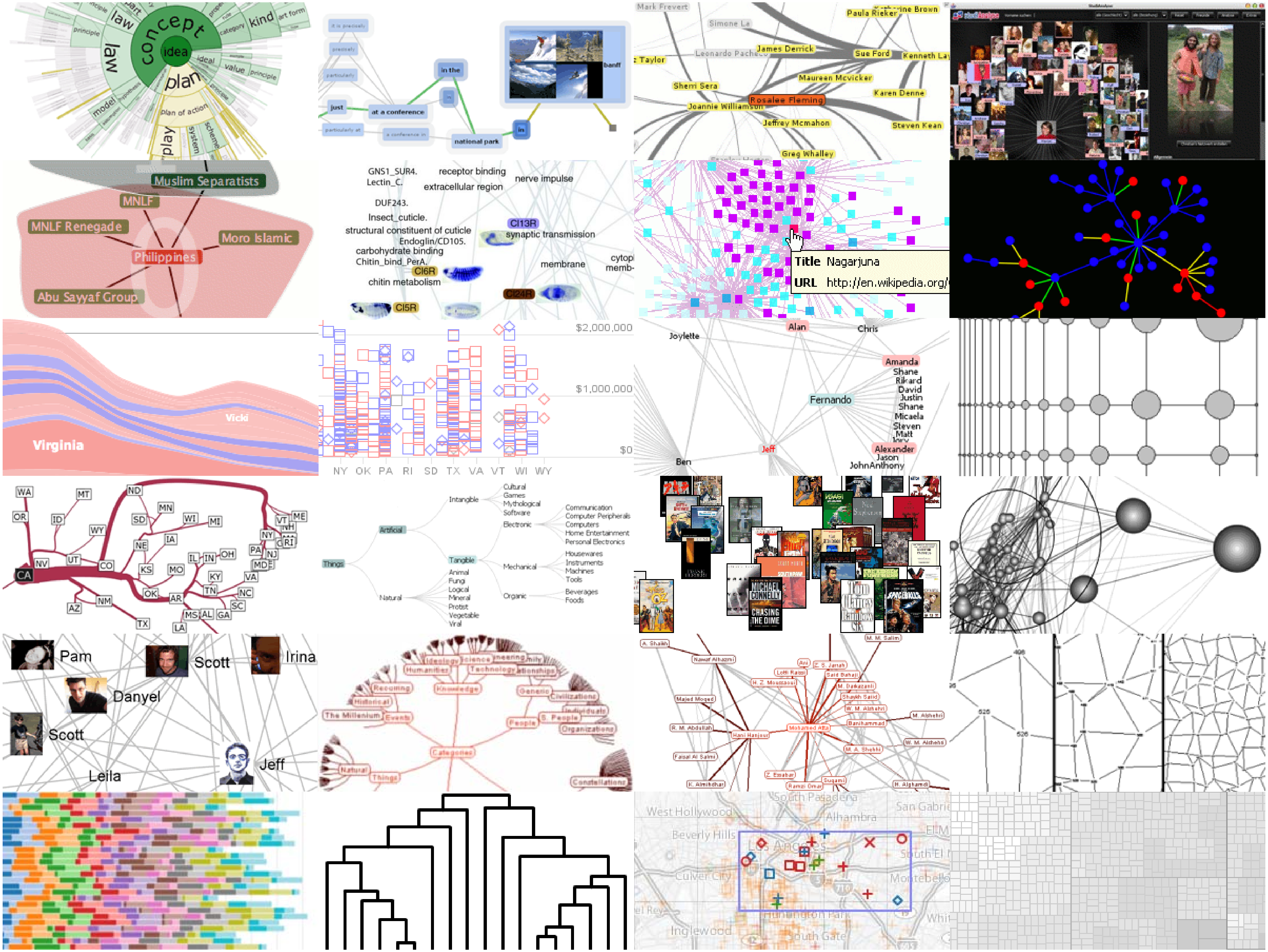


Set A

Set B

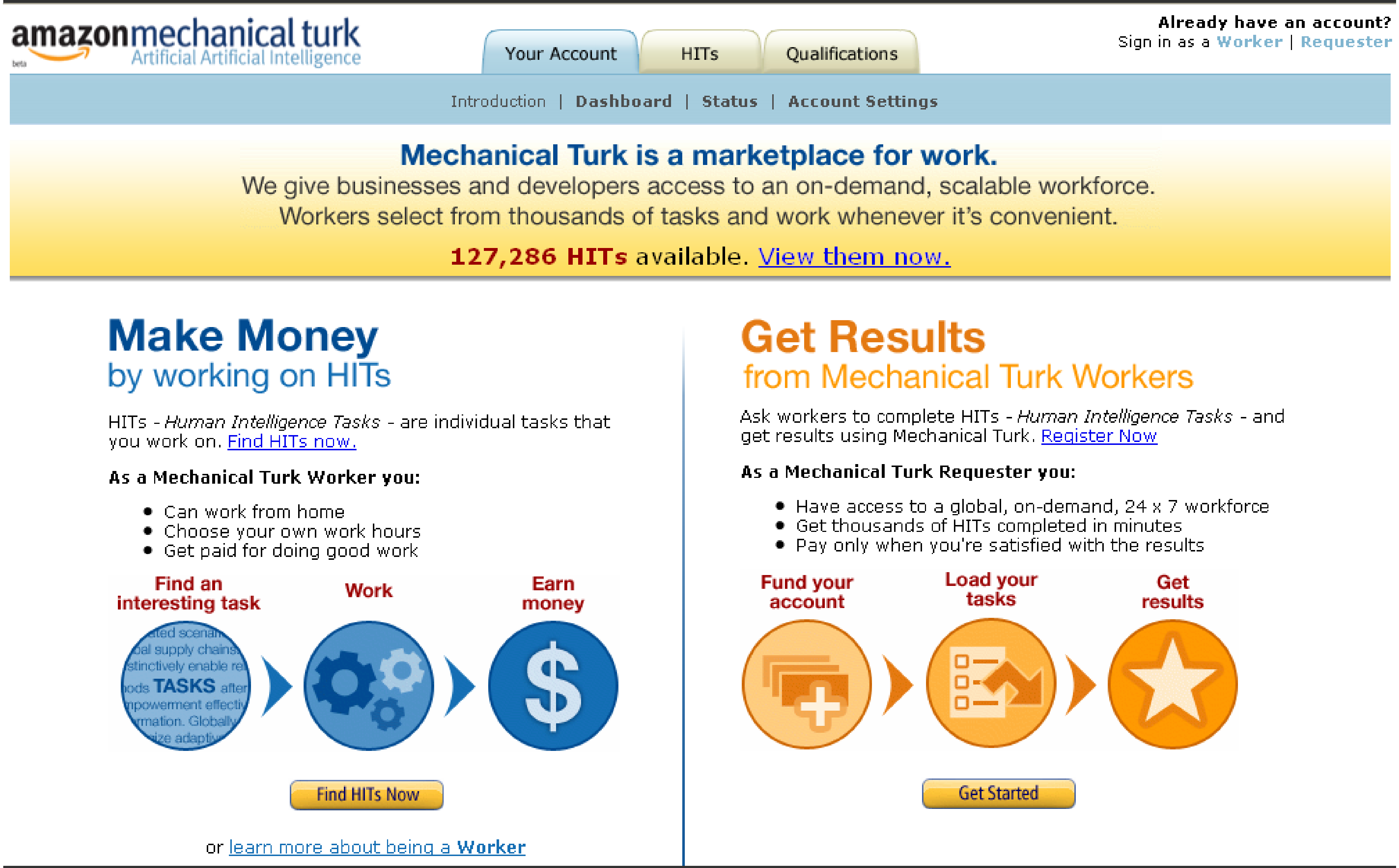
Set C

Set D



concept
idea
law
plan
kind
principle
ideal
value
model
play
plan of action
system
scheme
art form
in the
at a conference
just
national park
in
banff
Mark Prevert
Simone La
Paula Rieker
Katherine Brown
Leonardo Pacheco
James Derrick
Sue Ford
Kenneth Lay
Taylor
Maureen Mcvicker
Sherri Sera
Rosalee Fleming
Karen Denne
Joannie Williamson
Jeffrey Mcmahon
Steven Kean
Greg Whalley
Muslim Separatists
MNLF
MNLF Renegade
Moro Islamic
Philippines
Abu Sayyaf Group
GNS1_SUR4.
Lectin_C.
receptor binding
extracellular region
nerve impulse
DUF243.
Insect_cuticle.
structural constituent of cuticle
Endoglin/CD105.
carbohydrate binding
Chitin_bind_PerA.
chitin metabolism
CI13R
synaptic transmission
CI6R
CI5R
CI24R
membrane
cytopl
memb-
Title Nagarjuna
URL http://en.wikipedia.org/
Virginia
Vicki
$2,000,000
$1,000,000
$0
NY
OK
PA
RI
SD
TX
VA
VT
WI
WY
Joylette
Alan
Chris
Amanda
Shane
Rikard
David
Justin
Shane
Micaela
Steven
Matt
Fernando
Jeff
Alexander
Jason
JohnAnthony
Ben
WA
MT
ND
OR
ID
MN
SD
WI
MI
WY
VT
ME
NH
NY
RI
IA
NE
CT
PA
NJ
IL
IN
OH
NV
UT
CO
KS
MO
MD
CA
KY
VA
OK
TN
NC
AR
AZ
NM
SC
MS
AL
GA
TX
LA
Things
Artificial
Intangible
Cultural
Games
Mythological
Software
Electronic
Communication
Computer Peripherals
Computers
Home Entertainment
Personal Electronics
Tangible
Mechanical
Housewares
Instruments
Machines
Tools
Natural
Animal
Fungi
Logical
Mineral
Protist
People
Vegetable
Viral
Organic
Beverages
Foods
MICHAEL CONNELLY
CHASING THE DIME
Pam
Scott
Irina
Danyel
Scott
Jeff
Leila
Ideology
Science
Meaning
Humanities
Technology
Relationships
Family
Knowledge
Recurring
Historical
The Millennium
Events
Generic
Civilizations
Individuals
People
S. People
Organizations
Categories
Natural
Things
Constellations
A. Shaikh
Nawaf Alhazmi
Majed Moqed
Lotfi Raissi
H. Z. Moussaoui
S. S. Jarrah
Said Bahaji
Mohamed Atta
Marwan Al-Shehhi
Ramzi Omar
Ziad Jarrah
Hani Hanjour
M. M. Salim
Faisal Al Salmi
R. M. Abdullah
Z. Essabar
K. Almihdhar
M. A. Shehhi
Binalshibh
M. Alshehri
H. Alghamdi
West Hollywood
South Pasadena
San Gabriel
Alhambra
Beverly Hills
El M
Los Angeles
South El M
Culver City
Montebello
Huntington Park
Inglewood
South Gate
Whit

hawaii.naist.jp/research/visual_e.html



Already have an account?
Sign in as a Worker | Requester

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.

127,286 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

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

Andy Baio - Waxy.org

I TURK FOR MONEY FOR MY BOOKS!!!
I turk for Andy's amusement
I Turk for FUN!
I TURK FOR CASH
I Turk for money baby!!!
I Turk to Pass the Time
11/15/2008
I turk for to turk brings CHANGE
I TURK FOR EXTRA CASH
I TURK TO EARN $ FOR BOOKS!

Using MTurk for Research

Machine Learning, Comp. Vision & Info. Retrieval
User-Generated Metadata, Labeling Data

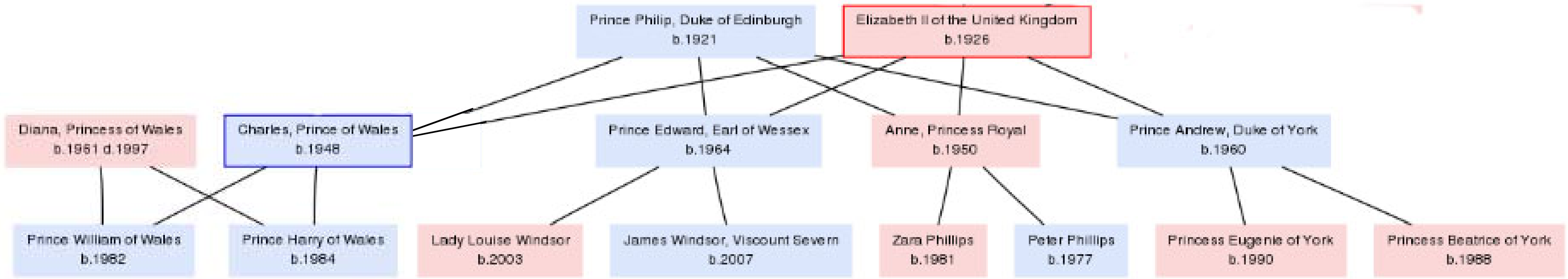
Kittur, Chi & Suh: Wikipedia Article Quality
Use verifiable questions to reduce gaming
Make sincere responses as easy as insincere ones

Mason & Watts: Financial Incentives
Higher reward → faster completion, same quality

AN EXAMPLE:

TimeNets for **Genealogical Data**

Visualizing Genealogical Graphs



Prince Philip, Duke of Edinburgh
b.1921

Elizabeth II of the United Kingdom
b.1926

Diana, Princess of Wales
b.1961 d.1997

Charles, Prince of Wales
b.1948

Prince Edward, Earl of Wessex
b.1964

Anne, Princess Royal
b.1950

Prince Andrew, Duke of York
b.1960

Prince William of Wales
b.1982

Prince Harry of Wales
b.1984

Lady Louise Windsor
b.2003

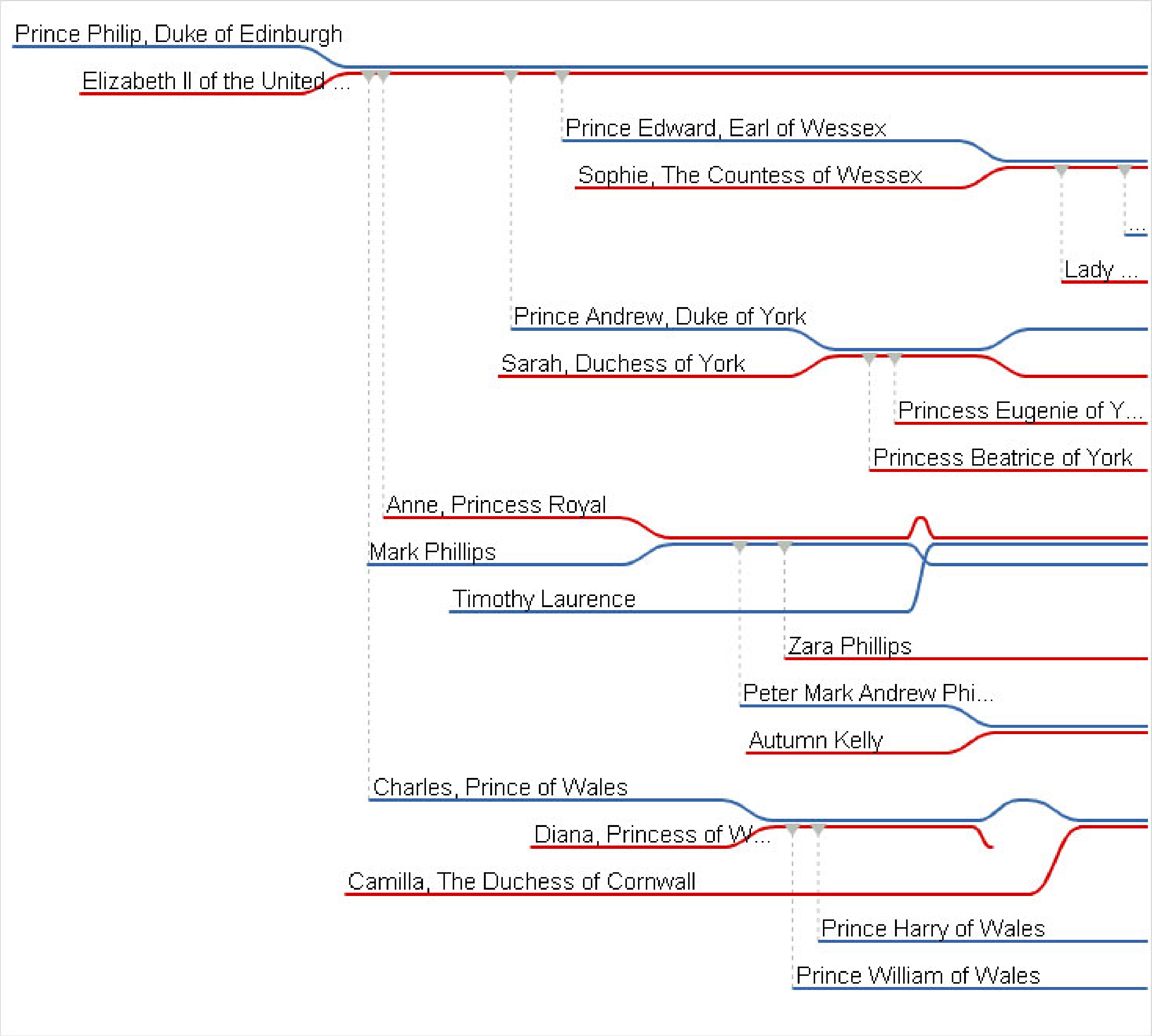
James Windsor, Viscount Severn
b.2007

Zara Phillips
b.1981

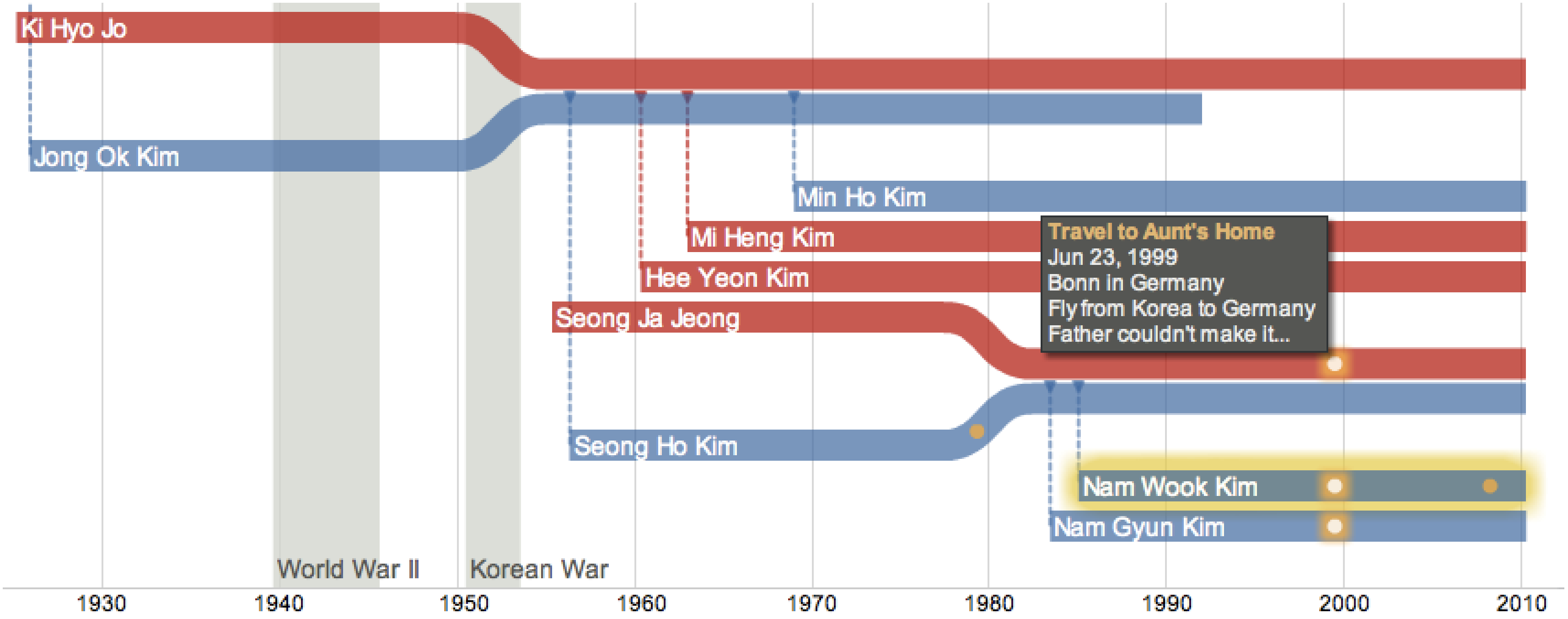
Peter Phillips
b.1977

Princess Eugenie of York
b.1990

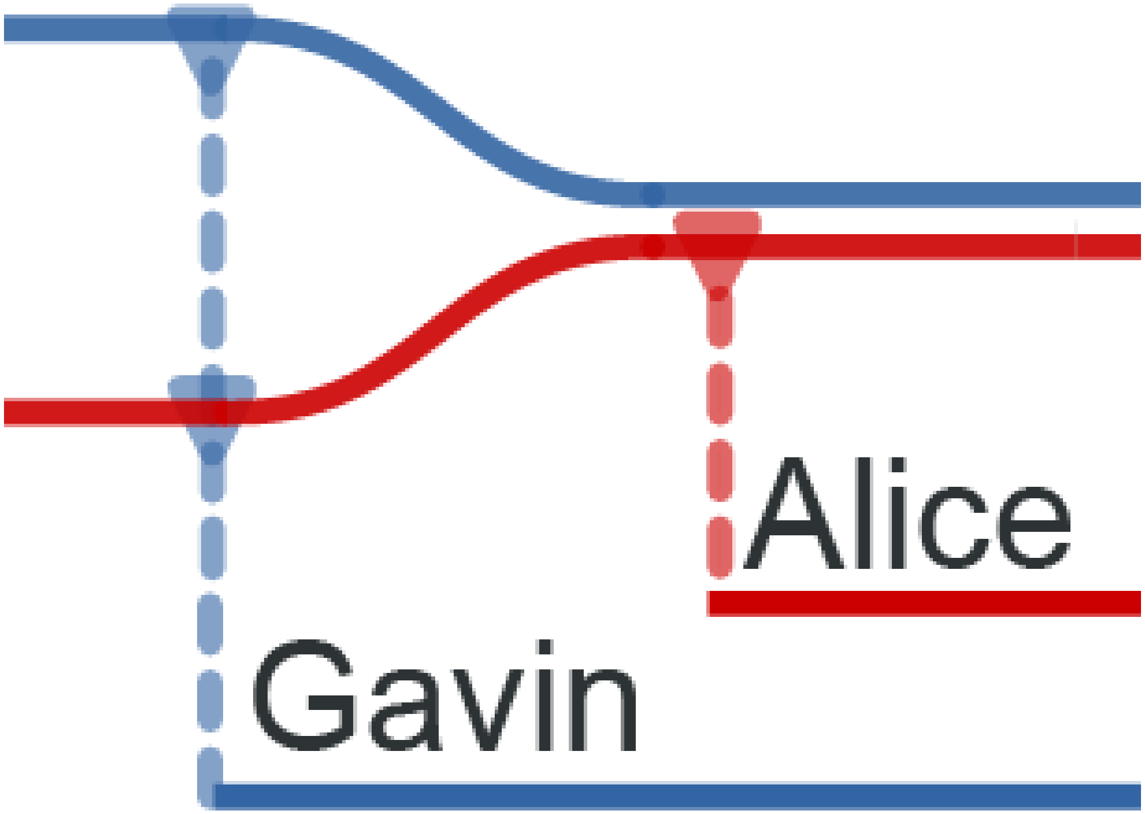
Princess Beatrice of York
b.1988


Prince Philip, Duke of Edinburgh
Elizabeth II of the United ...
Prince Edward, Earl of Wessex
Sophie, The Countess of Wessex
...
Lady ...
Prince Andrew, Duke of York
Sarah, Duchess of York
Princess Eugenie of Y...
Princess Beatrice of York
Anne, Princess Royal
Mark Phillips
Timothy Laurence
Zara Phillips
Peter Mark Andrew Phi...
Autumn Kelly
Charles, Prince of Wales
Diana, Princess of W...
Camilla, The Duchess of Cornwall
Prince Harry of Wales
Prince William of Wales

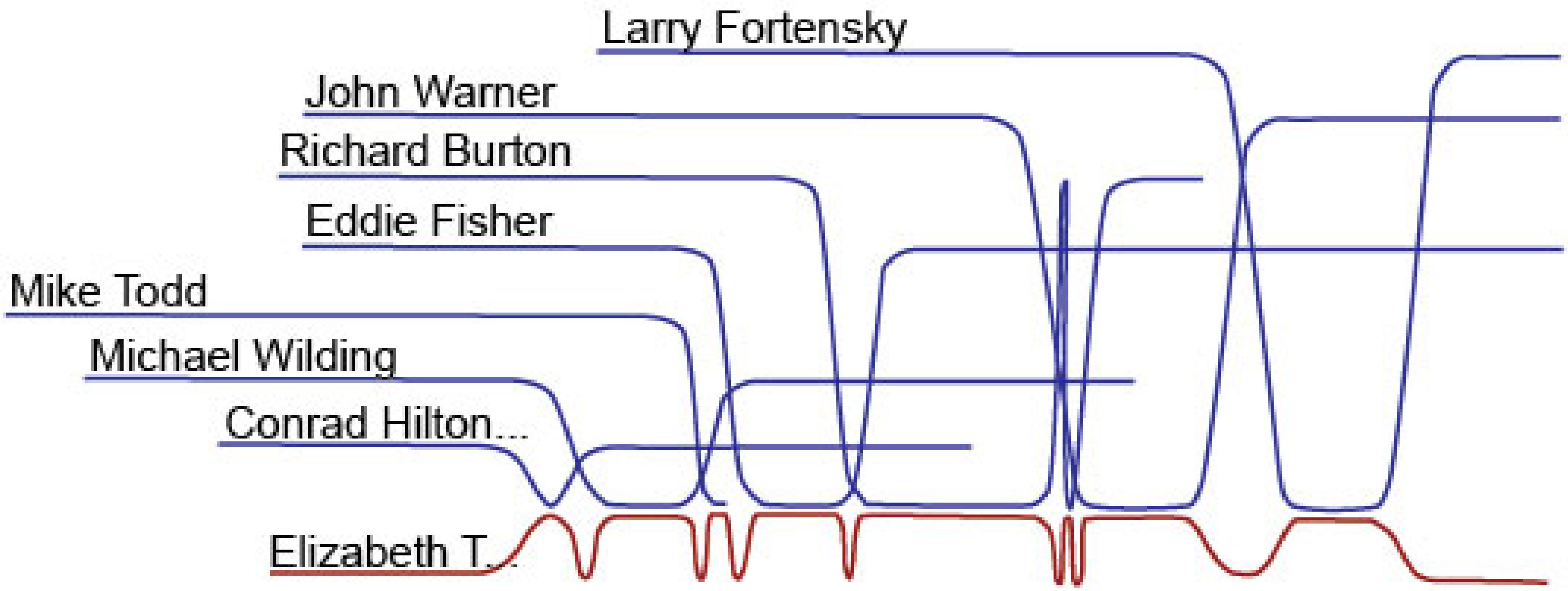
TimeNets = Time x Family Trees



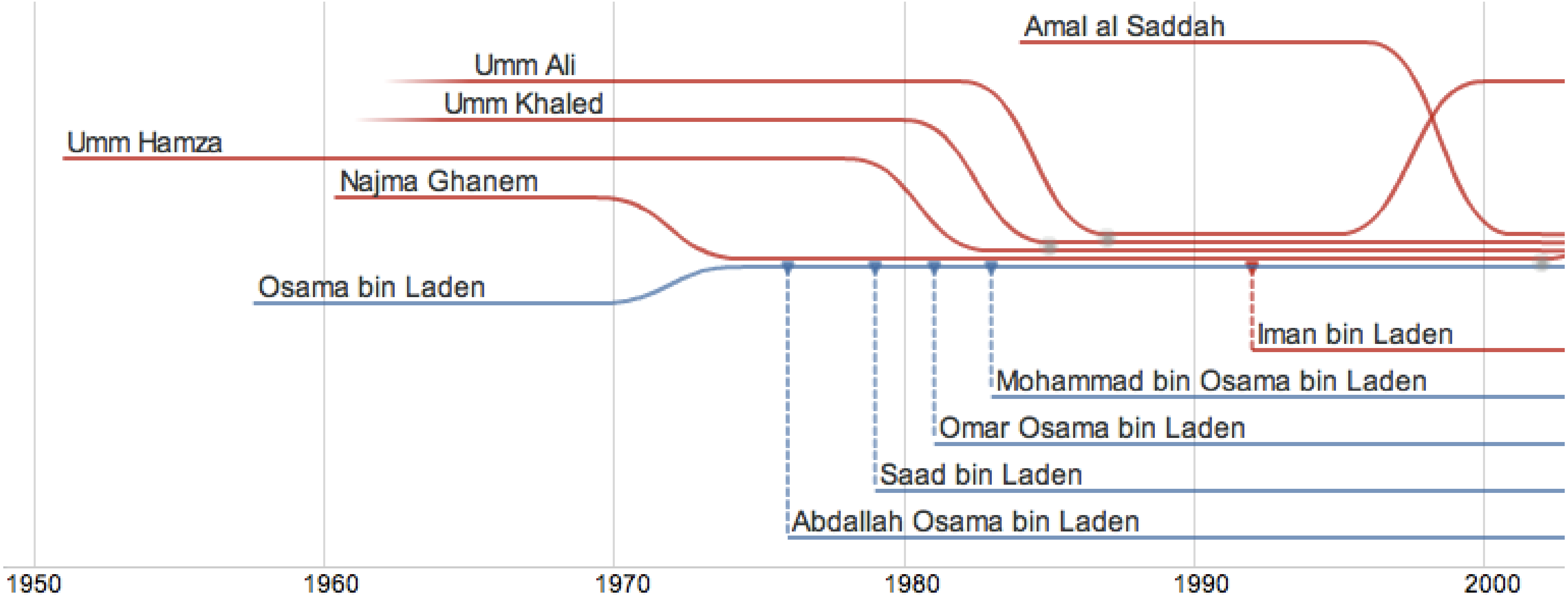
(Out-of-Wedlock Births)

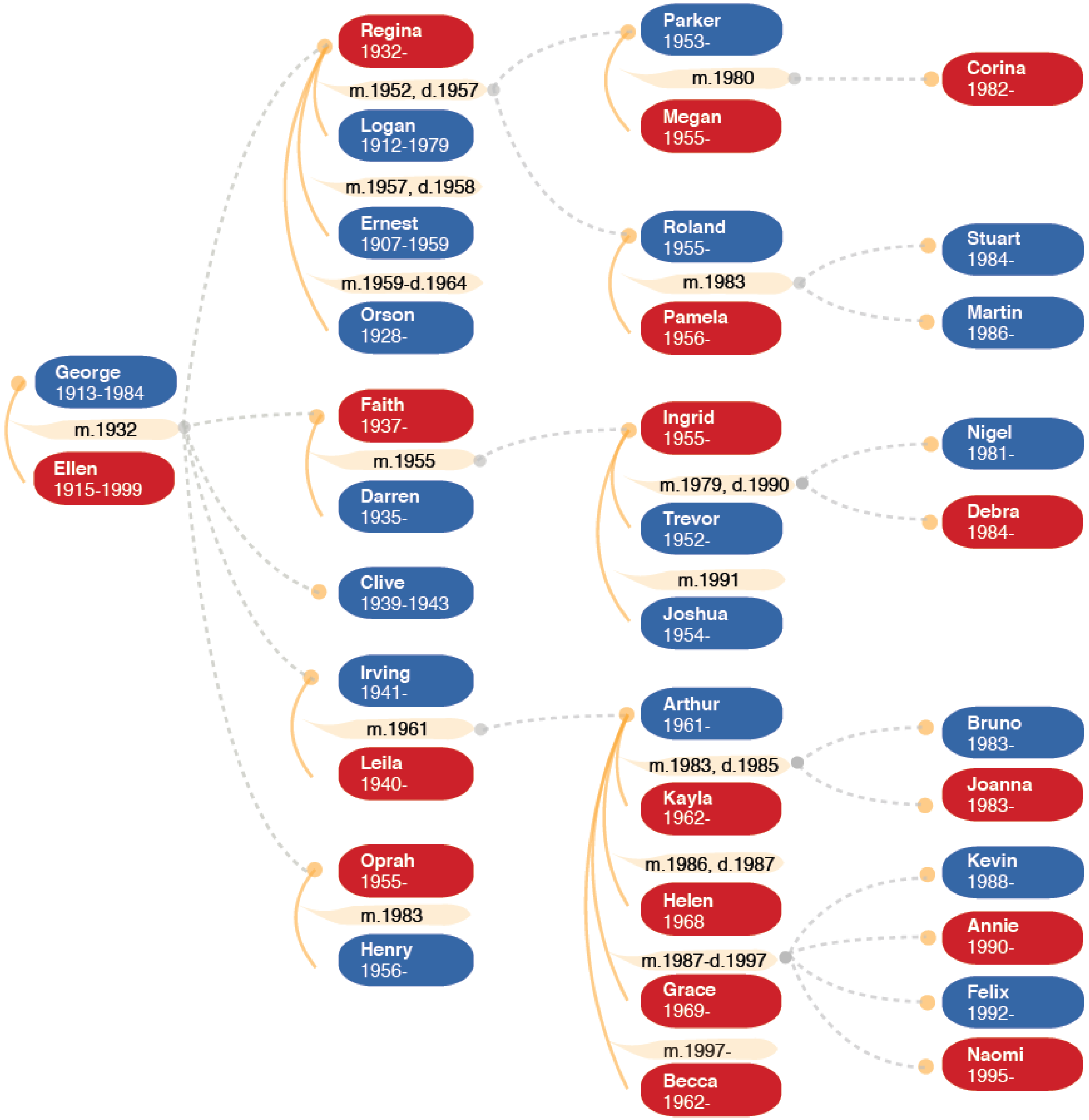


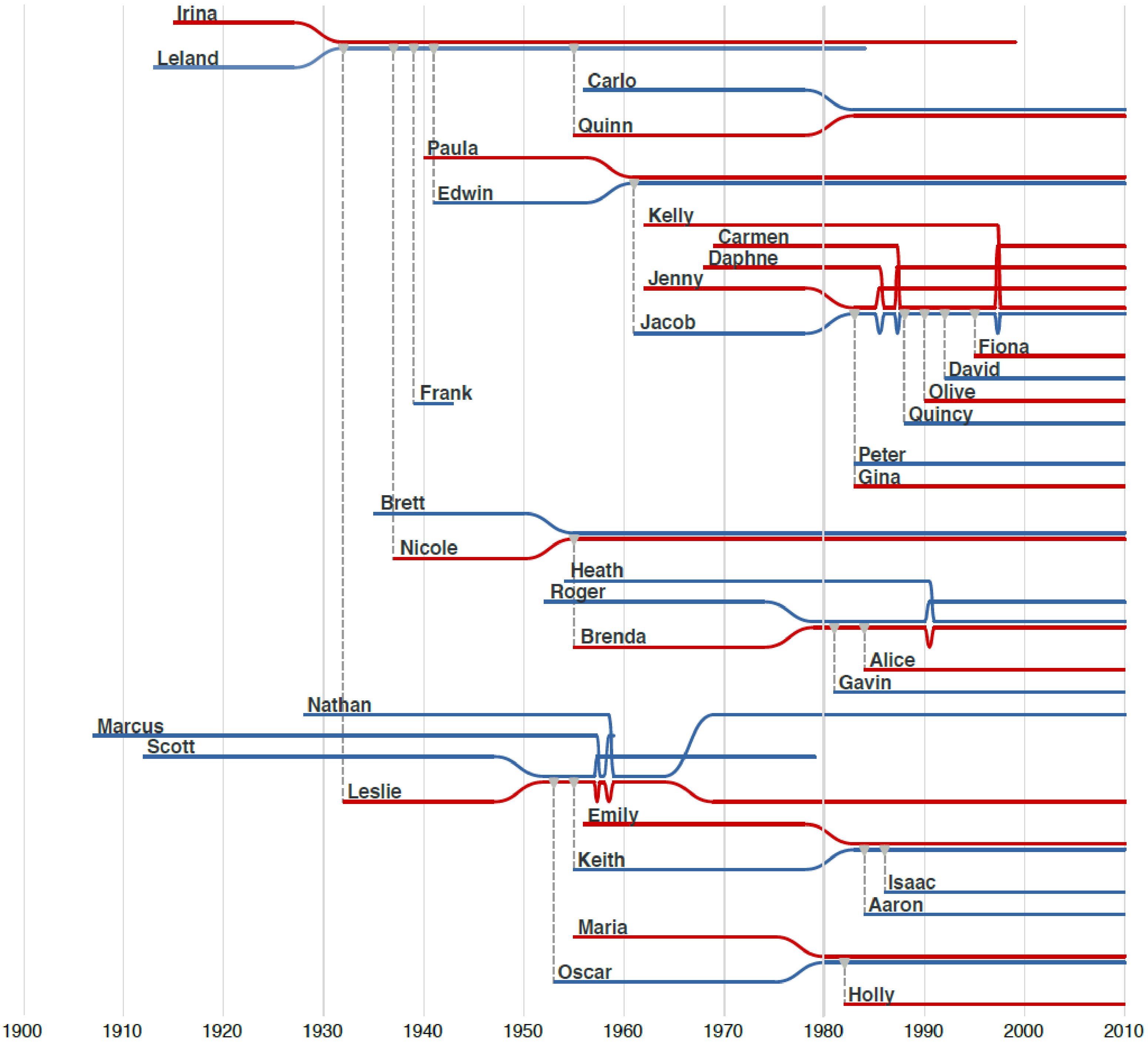
Elizabeth Taylor (Remarriage)

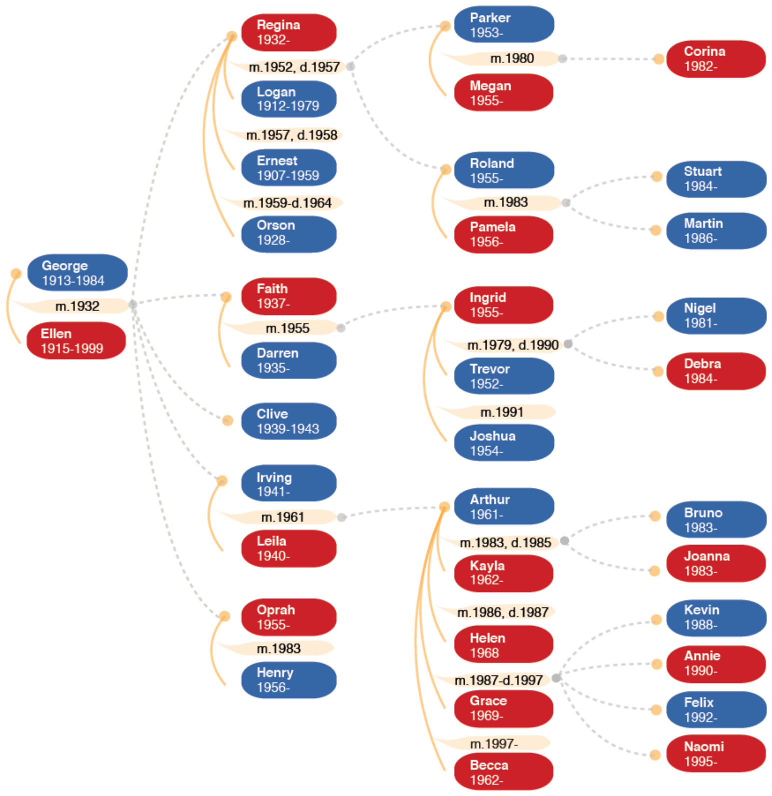


Osama bin Laden (Polygamy)

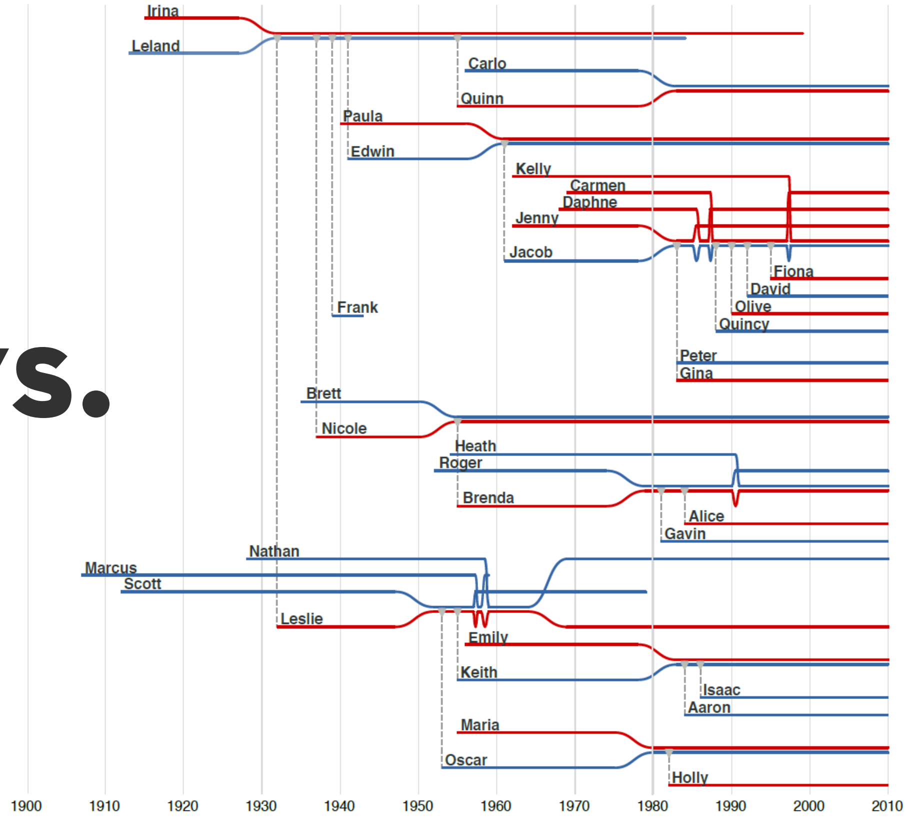



George
1913-1984
m.1932
Ellen
1915-1999
Regina
1932-
m.1952, d.1957
Logan
1912-1979
m.1957, d.1958
Ernest
1907-1959
m.1959-d.1964
Orson
1928-
Faith
1937-
m.1955
Darren
1935-
Clive
1939-1943
Irving
1941-
m.1961
Leila
1940-
Oprah
1955-
m.1983
Henry
1956-
Parker
1953-
m.1980
Megan
1955-
Roland
1955-
m.1983
Pamela
1956-
Ingrid
1955-
m.1979, d.1990
Trevor
1952-
m.1991
Joshua
1954-
Arthur
1961-
m.1983, d.1985
Kayla
1962-
m.1986, d.1987
Helen
1968
m.1987-d.1997
Grace
1969-
m.1997-
Becca
1962-
Corina
1982-
Stuart
1984-
Martin
1986-
Nigel
1981-
Debra
1984-
Bruno
1983-
Joanna
1983-
Kevin
1988-
Annie
1990-
Felix
1992-
Naomi
1995-


Irina
Leland
Carlo
Quinn
Paula
Edwin
Kelly
Carmen
Daphne
Jenny
Jacob
Fiona
David
Olive
Quincy
Frank
Peter
Gina
Brett
Nicole
Heath
Roger
Brenda
Alice
Gavin
Nathan
Marcus
Scott
Leslie
Emily
Keith
Isaac
Aaron
Maria
Oscar
Holly
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010



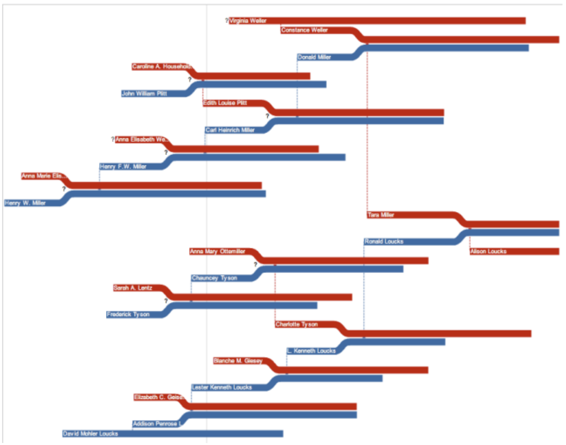
vs.

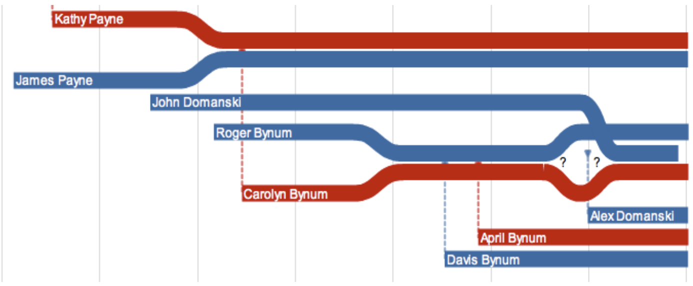


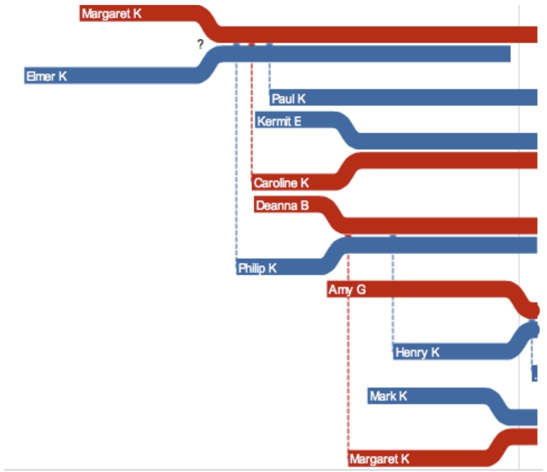
Asked *structural*, *temporal* & *struct x temp* tasks

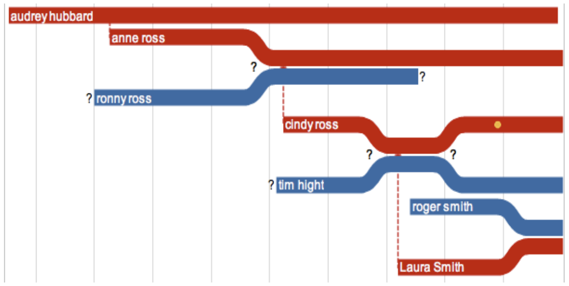
No accuracy differences between visuals

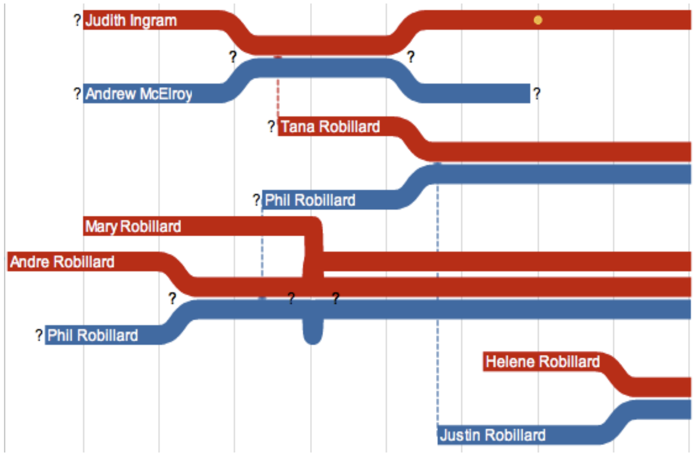
TimeNets were significantly faster (~25%) for tasks with a *temporal* component

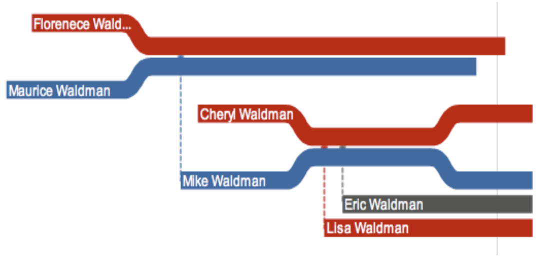












I love the idea of this tool. I love the look and ease of this program! I

I think that the concept is very good and the effort taken is commendable.

Please don't delete my data! Very cool. This is a very interesting idea

i was having a lot of fun with this.i love how it shows everything simply

hawaii.naist.jp/research/visual_e.html

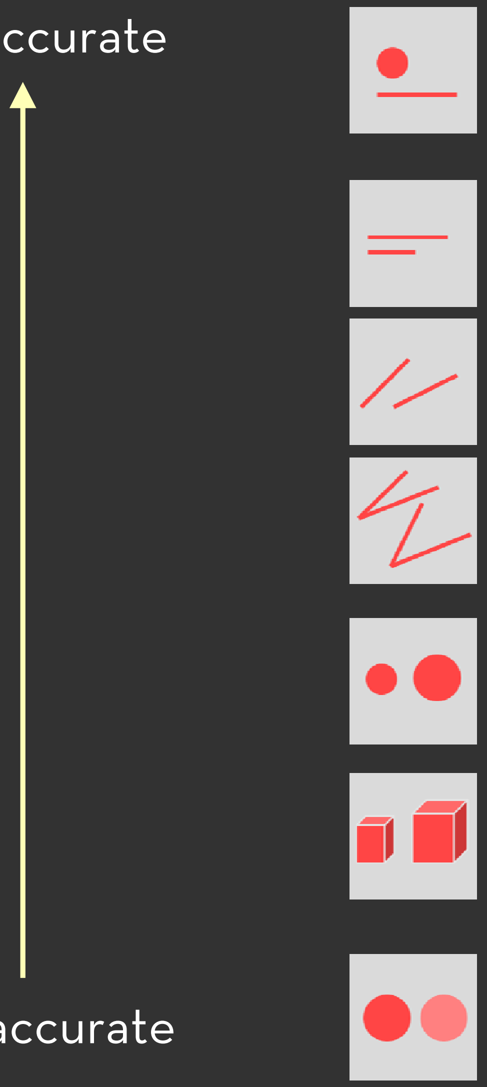
Research Goals

1. **Assess the viability** of crowdsourced perception experiments on Mechanical Turk.

2. Demonstrate the use of MTurk to **gain novel insights** for visualization design.

3. Analyze experimental data to **characterize MTurk as an experimental platform**.

Experiment 1:
Proportional Judgments of Spatial Data Encodings



Most accurate

Position (common) scale

Position (non-aligned) scale

Length

Slope

Angle

Area

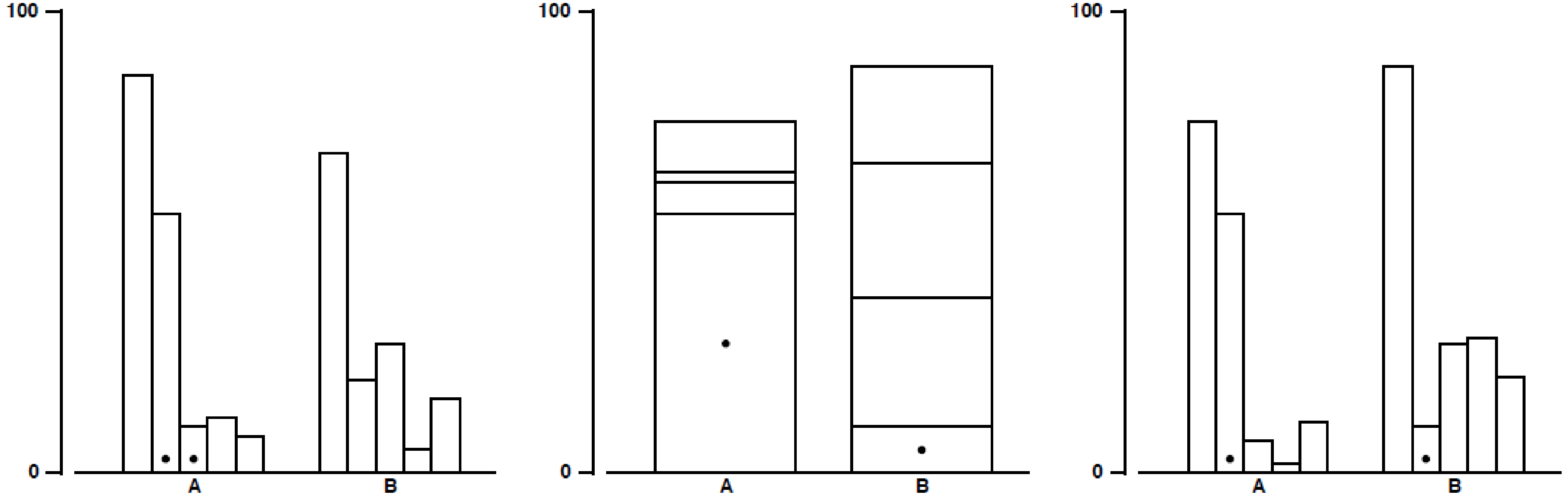
Volume

Color hue-saturation-density

Least accurate

Cleveland & McGill ‘84

Cleveland & McGill, 1984



Stimuli for position encodings.

Task: estimate **%** smaller element is of the larger

Experiment 1A: Proportions

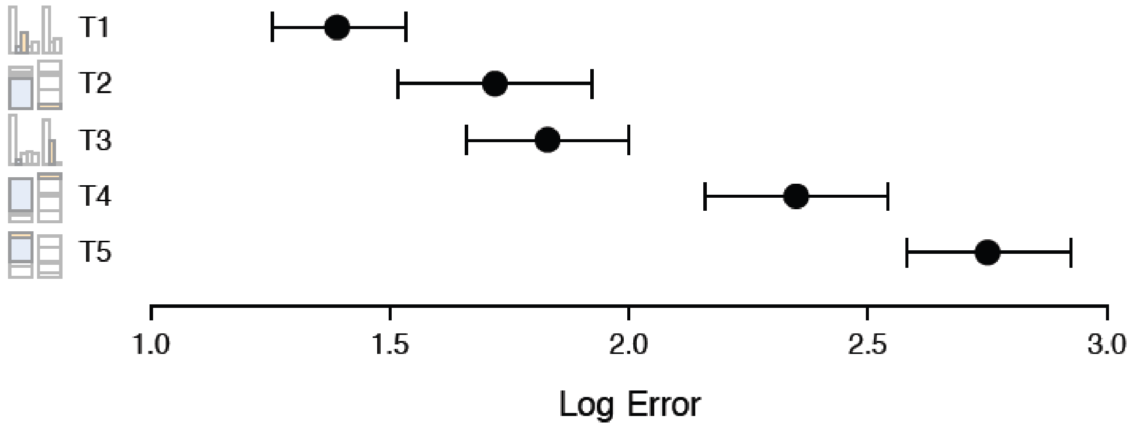
Goal: replicate Cleveland & McGill, 1984

5 *original types*: position (3) + length (2)
+ 2 *new types*: angle + circular area
x 10 *proportional differences*

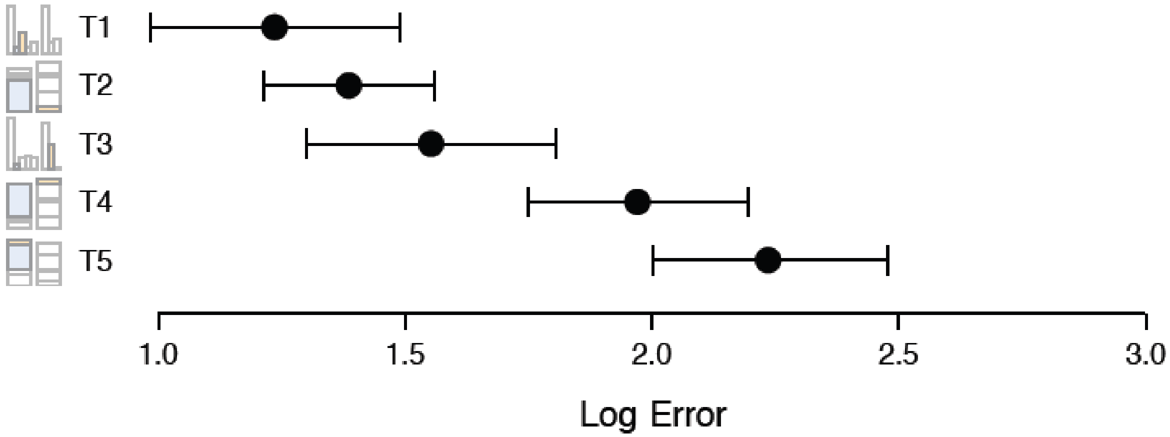
N=50 assignments, $0.05 per HIT

Task: estimate **%** smaller element is of the larger
Error = $\log_2(\,|\,\text{true\%} - \text{estimated\%}\,| + 1/8)$

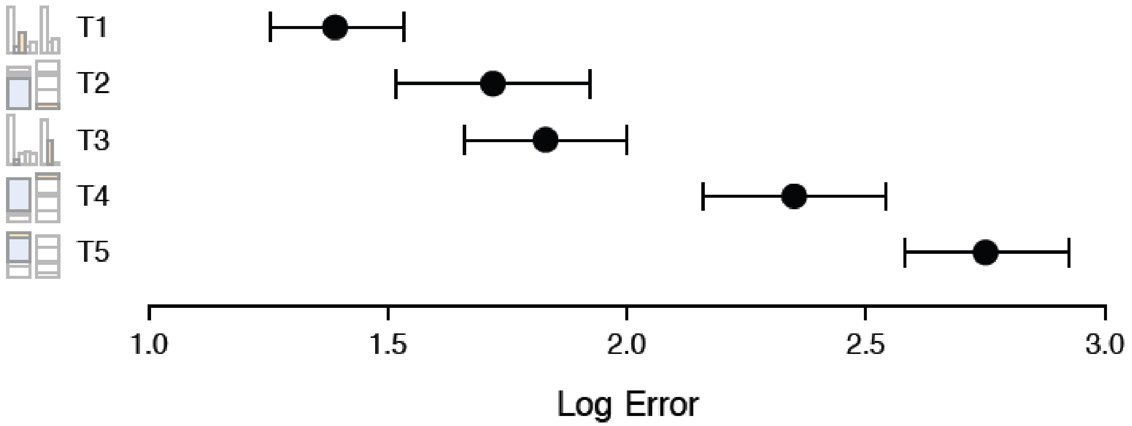
Cleveland & McGill, 1984 (Lab Study)



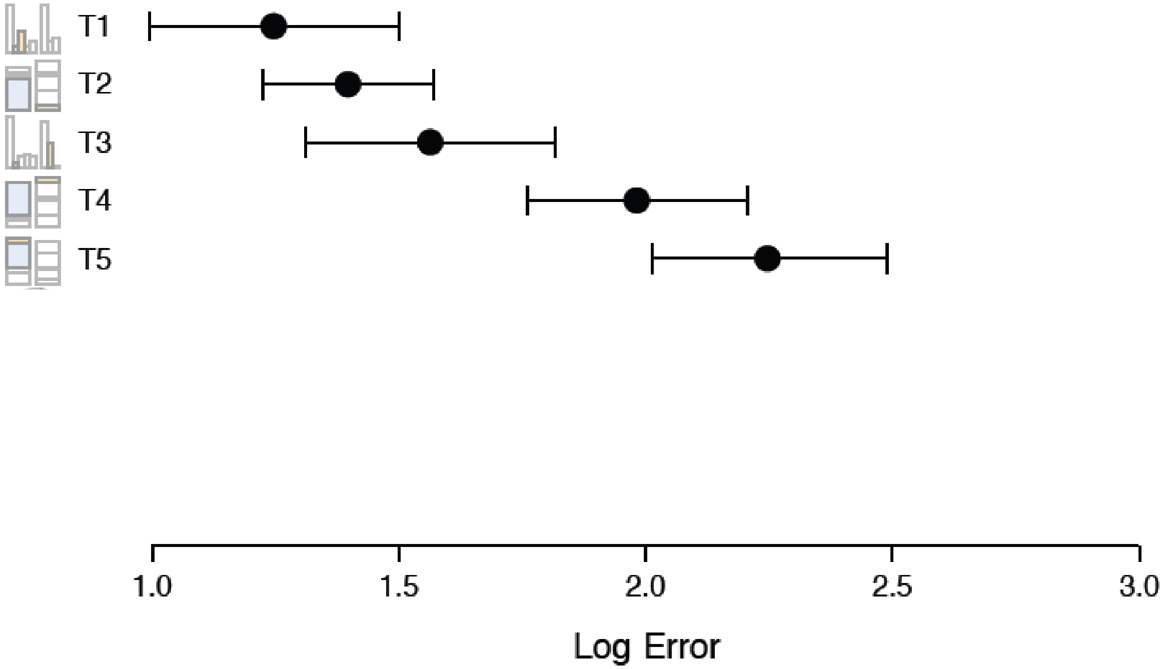
Our Crowdsourced Study



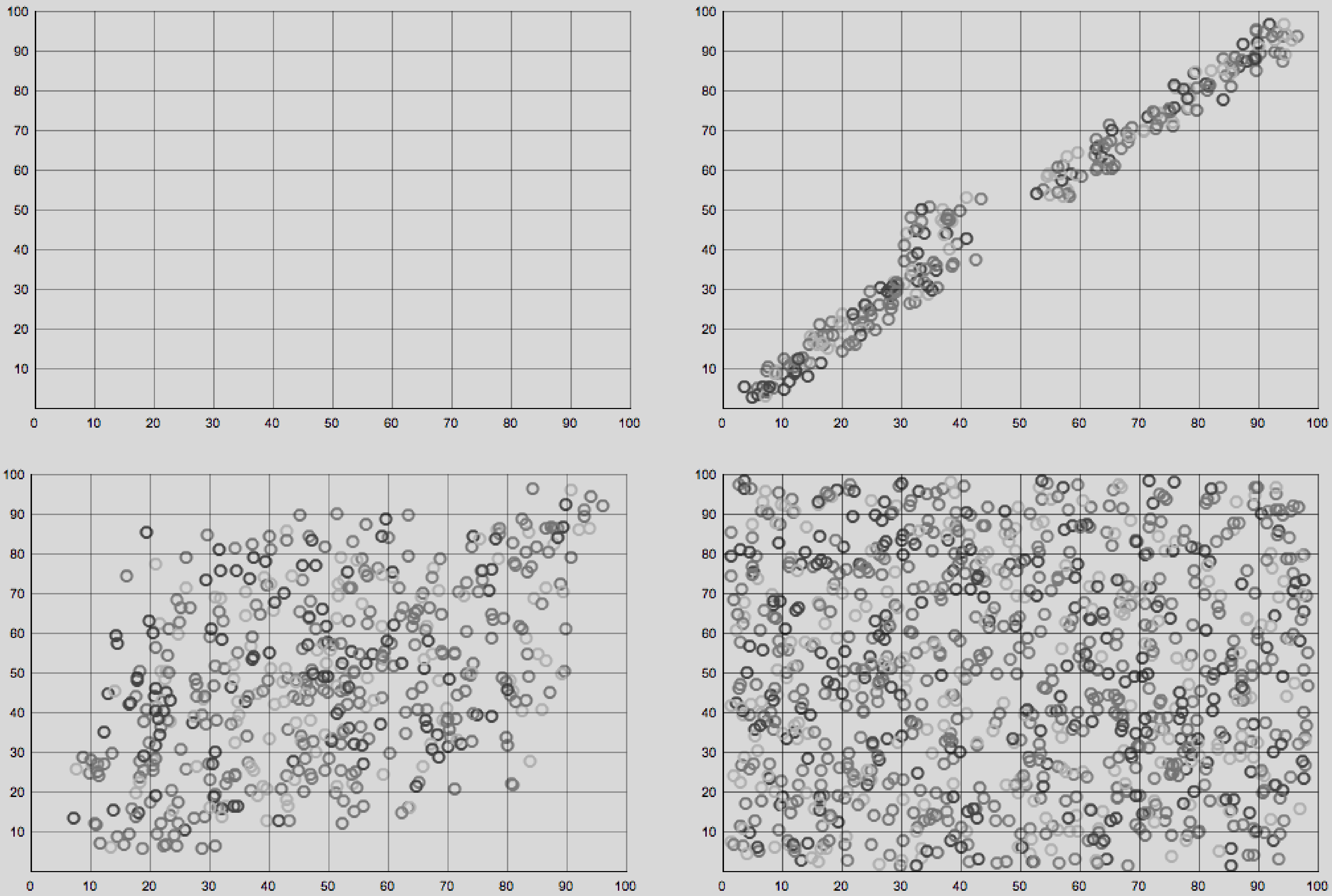
Cleveland & McGill, 1984 (Lab Study)



Our Crowdsourced Study



Experiment 2:
Gridline Alpha Contrast



100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100

Experiment 2 Tasks

2L: Adjust the grid so that it is as light as possible while still being usably perceptible.

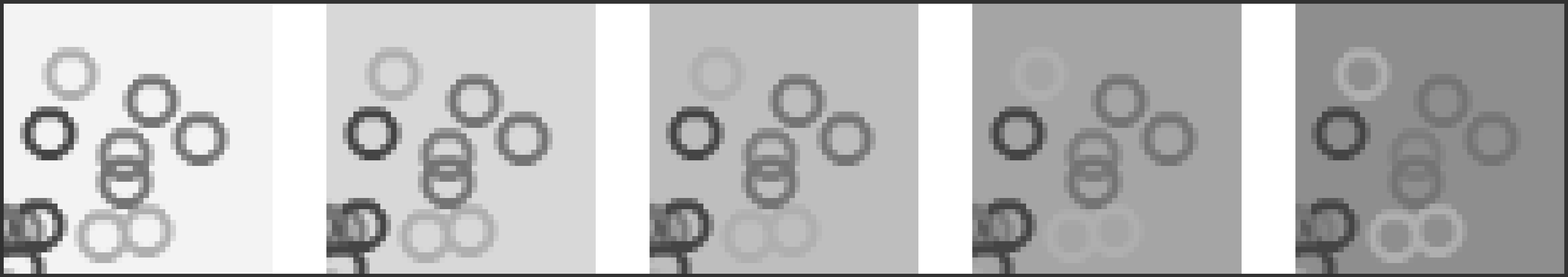
2D: Adjust the grid strength to meet your best judgment of how obvious it can be before it becomes too intrusive and sits in front of the image; some users have called this a ‘fence’.

Experiment 2: Gridline Alpha

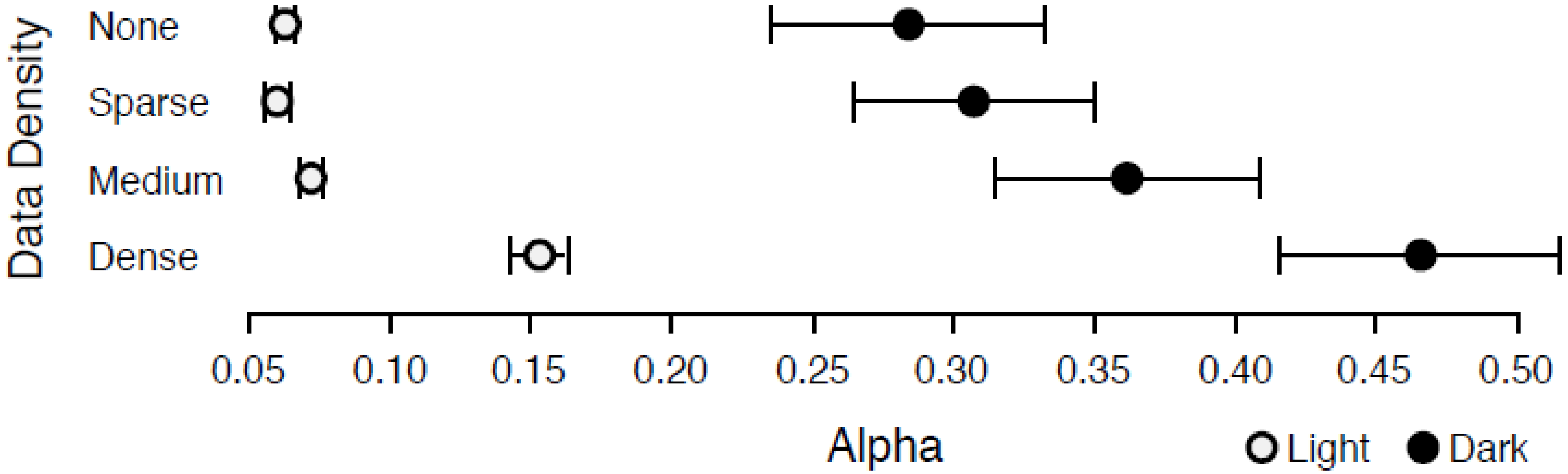
4 *plot density*: none, sparse, medium, dense
x 5 *background*: #f3, #d8, #be, #a5, #8e
x 3 *replications*

N=24 assignments, $0.02 per HIT

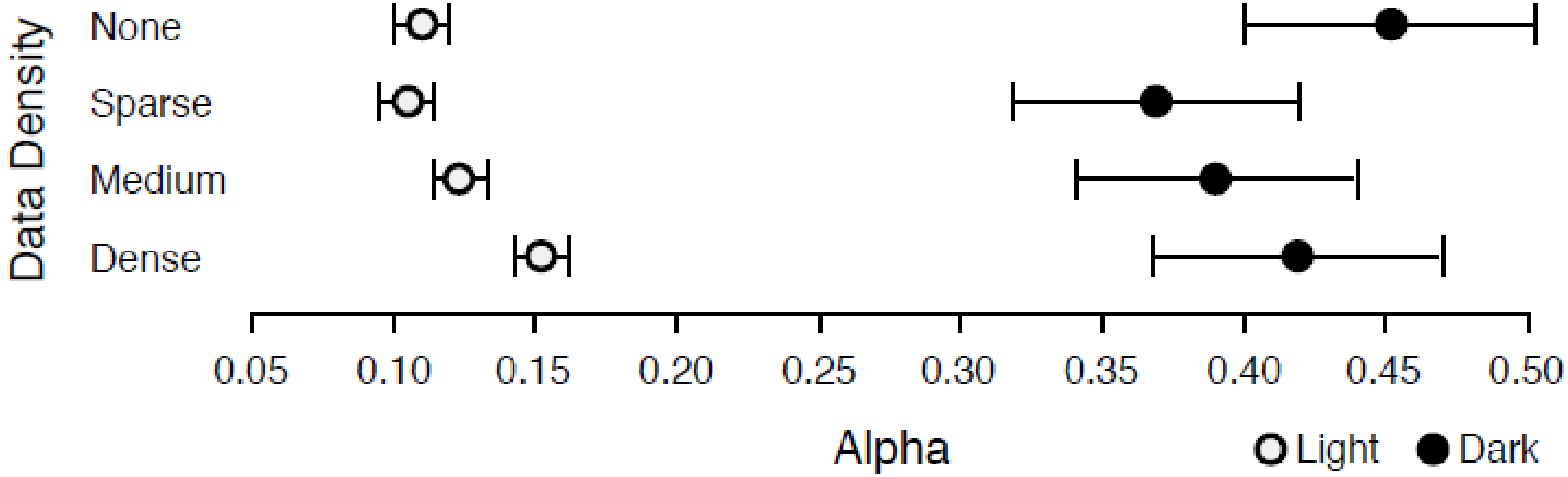
Record alpha value, User-Agent, JS “screen” info



Stone & Bartram, 2009 (Lab Study)

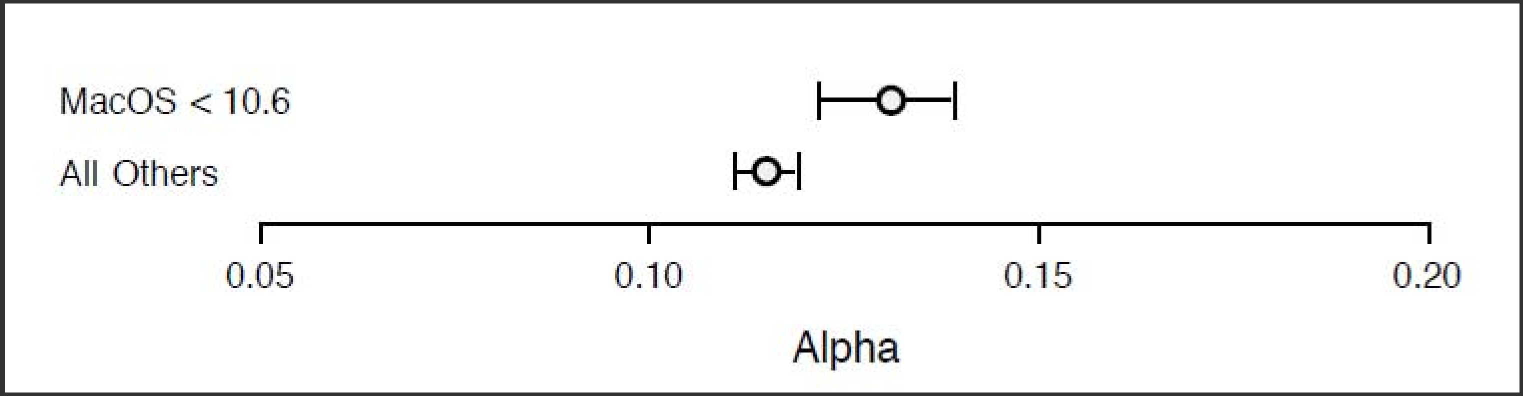


Our Crowdsourced Study



Inferred Display Configuration

Operating System (& gamma?) from User-Agent

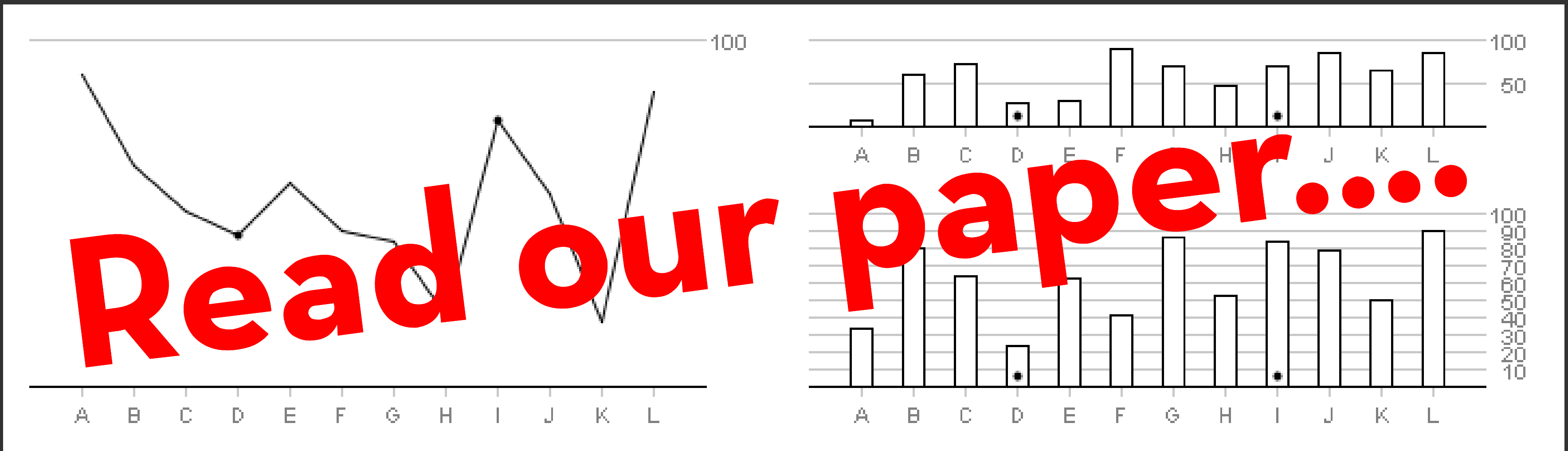


MacOS < 10.6: $\gamma = 1.8$ vs. PC: $\gamma = 2.2$

Alpha x pixel resolution: $r = 0.07$, $p < 0.01$

Alpha x color depth: $r = -0.18$, $p < 0.01$

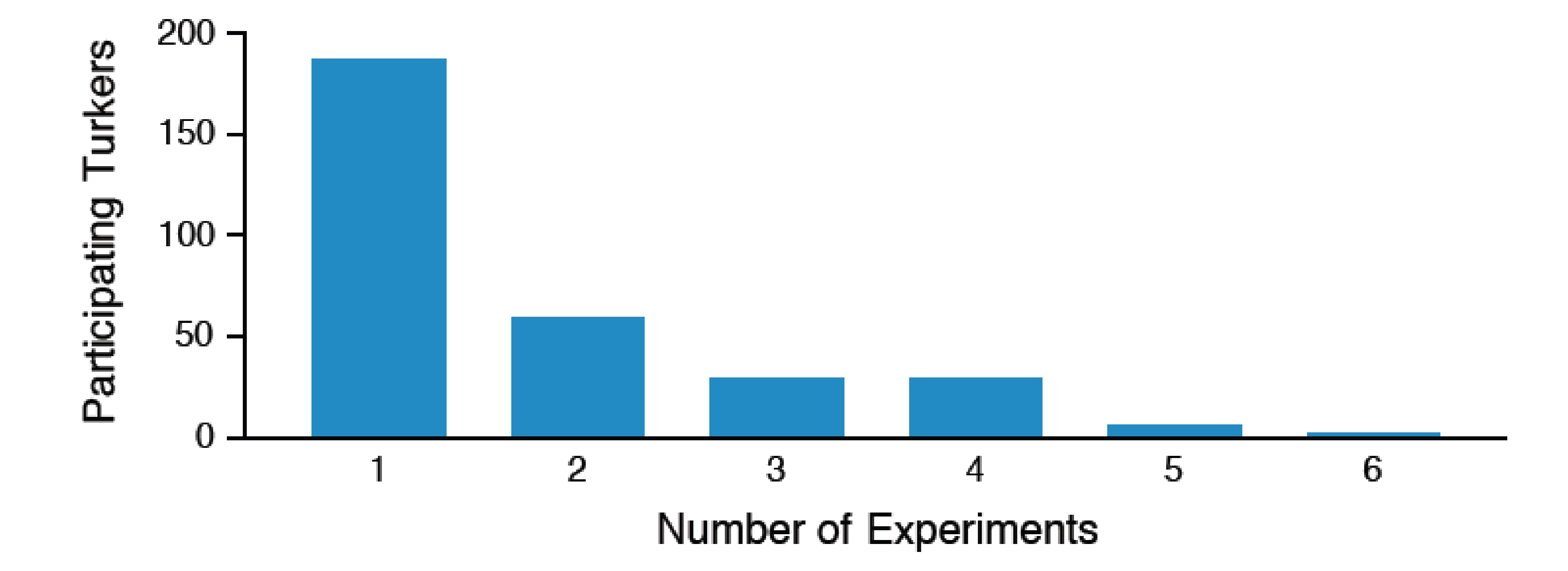
Experiment 3:
Chart Size & Gridline Spacing



Read our paper....

Mechanical Turk: Performance and Cost

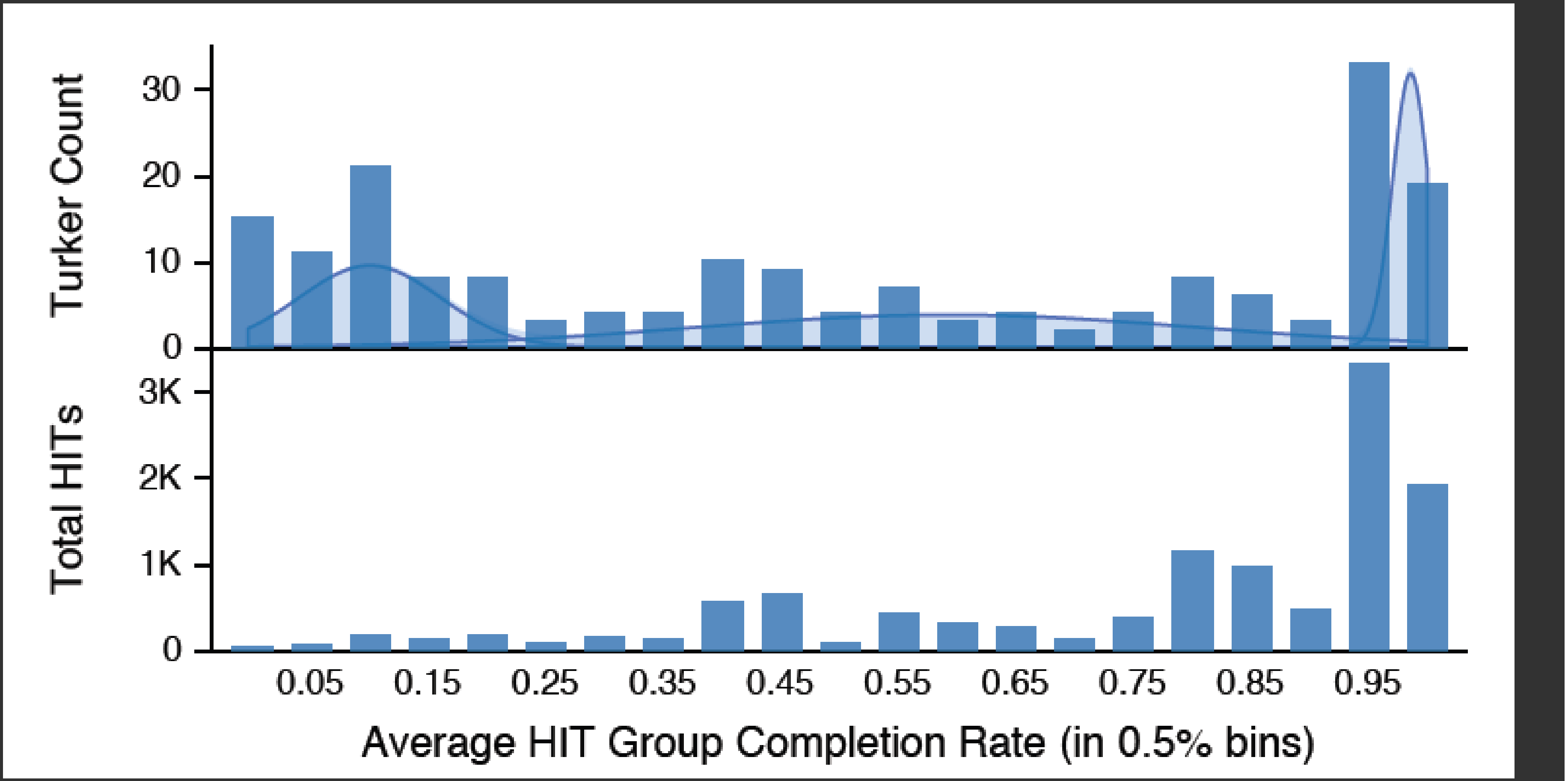
Turkers Overlap Across Studies



31% (51/186) Turkers participated in 2 or more

Only 7% (13) from Exp. 1A participated later

Samplers and *Streakers*



Quality with Qualification

High quality results: Only 0.75% of responses were rejected outliers.

Removing qualification tasks resulted in ***over* 10% unusable responses.**

Verifiable answers reduce gaming incentive and insincere responses.

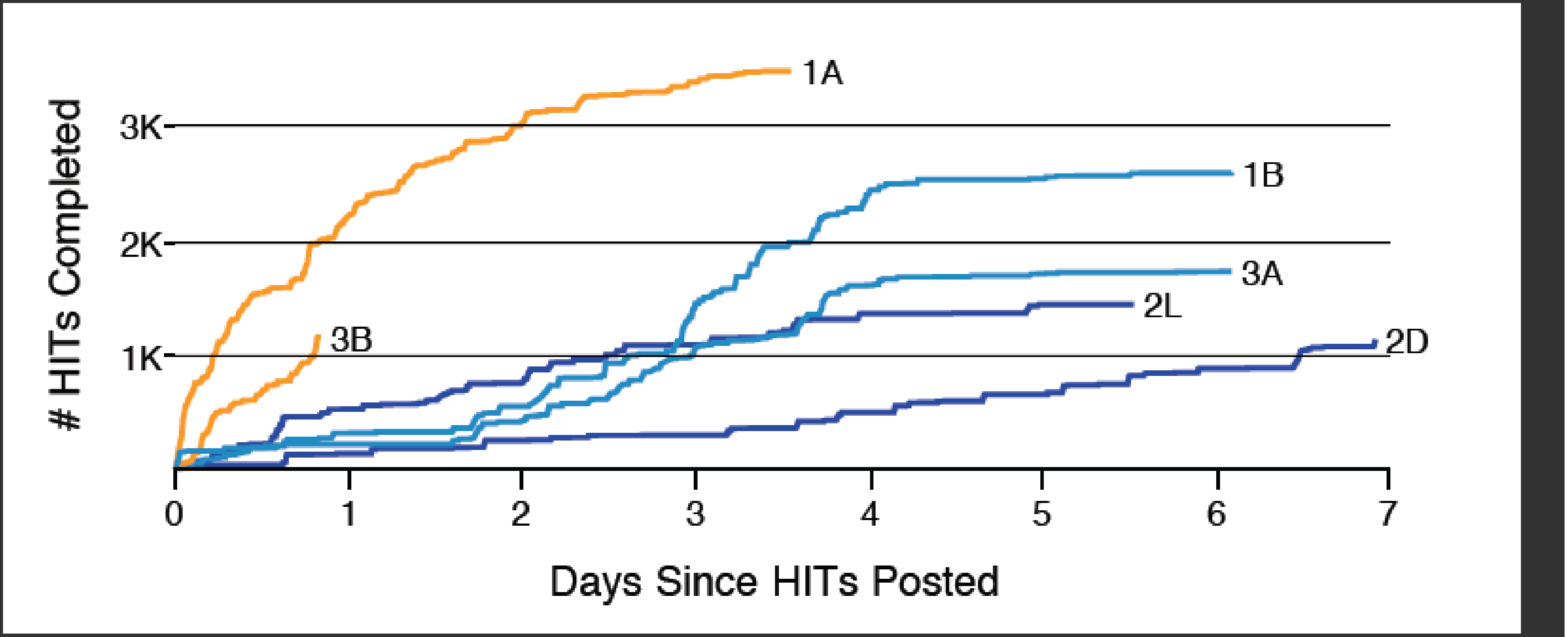
Standard HITs Frustrate Timing

Expected time per HIT: 10s
Observed time per HIT: 42s (μ=54s, σ=41s)
→ **Timing data is not reliable.**

Strategies for Fine-Grained Timing

- *Macro-Task* (batch of micro-tasks)
- *Ready-Set-Go* HIT interface
  → Successful in subsequent studies.

HIT Completion Rates



Orange ≥ 4¢ Blue = 2¢

Raise reward → faster results; ≅ quality

Crowdsourcing Reduces Costs

6x cost savings (vs. $15/subject lab rate)
9x savings possible (using $0.02 rewards)
Study time drops **from 2 weeks to 1-3 days**

→ **Crowdsourcing provides up to an order of magnitude $$ and time savings**

→ With constant cost, it enables **more studies, more variables, more subjects**

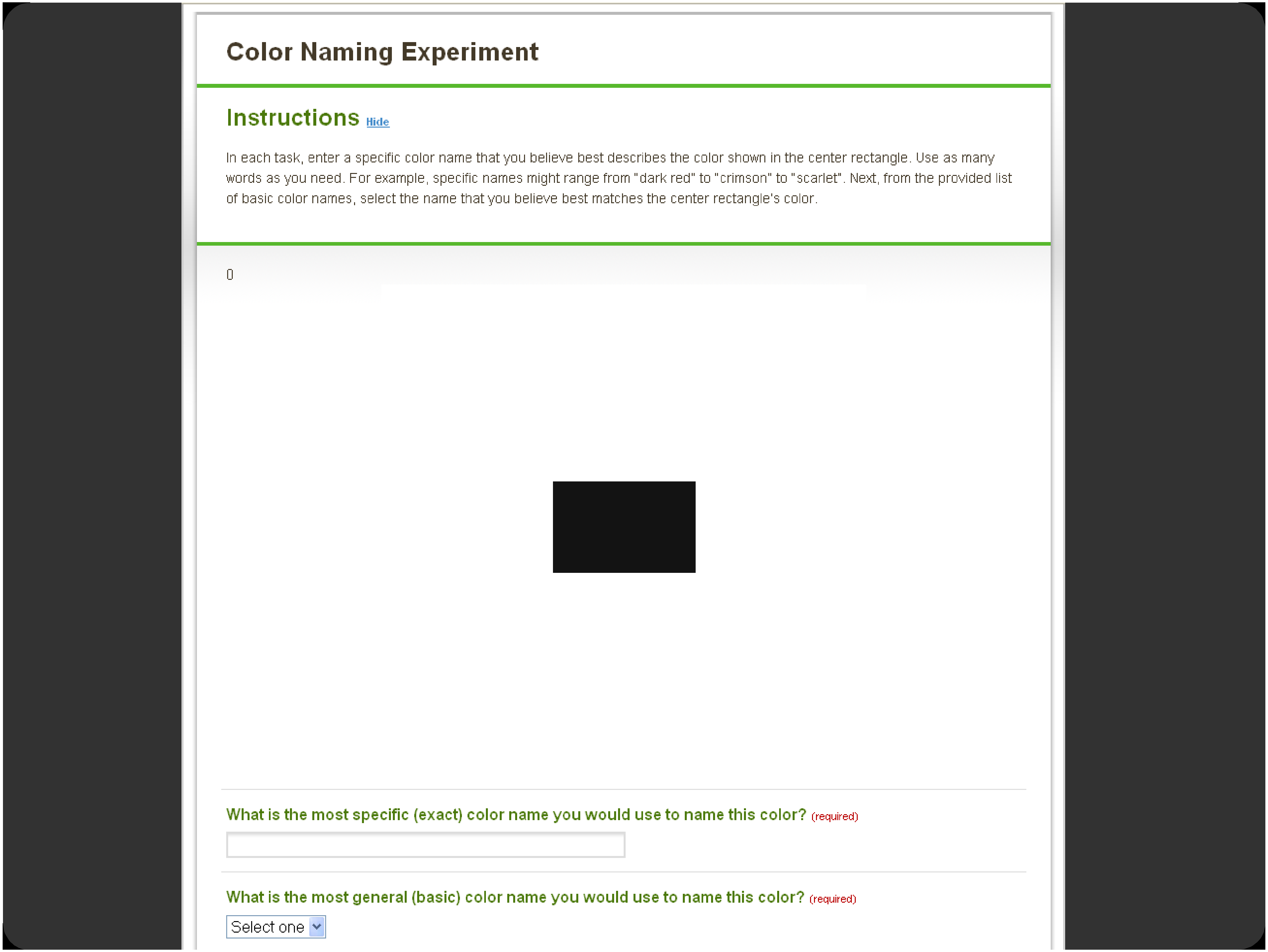
Future Work

Multiple methods studies: how to best balance the laboratory with online crowdsourcing?

Better tools for crowdsourced experimentation. Facilitate experimental control and adaptation.

Community resources for evaluation: share “market” data, share experimental designs, facilitate replication and meta-analysis.

Extend crowdsourcing methods to an even greater diversity of experimental designs.



Color Naming Experiment

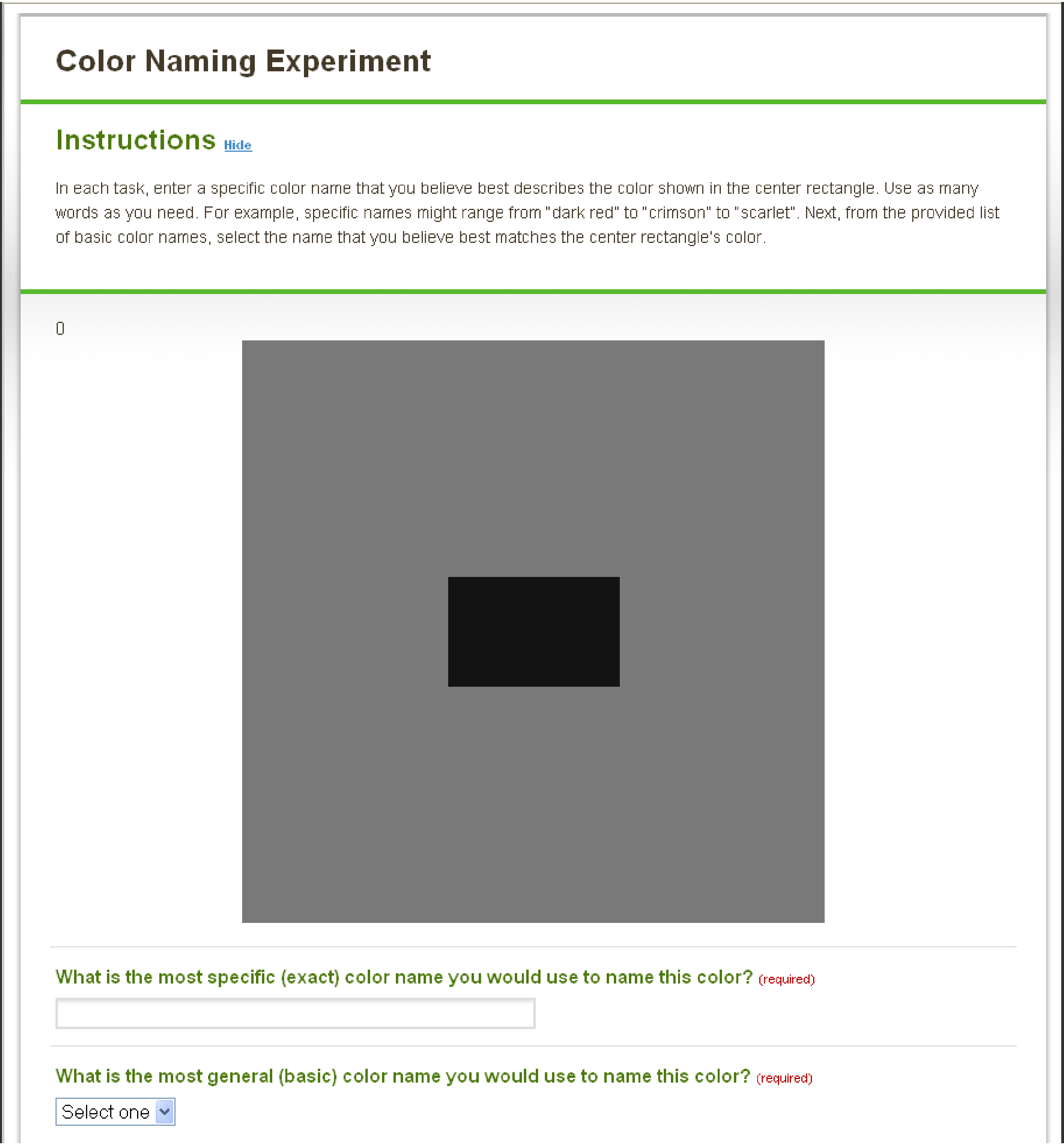
Instructions

In each task, enter a specific color name that you believe best describes the color shown in the center rectangle. Use as many words as you need. For example, specific names might range from "dark red" to "crimson" to "scarlet". Next, from the provided list of basic color names, select the name that you believe best matches the center rectangle's color.

0

What is the most specific (exact) color name you would use to name this color? (required)

What is the most general (basic) color name you would use to name this color? (required)



Color Naming Experiment

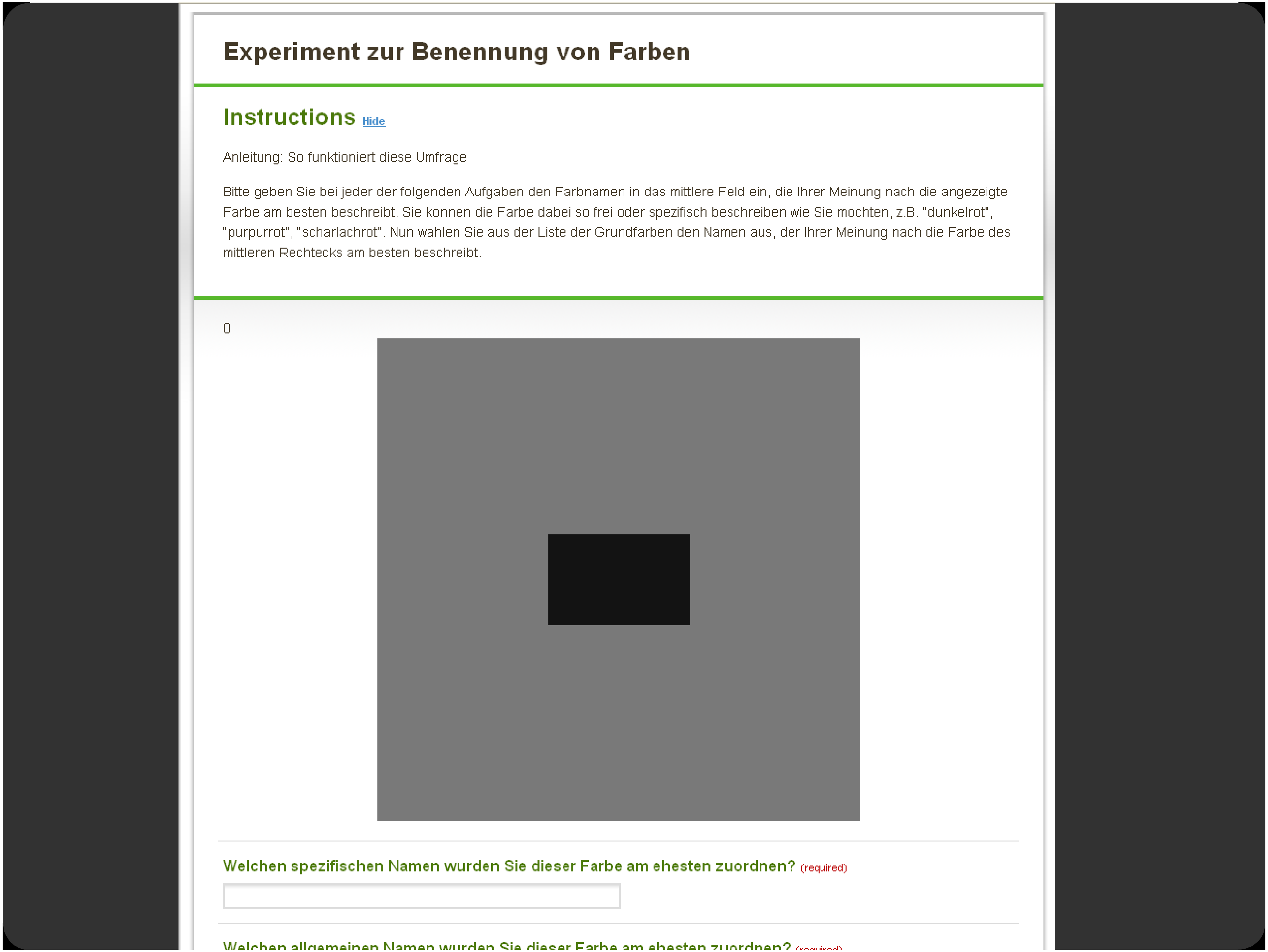
Instructions Hide

In each task, enter a specific color name that you believe best describes the color shown in the center rectangle. Use as many words as you need. For example, specific names might range from "dark red" to "crimson" to "scarlet". Next, from the provided list of basic color names, select the name that you believe best matches the center rectangle's color.

0

What is the most specific (exact) color name you would use to name this color? (required)

What is the most general (basic) color name you would use to name this color? (required)



Experiment zur Benennung von Farben

Instructions

Anleitung: So funktioniert diese Umfrage

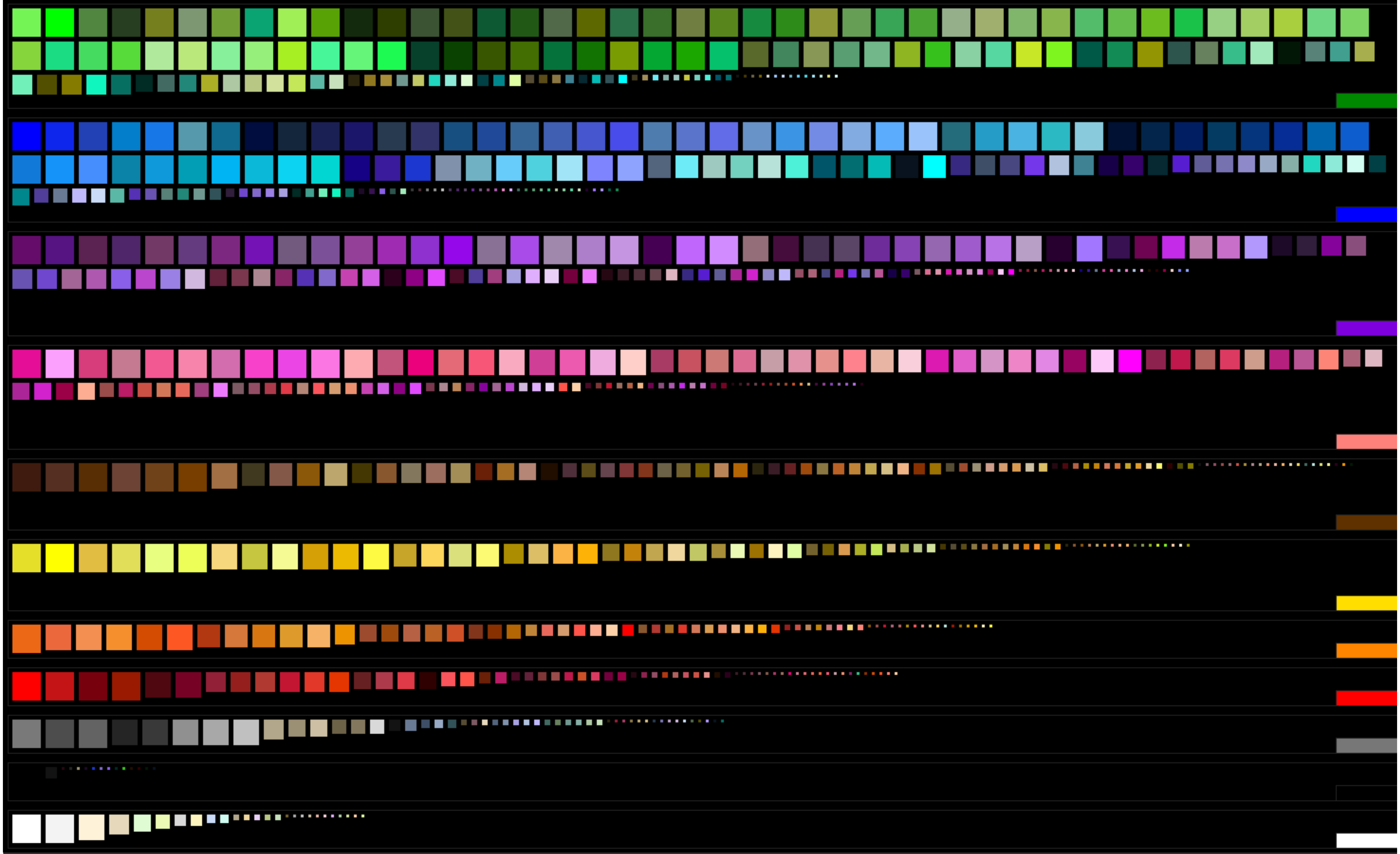
Bitte geben Sie bei jeder der folgenden Aufgaben den Farbnamen in das mittlere Feld ein, die Ihrer Meinung nach die angezeigte Farbe am besten beschreibt. Sie konnen die Farbe dabei so frei oder spezifisch beschreiben wie Sie mochten, z.B. "dunkelrot", "purpurrot", "scharlachrot". Nun wahlen Sie aus der Liste der Grundfarben den Namen aus, der Ihrer Meinung nach die Farbe des mittleren Rechtecks am besten beschreibt.

0

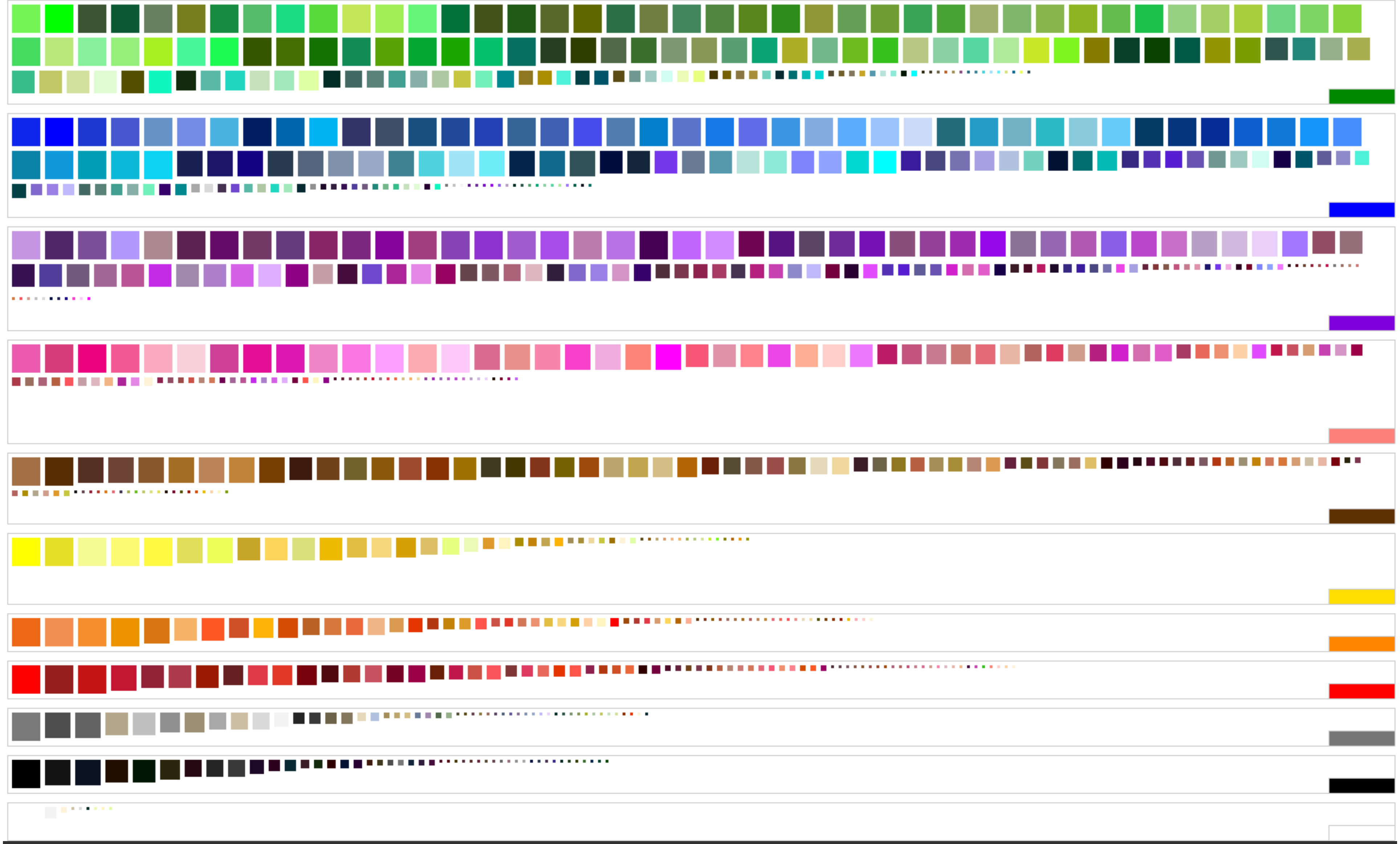
Welchen spezifischen Namen wurden Sie dieser Farbe am ehesten zuordnen? (required)

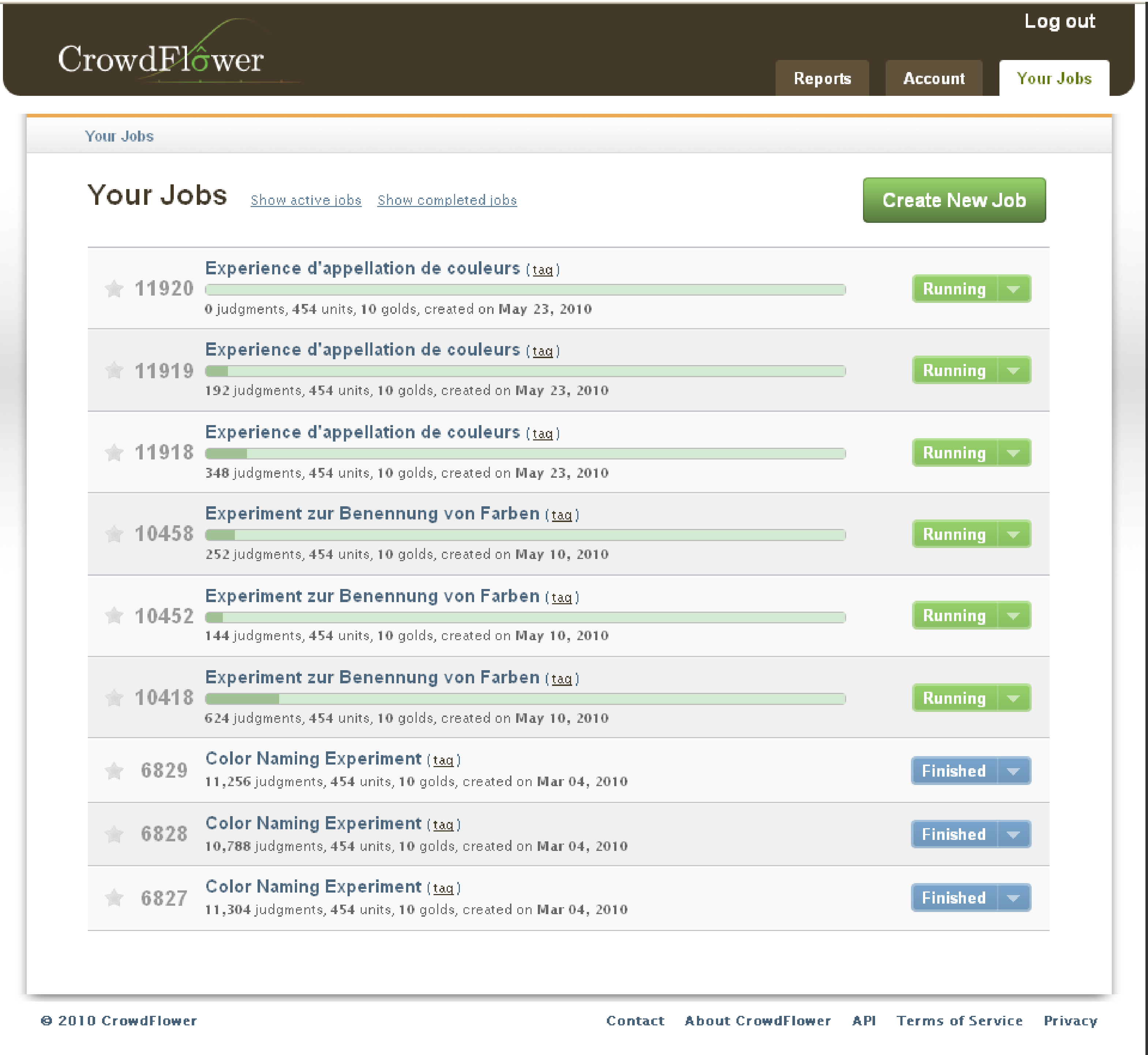
Welchen allgemeinen Namen wurden Sie dieser Farbe am ehesten zuordnen? (required)

Visualizing common color terms



Visualizing common color terms





Your Jobs

Your Jobs

Show active jobs Show completed jobs

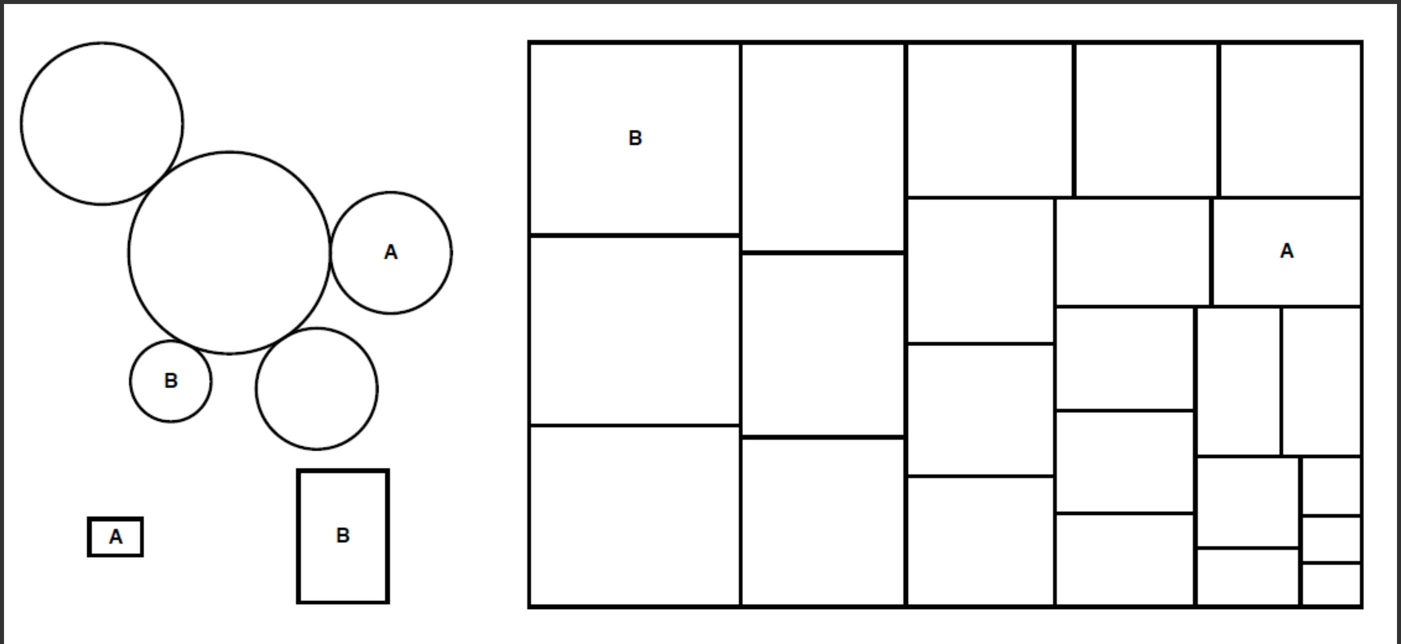
☆	11920	Experience d'appellation de couleurs (tag) 0 judgments, 454 units, 10 golds, created on May 23, 2010	Running ▾
☆	11919	Experience d'appellation de couleurs (tag) 192 judgments, 454 units, 10 golds, created on May 23, 2010	Running ▾
☆	11918	Experience d'appellation de couleurs (tag) 348 judgments, 454 units, 10 golds, created on May 23, 2010	Running ▾
☆	10458	Experiment zur Benennung von Farben (tag) 252 judgments, 454 units, 10 golds, created on May 10, 2010	Running ▾
☆	10452	Experiment zur Benennung von Farben (tag) 144 judgments, 454 units, 10 golds, created on May 10, 2010	Running ▾
☆	10418	Experiment zur Benennung von Farben (tag) 624 judgments, 454 units, 10 golds, created on May 10, 2010	Running ▾
☆	6829	Color Naming Experiment (tag) 11,256 judgments, 454 units, 10 golds, created on Mar 04, 2010	Finished ▾
☆	6828	Color Naming Experiment (tag) 10,788 judgments, 454 units, 10 golds, created on Mar 04, 2010	Finished ▾
☆	6827	Color Naming Experiment (tag) 11,304 judgments, 454 units, 10 golds, created on Mar 04, 2010	Finished ▾

© 2010 CrowdFlower

Contact About CrowdFlower API Terms of Service Privacy

Crowdsourcing Graphical Perception

Using Mechanical Turk to Assess Visualization Design



Jeffrey Heer & Michael Bostock

http://hci.stanford.edu/jheer