

A HOLISTIC VIEW OF HUMAN FACTORS IN CROWDSOURCING

Sihem Amer-Yahia

CNRS Research Director, France

Keynote at SBBD 2017

October 3rd, 2017



A definition

- The act of posting an open call to hire *cheap, immediate, skilled, and easily accessible* labor online
- A place where one finds work, possibly with *remuneration*
- Micro-tasks often easier to complete by humans than by machines

OTHER LAYERS [[HIDE](#)]

HOW TO REPORT

By using an app:

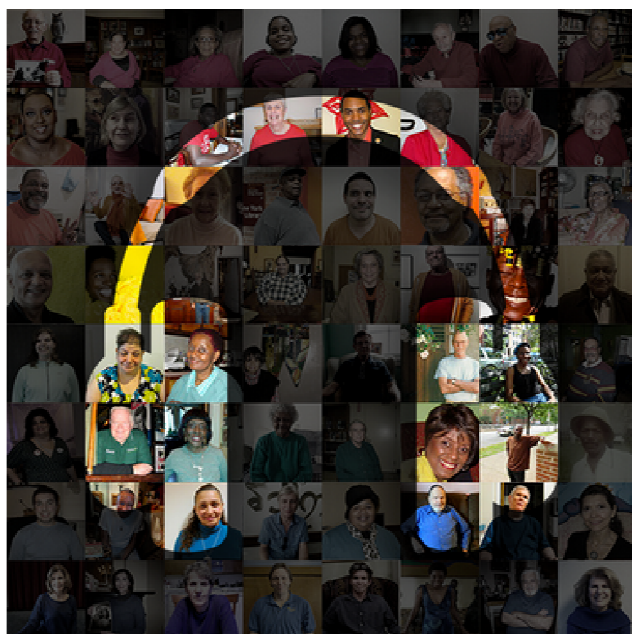
iPhone
Android

By sending an email:

2. Audio Transcription

NYPL Lab

Together We Listen



Help [The New York Public Library](#) fix computer-generated transcripts from hundreds of stories from the library's [Community Oral History Project](#).

You have edited this line


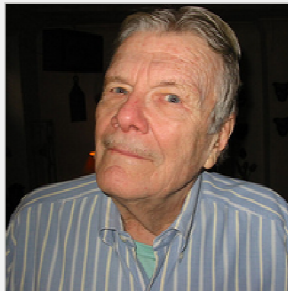


0:10	<input checked="" type="radio"/>	Yea- yeah. Um, [laughs] where to start?
	<input type="radio"/>	Uh, wh want to show
0:17	<input type="radio"/>	it's a community driven project

An example of editing a transcript

An example of how the transcript editor works (click for sound)

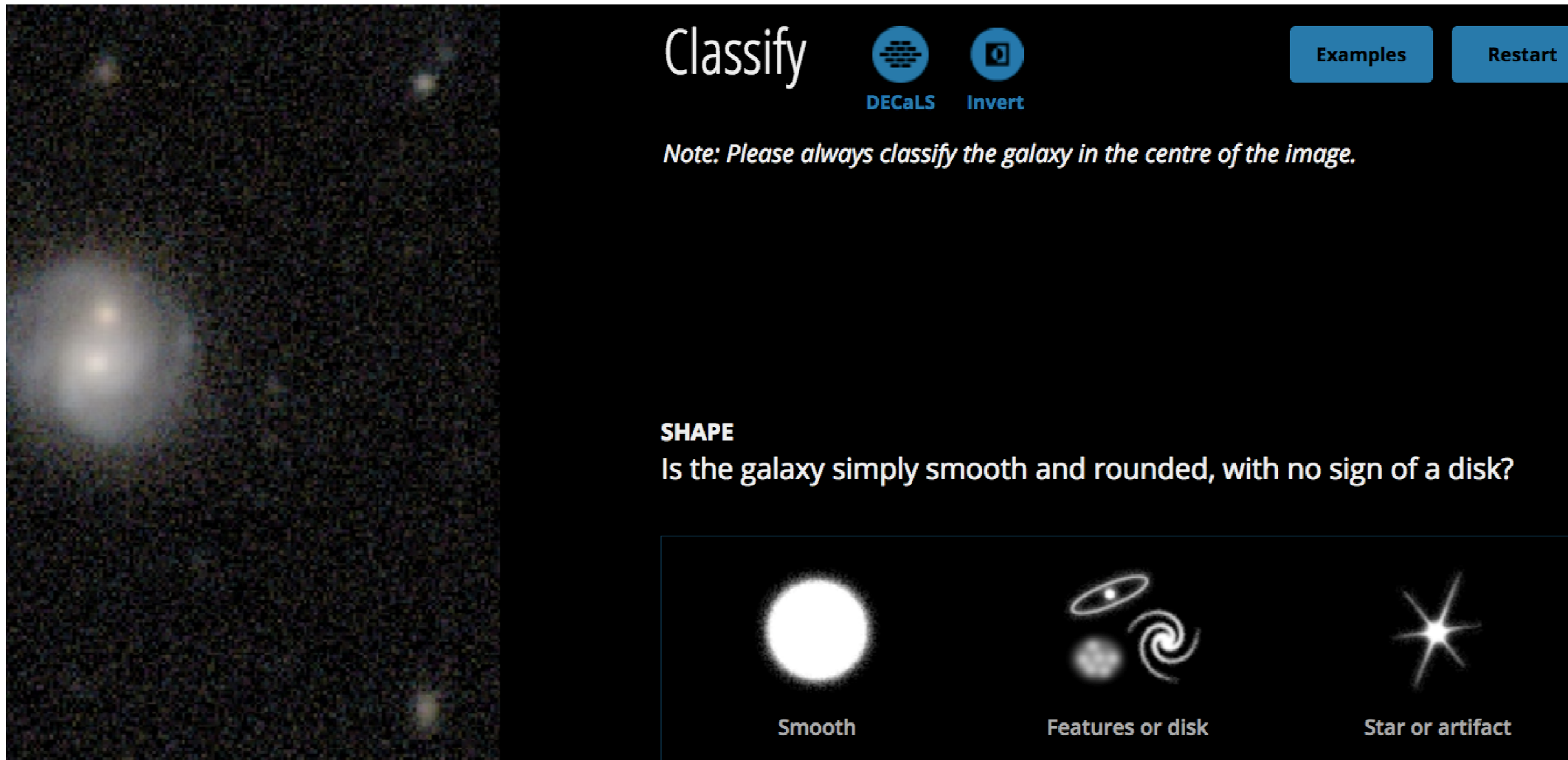
Select an interview to get started.

Filter by Collection: [All Collections](#) Sort by: [Title \(A to Z\)](#) Search Title/Description

 <p>VISIBLE LIVES Adam Payne Interviewed by Monica Diaz 57m 53 contributors 61% reached consensus</p>	 <p>YOUR VILLAGE, YOUR STORY Addis Williams Addis Williams, who began working in show business at age seven or eight, discusses his 1h 4m 43 contributors 34% reached consensus 2% awaiting review</p>	 <p>VOICES FROM EAST OF BRONX P... Adele Acampora Pasmantier Long-time Bronx resident Adele Acampora Pasmantier shares memories of her close-knit Italian 1h 10m 20 contributors 28% reached consensus 1% awaiting review 8% have edits</p>	 <p>A PEOPLE'S HISTORY OF HARLEM Aden Seraile Aden Seraile was born in Harlem where he lives now. He recalls the neighborhood's bad 31m 26 contributors 83% reached consensus</p>
--	--	---	---

3. Galaxy classification

Galaxy Zoo



Classify

DECaLS Invert

Examples Restart

Note: Please always classify the galaxy in the centre of the image.

SHAPE
Is the galaxy simply smooth and rounded, with no sign of a disk?

Smooth Features or disk Star or artifact

4. Receipt Transcription on AMT

KEYBOARD SHORTCUTS: Scroll: Shift + up/down [Open Image](#)



Classify Receipt

Hit Reward: \$0.02

Real readable original receipt

Not a receipt or not readable

The following details can often be found at the top or bottom of the receipt. Enter as much information as you can find.

Find and enter the business phone number:

Phone

Example: (888) 555-1234 or 8885551234

Find and enter the business address:

Address

City

State

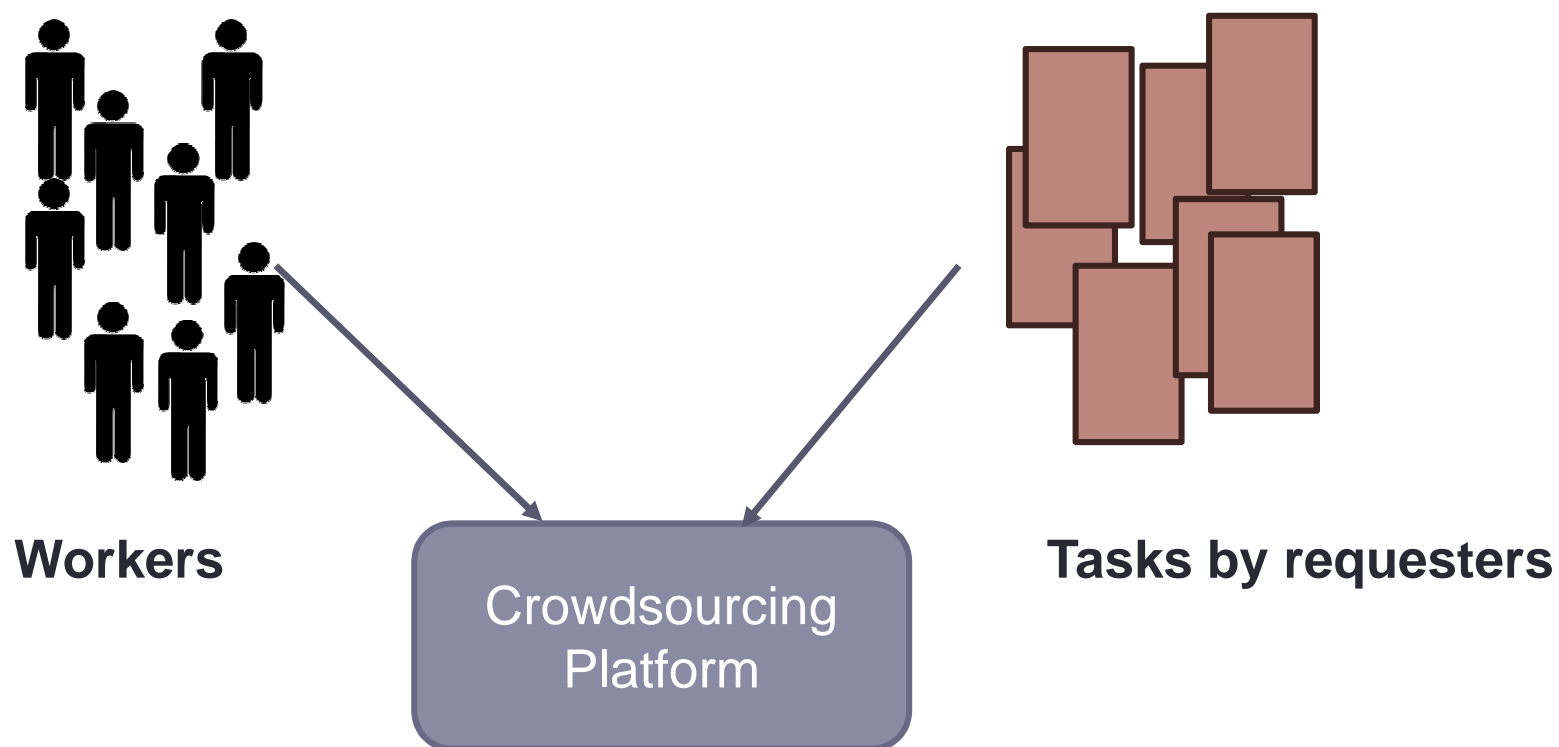
Postal code

Example: 321 Fake Street, Los Angeles, CA, 90210

Next



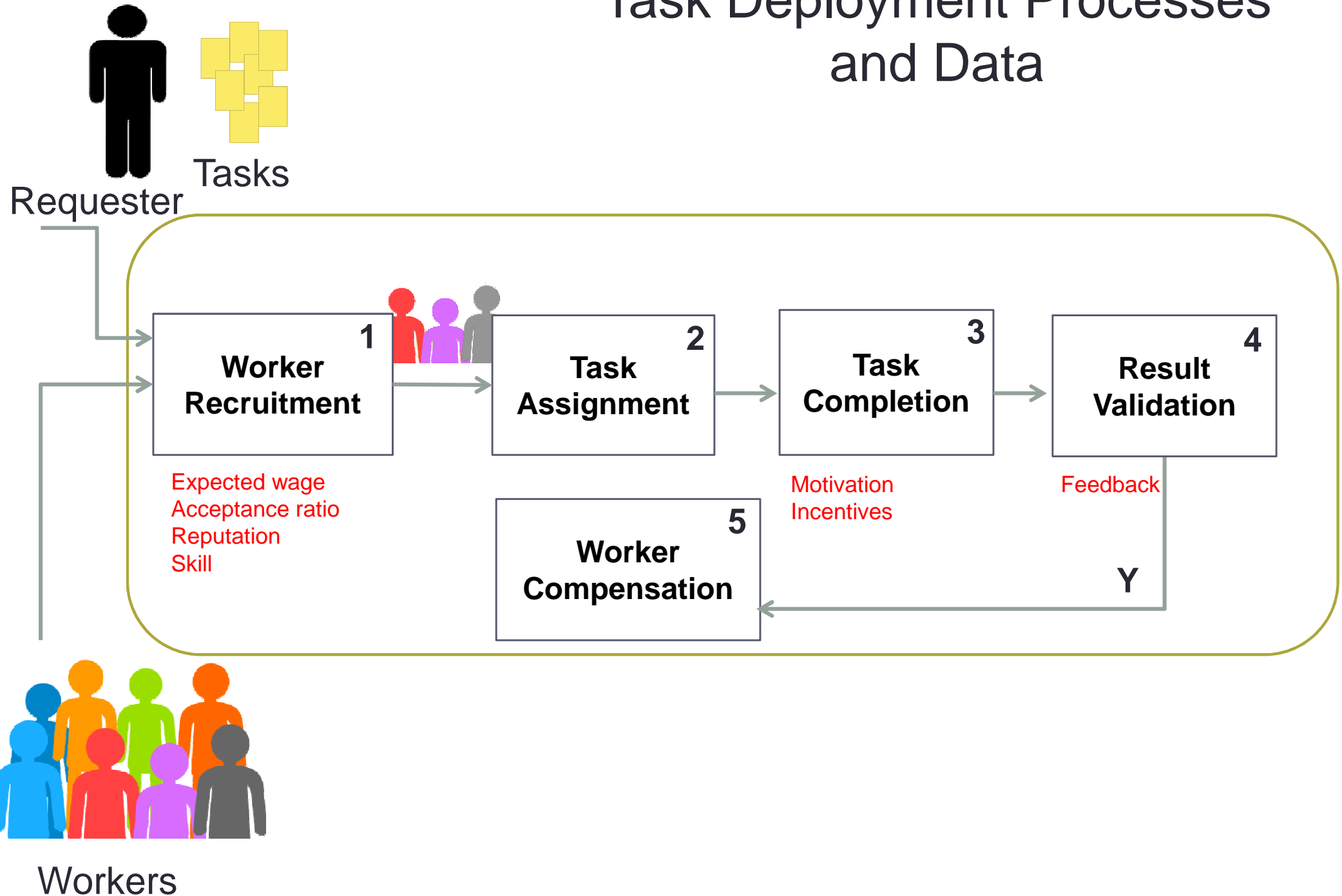
5. Generic Platforms



Many more examples

- Micro-tasks
 - audio transcription, text translation, image tagging, citizen science
 - implicit collaboration
 - consensus usually achieved with majority voting
- Collaborative tasks
 - a group of individuals *collectively working* to achieve a goal
 - collaborative editing, fan-subbing, solution outsourcing (e.g., Netflix contest)
 - Consensus achieved when crowd converges

Task Deployment Processes and Data



Human Factors, a rough characterization

1. Worker-specific

- *Micro-tasks: Skill, Reputation/Trust, Expected Pay, Acceptance Ratio*
- *Collaborative tasks: Affinity, Critical Mass, Interaction model*

1. Task-specific

- *Expected Quality, Budget, Desired Expertise, Incentives,*

2. Workers and tasks

- *Motivation, Feedback*

Human Factors

- They are pervasive in crowdsourcing processes
- Their acquisition, inference and evolution affect performance
- They have been studied in isolation: one process at a time and one factor at a time

This talk's purpose and outline

Advocate a holistic approach to human factors

- because of the unpredictability of humans as resources
 - because human factors are evolving in nature
-
- Reason 1: Human factors are unpredictable
 - Reason 2: Human factors evolve
 - Reason 3: Looking beyond

Human Factors Are Unpredictable

Self-appointment to tasks in AMT


[Sign In](#)

Your Account

HITs

Qualifications

190,184 HITs
available now

[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 1518 Results

Sort by:
[Show all details](#) | [Hide all details](#)

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

Items per Page:

CTRP: Type name, date and total of a receipt

[View a HIT in this group](#)
Requester: CopyText Inc.

HIT Expiration Date: Jan 14, 2016 (9 minutes 46 seconds) **Reward:** \$0.01

Time Allotted: 4 minutes

Improve a Transcript

[View a HIT in this group](#)
Requester: CastingWords

HIT Expiration Date: Jan 14, 2016 (10 minutes 50 seconds) **Reward:** \$0.19

Time Allotted: 8 hours

Identify groups announced from audio recording (Level 2)

[View a HIT in this group](#)
Requester: TestNotice

HIT Expiration Date: Jan 14, 2016 (19 minutes 59 seconds) **Reward:** \$0.25

Time Allotted: 10 minutes

Human Factors in Task Assignment

with Rahman et al. *VLDBJ* 2015

- **Input:** tasks to complete, human workers
- **Goal:** for one team per task
- **Output:** completed tasks
- Each task has *Expertise, Quality, Budget*
 - *English comprehension* for audio transcription
- Each worker has human factors: *Skill, Expected wage, Acceptance ratio*

Objective: maximize crowd-work quality

$$\text{Maximize } \mathcal{V} = \sum_{t \in T} v_t$$

aggregated worker skills and wages

task quality constraint

$$v_t = \begin{cases} W_1 \times \sum_{j \in \{1..m\}} q_{t_j} + W_2 \times (1 - \frac{w_t}{W_t}) & \text{if } q_{t_j} \geq Q_{t_j} \\ & \wedge w_t \leq W_t \\ 0 & \text{if } q_{t_j} < Q_{t_j} \\ & \vee w_t > W_t \end{cases}$$

task budget

where $W_1, W_2 \geq 0$ and $W_1 + W_2 = 1$.

ILP formulation of task quality

worker selected or not

worker's acceptance ratio

worker's skill

$$q_{t_j} = \sum_{u \in U} u_t \times p_u \times u_{s_j} \geq Q_{t_j}, \forall j \in \{1..m\}$$
$$w_t = \sum_{u \in U} u_t \times p_u \times w_u \leq W_t$$
$$u_t = [0/1]$$
$$X_l \leq \sum_{t \in T} \{u_t\} \leq X_h$$

bounds on number of tasks

The diagram illustrates the ILP formulation of task quality. It features three equations and one constraint, with arrows pointing from descriptive text to specific variables or terms within the equations. The first equation, $q_{t_j} = \sum_{u \in U} u_t \times p_u \times u_{s_j} \geq Q_{t_j}, \forall j \in \{1..m\}$, is annotated with 'worker selected or not' pointing to u_t , 'worker's acceptance ratio' pointing to p_u , and 'worker's skill' pointing to u_{s_j} . The second equation, $w_t = \sum_{u \in U} u_t \times p_u \times w_u \leq W_t$, is also annotated with 'worker selected or not' pointing to u_t and 'worker's acceptance ratio' pointing to p_u . The third equation, $u_t = [0/1]$, defines the binary nature of the selection variable. The final constraint, $X_l \leq \sum_{t \in T} \{u_t\} \leq X_h$, is annotated with 'bounds on number of tasks' pointing to the summation term.

Task Assignment Solution Overview

NP-hard (reduction using Multiple-Knapsack Problem)

- Our approach: offline indexing for a workload of tasks
 - Our implementation uses the primal-dual barrier method to solve the ILP
- Solution:
 - A greedy randomized algorithm with a $2/5$ approximation factor when objective function is sub-modular
 - A greedy deterministic algorithm with a $1-1/e$ approximation factor when objective function is sub-modular and monotonic

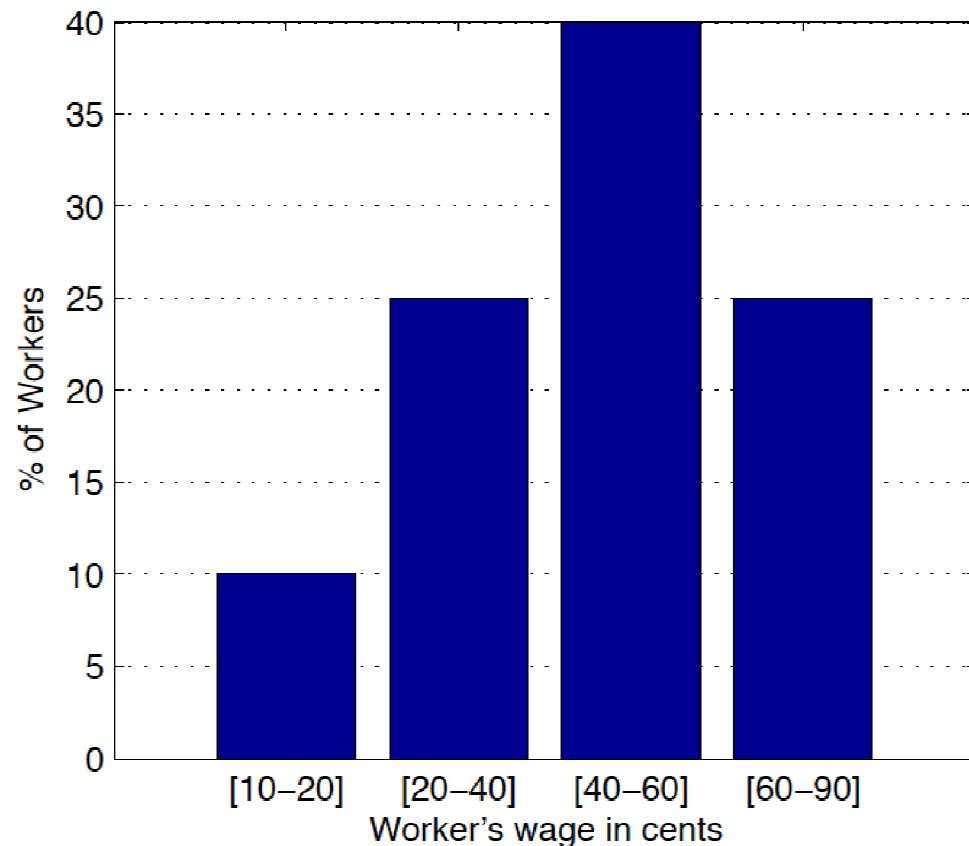
Human Factors are unpredictable

- In practice:
 - Workers may not accept tasks assigned to them
 - Worker and Task churn
- Challenges:
 - How to replace a worker who is not available for a task?
 - How to handle new workers/tasks?
- Our approach: Online index maintenance
 - Solve a marginal ILP problem (on a smaller instance)

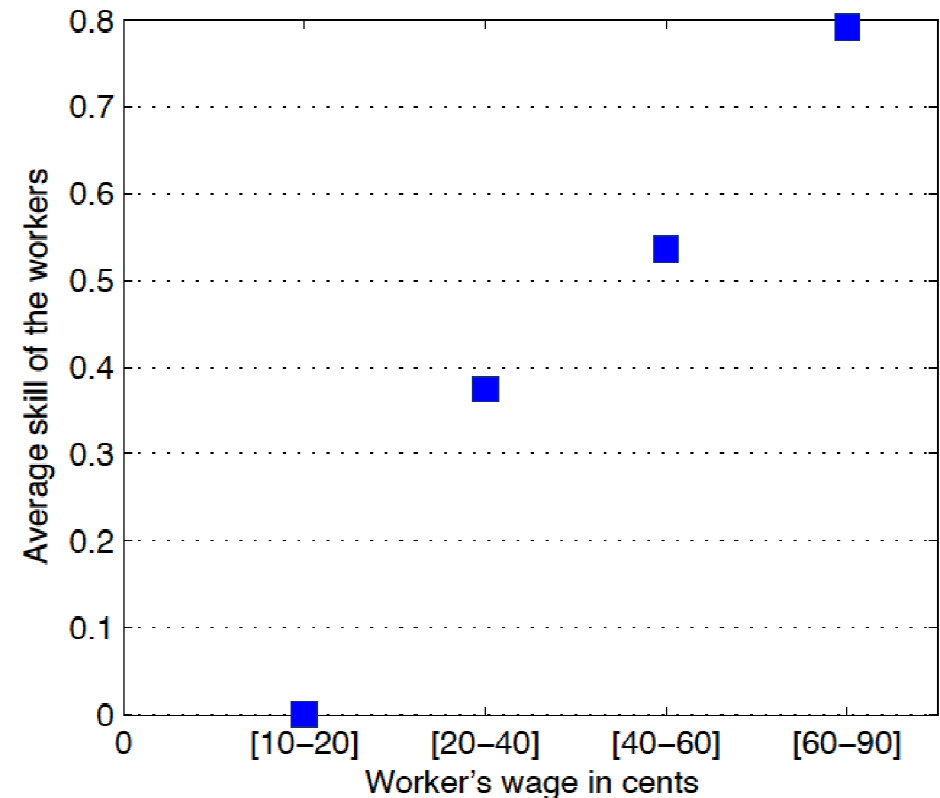
Quality Experiments

- **Phase 1:** 8 multi-choice questions/task, to assess skills
- **Phase 2:** Collaborative Document Editing task
 - 20 workers asked to produce reports on 5 different topics:
 - 1) *Political unrest in Egypt,*
 - 2) *NSA document leakage,*
 - 3) *Playstation games,*
 - 4) *All electric cars*
 - 5) *Global warming*
- **Phase 3:** Completed tasks evaluated by crowd workers
 - 150 AMT workers evaluated Completeness, Grammar, Neutrality, Clarity, Timeliness, Added-Value

AMT Worker Distributions (Egypt task)



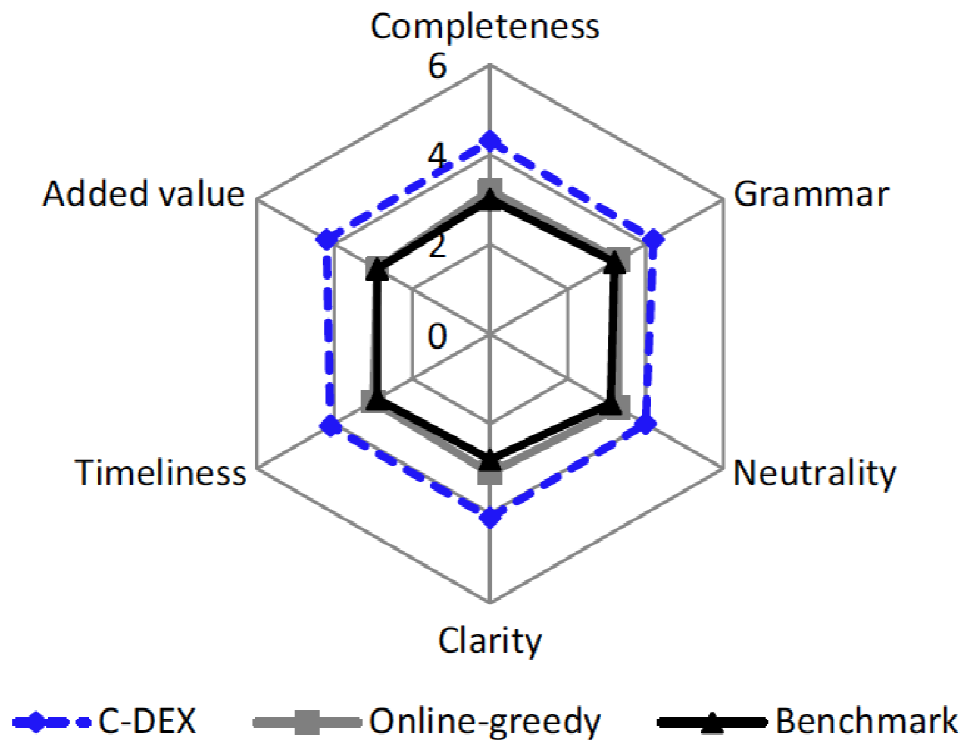
(c) Wage distribution



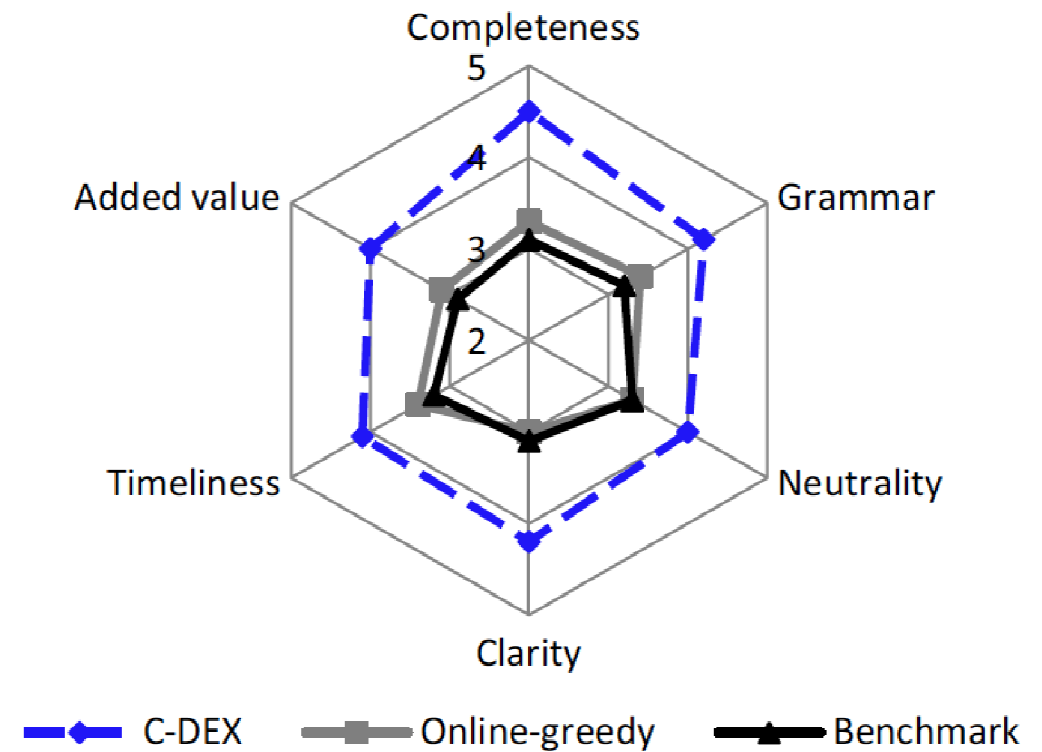
(d) Strong positive correlation between worker skill and wage

Outcome Quality Assessment

Playstation Games



Egypt Political Unrest



Group-aware Human Factors

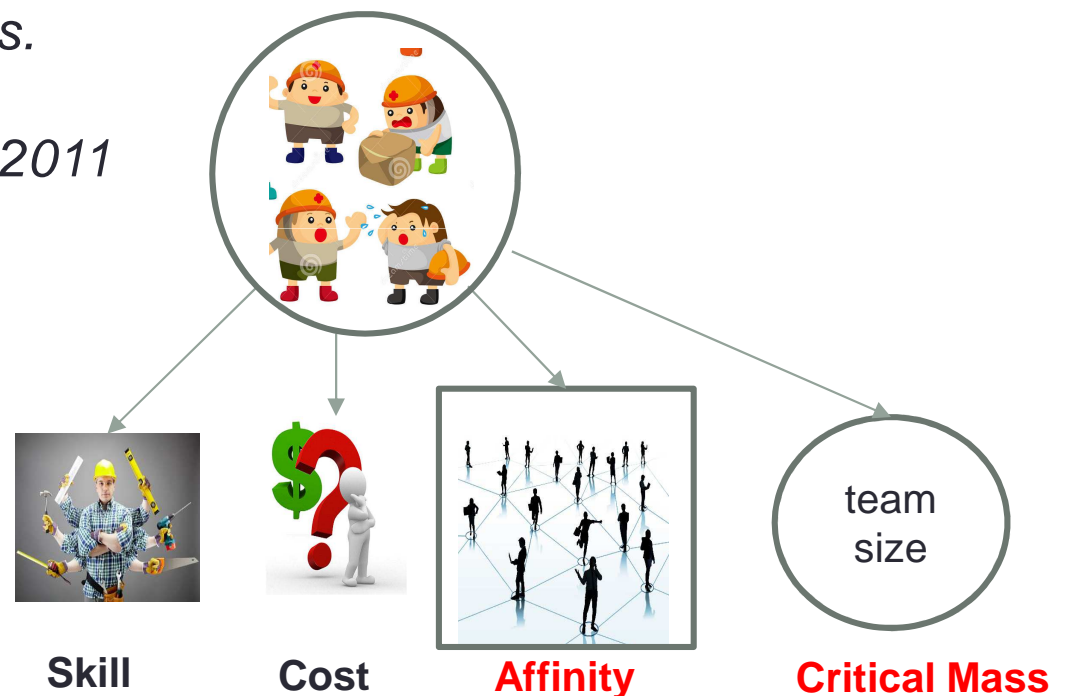
with Rahman et al. *ICDM 2015*

- In some cases, outcome quality was low
 - Conflicting opinions
 - Edit wars

Synergetic effects in working teams.

G. Hertel and G. Hertel.

Journal of Managerial Psychology 2011



Affinity

- *Type Indicator: MBTI. Myers and Briggs. Consulting Psychologists Press 1988*
- *Are two heads better than one? Crowdsourced translation via a two-step collaboration of non-professional translators and editors. R. Yan et. al. ACL 2014*

- Intra-team distance: e.g., edit wars

$$DiaDist(\mathcal{G}) = \text{Max}_{\forall u_i, u_j \in \mathcal{G}} \text{dist}(u_i, u_j)$$

Critical Mass

Managing research quality: critical mass and optimal academic research group size. R. Kenna et. al. IMA Journal of Management Mathematics 2012

Objective, revisited

$$\text{Minimize } \{DiaDist(\mathcal{G}) + \sum_{\forall G_i, G_j \in \mathcal{G}} SumInterDist(G_i, G_j)\}$$

- Under:

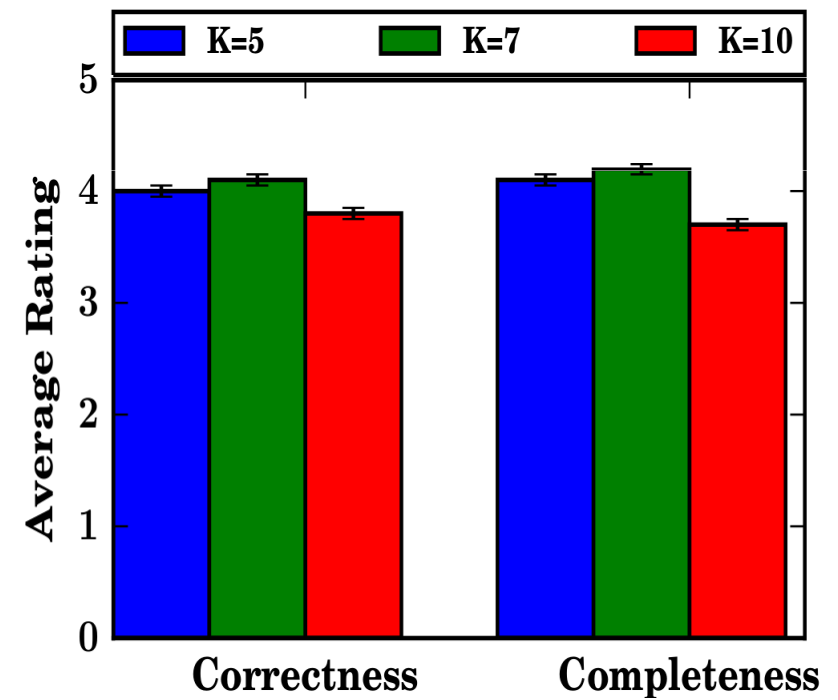
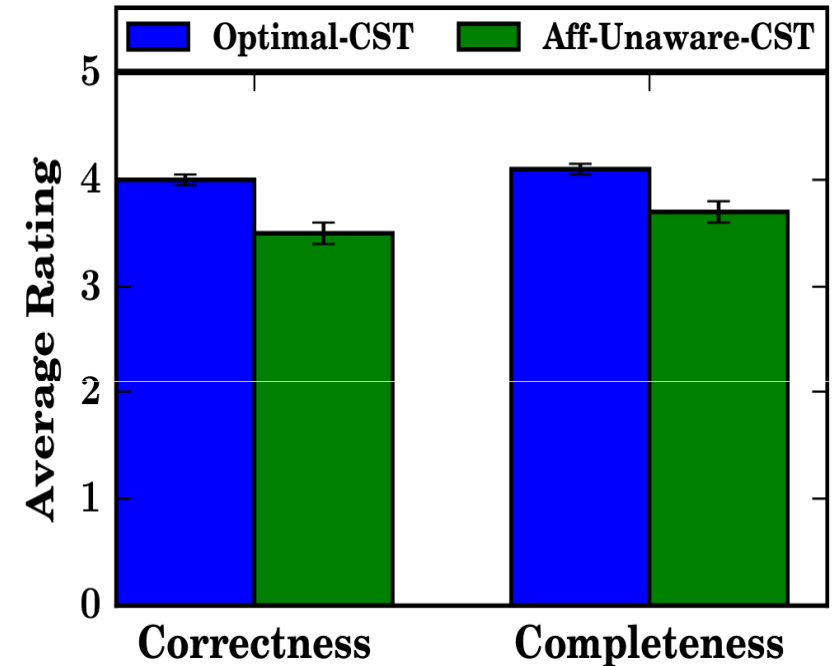
$$\begin{aligned} \sum_{\forall u_i \in \mathcal{G}} u_{d_i} &\geq Q_i \quad \forall d_i && \text{skill} \\ \sum_{\forall u \in \mathcal{G}} w_u &\leq C && \text{cost} \\ |G_i| &\leq K \quad \forall i = \{1, 2, \dots, x\} && \text{Critical mass} \end{aligned}$$

Overview of Algorithmic Solutions

- A two-stage approach
 1. Form a single team that **maximizes intra-affinity**, and satisfies skill and cost (*NP-hard, reduction of Min-Dia, a variant of Compact Location*)
 2. Decompose into smaller teams, each satisfies critical mass and **inter-affinity across teams is maximized** (*NP-hard, reduction of Minimum Bisection*)
- Algorithms
 1. An instance optimal exact algorithm and a 2-approximation algorithm (when distance is a metric)
 2. A 3-approximation algorithm (akin to Min k-cut)

Experiments with Affinity and Critical Mass

- Translation task with 120 AMT workers
- Region- and age/gender-based affinities
- Results
 - Higher affinity impacts positively quality
 - A group beyond size 10 is less effective
 - Region more effective than age/gender





Human Factors Evolve

In practice...

- Workers are involved in a series of tasks
- Their motivation evolves over time

Motivation in the Social Sciences

Motivation through the design of work: Test of a theory. **J. Hackman and G. R. Oldham.** *Organizational behavior and human performance*, 1976

- 658 employees in 62 heterogeneous jobs (*white collar, blue collar, industry, services, urban and rural settings*) in 7 organizations.
- Goal: which Job Dimensions stimulate which Psychological States:
experienced meaningfulness of the work, experienced responsibility for the work outcomes, knowledge of the actual results of the work.
- Proposed model good for job design, i.e.
 - in determining the potential of a job to engender motivation,
 - in identifying which jobs need improvement,
 - in assessing the readiness of employees to respond to a redesigned job

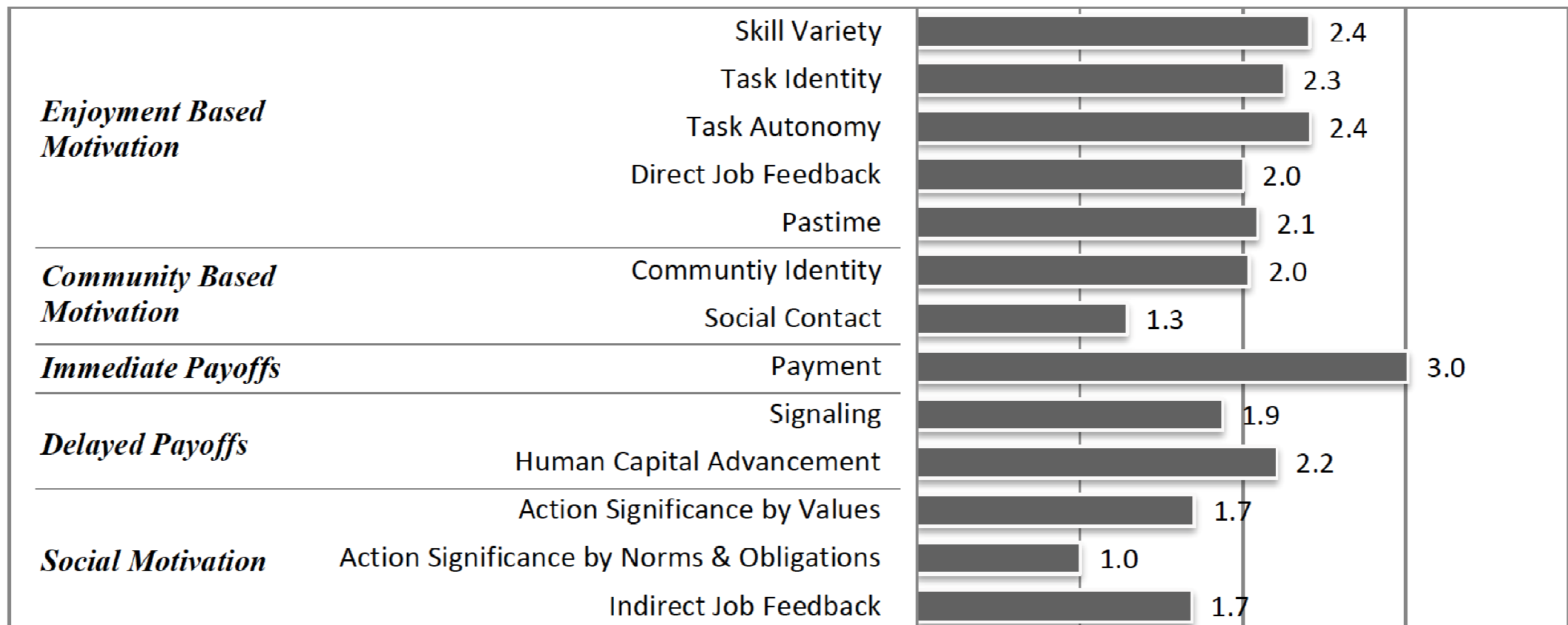
Motivation in the Social Sciences

Motivation through the design of work: Test of a theory. **J. Hackman and G. R. Oldham.** *Organizational behavior and human performance*, 1976

$$\text{Motivating Potential Score (MPS)} = \left[\frac{\text{Skill Variety} + \text{Task Identity} + \text{Task Significance}}{3} \right] \times \text{Autonomy} \times \text{Feedback}$$

What is motivation (in AMT)?

*More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. N. Kaufmann, T. Schulze, and D. Veit. **AMCIS 2011***

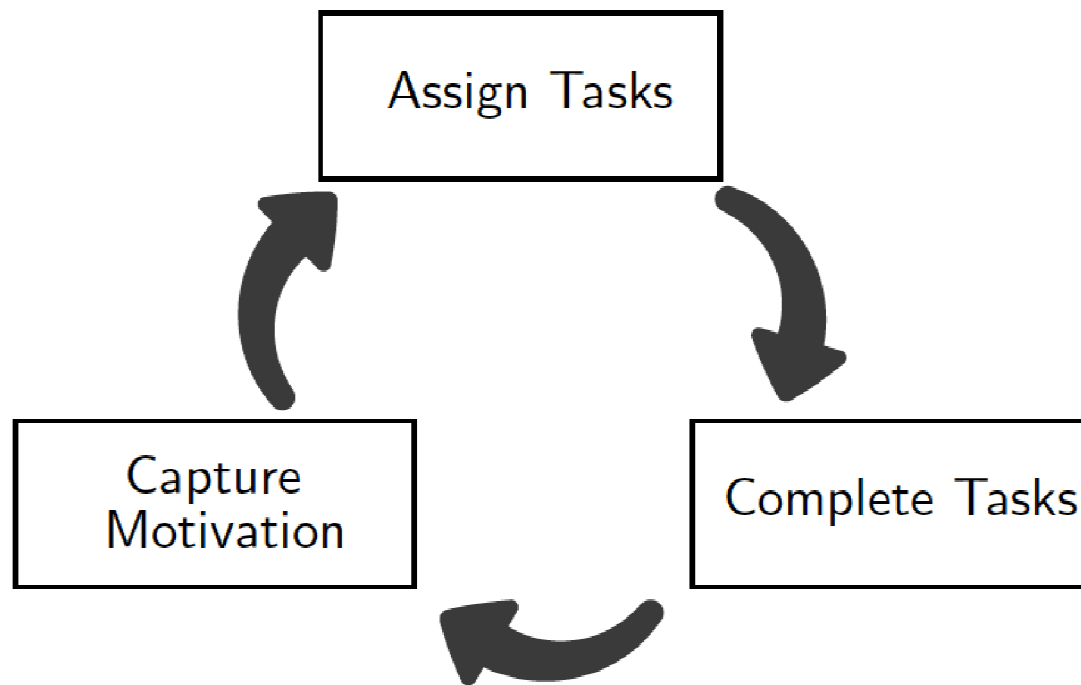


In the related work

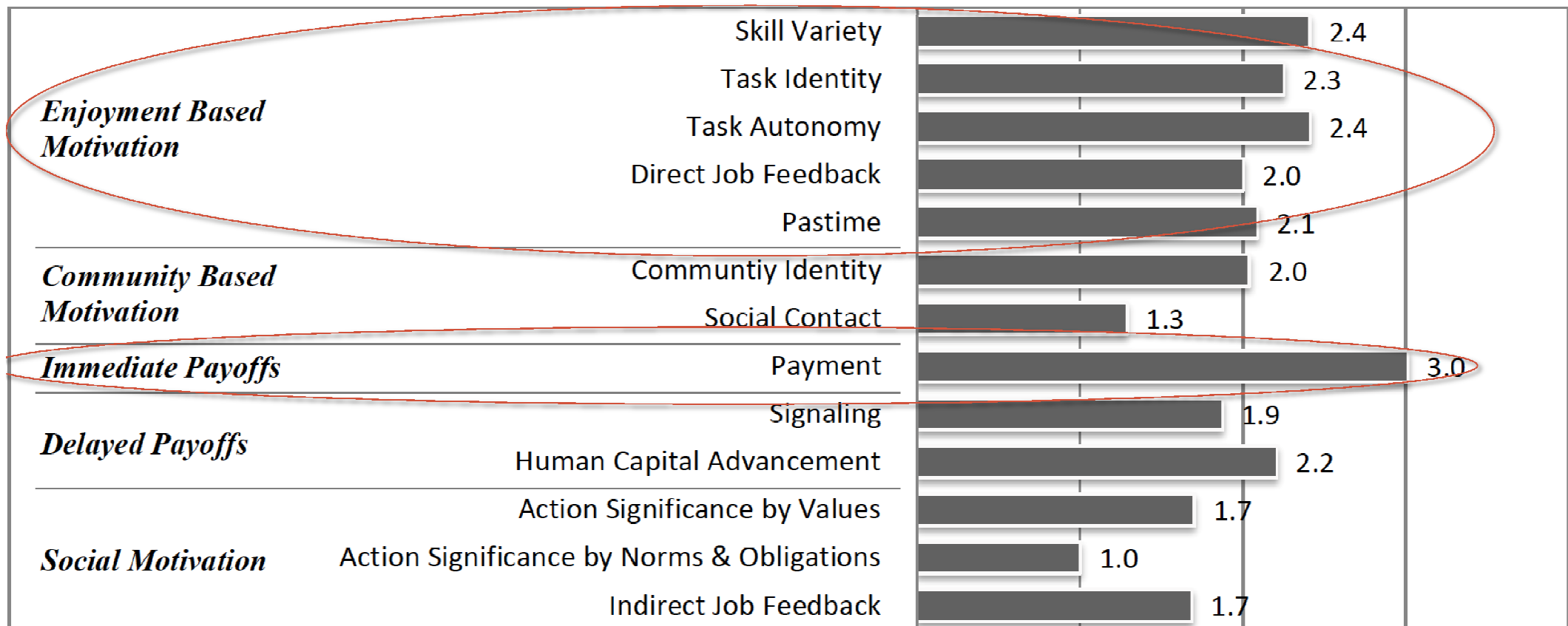
- Incentives
 - with gradually increasing pay (*Gao et al, PVLDB 2014*)
 - with a bonus (*Shaw et al. CSCW 2011, Slivkins et al. WWW 2015*)
 - with feedback on others' performance (*Shaw et al. CSCW 2011*)
 - with entertainment during task completion (*Dai et al, CSCW 2015*)
- Feedback
 - CrowdFlower displays a panel above the task with the worker's estimated accuracy so far
 - Encouraging text and a board showing how well a worker is doing. Diversions improve worker retention rate while retaining the same work quality (*Rzeszotarski et al, AAA 2013*)

Proposed Approach

Observe workers and adaptively assign tasks that maximize their motivation



Focus on two Factors



Adaptive Task Assignment (motivation)

with J. Pilourdault, S. B. Roy, D. Lee. *EDBT 2017*

balance between 2 factors, e.g.,
intrinsic factor, *task diversity*, and
extrinsic factor, *task reward*

$$\begin{aligned} \text{motiv}(\mathcal{T}, w) = & \alpha_w TD(\mathcal{T}) \\ & + \beta_w \times TR(\mathcal{T}, w) \end{aligned}$$

Adaptive Task Assignment (optimization)

For a worker, find a set of tasks:

$$\arg \max \sum_{w \in \mathcal{W}^i} \text{motiv}(\mathcal{T}_w^i, w)$$

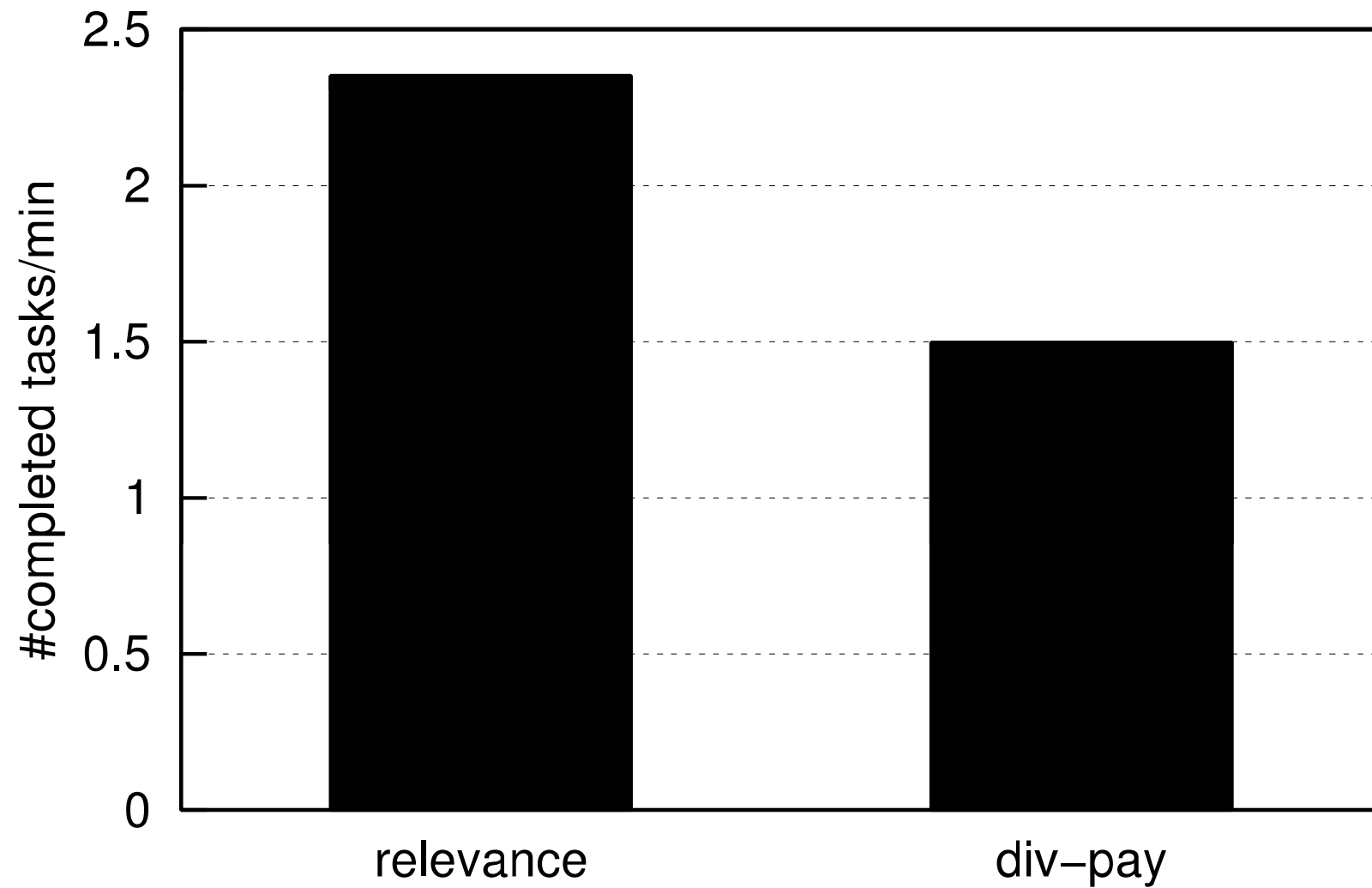
$$\forall w \in \mathcal{W}^i, |\mathcal{T}_w^i| \leq X_{max} \quad (C_1)$$

$$\forall w, w' \in \mathcal{W}^i, \mathcal{T}_w^i \cap \mathcal{T}_{w'}^i = \emptyset \quad (C_2)$$

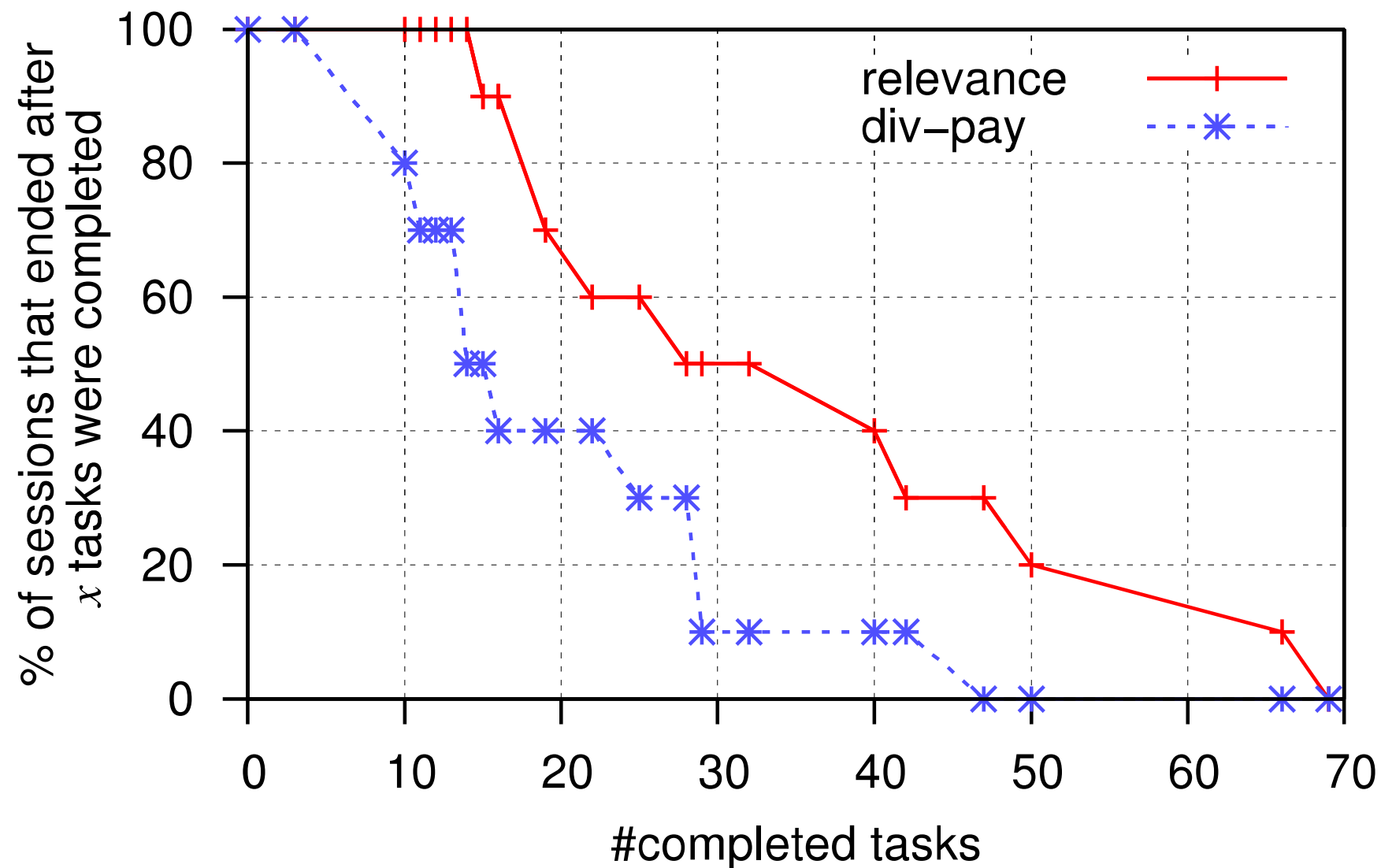
Adaptive Task Assignment (performance)

- 158,018 tasks from CrowdFlower in 22 kinds
- 23 workers in AMT
- 2 task assignment strategies:
 - Tasks matching a worker's profile: RELEVANCE
 - Tasks achieving a diversity/payment balance: DIV-PAY

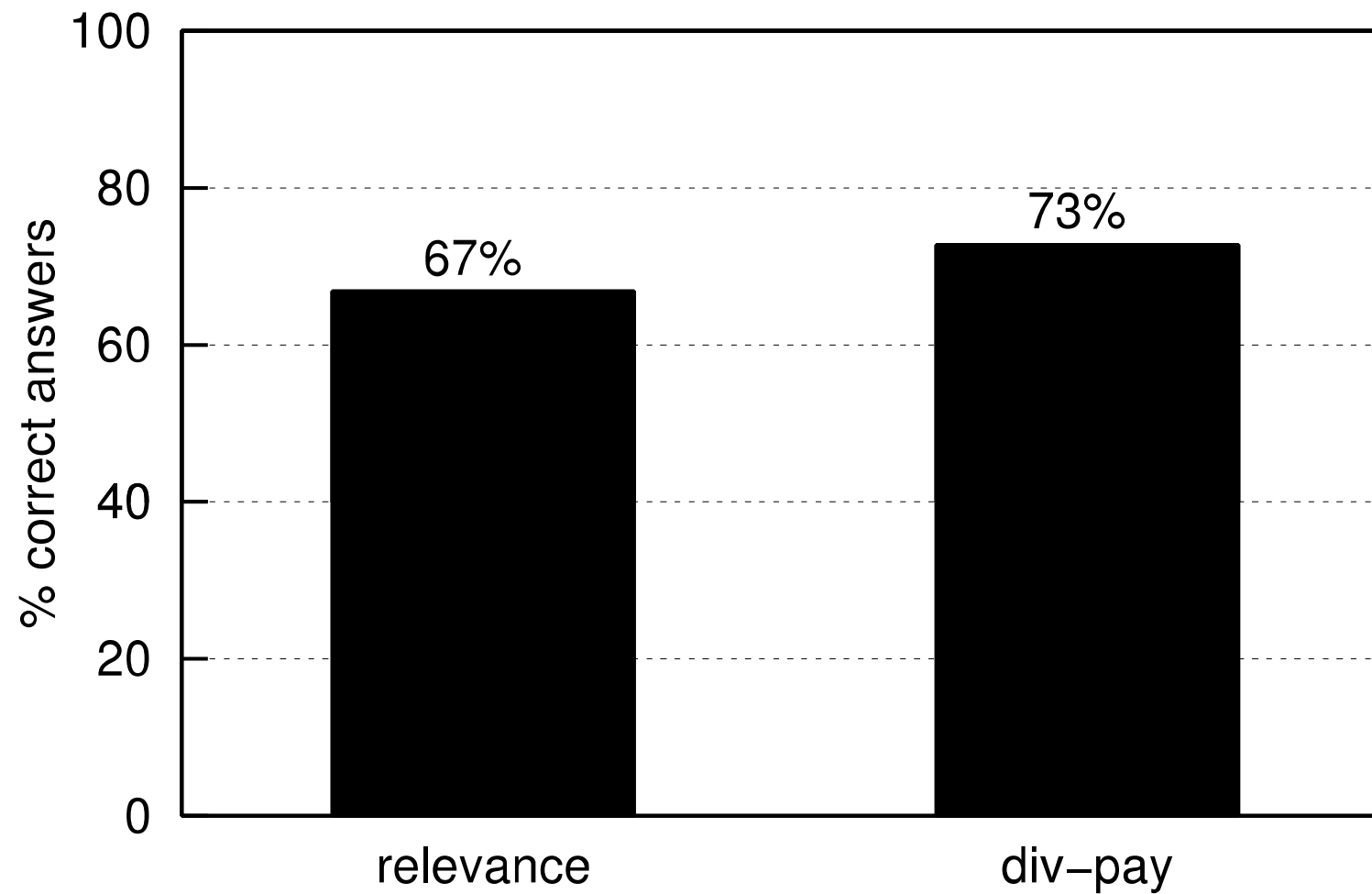
Task Throughput



Worker Retention



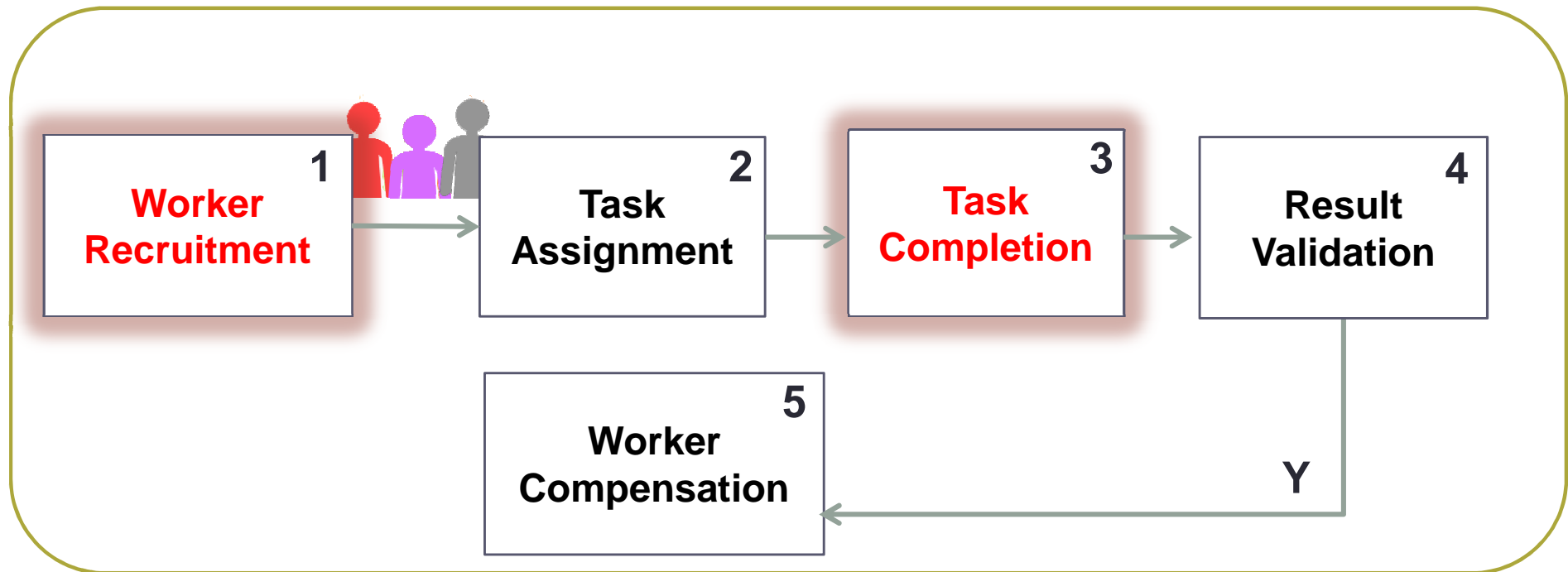
Outcome Quality



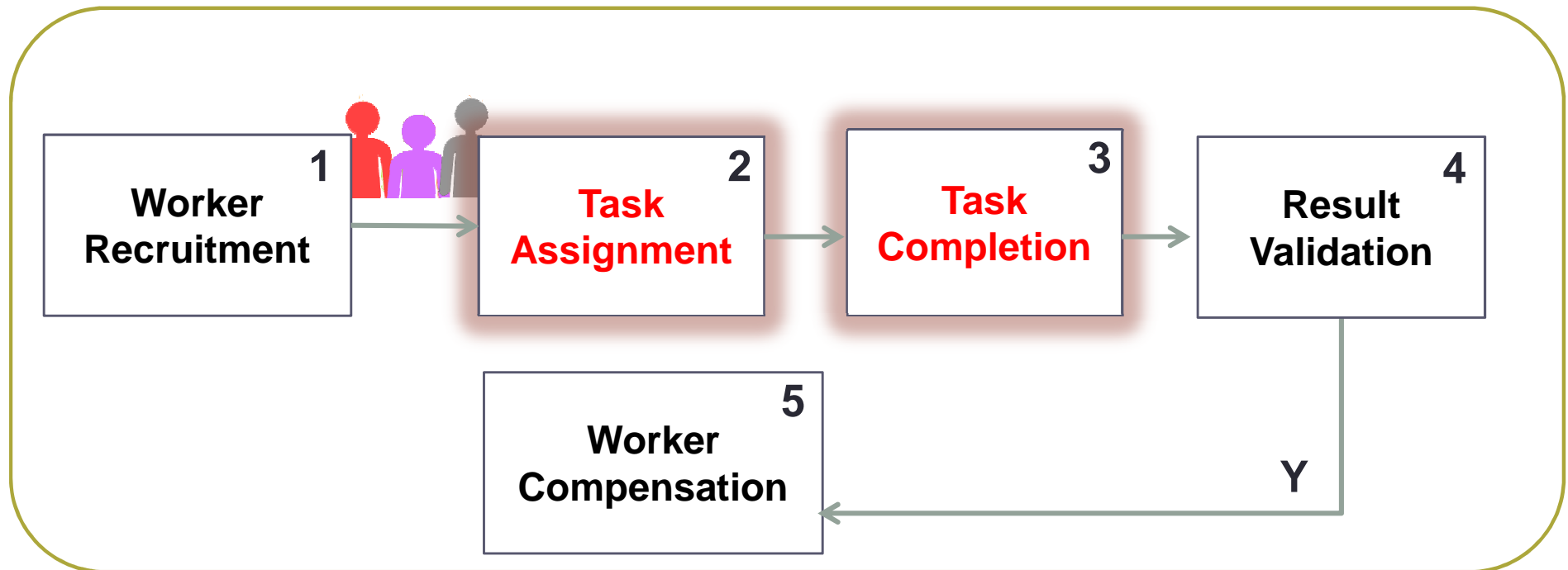
Summary and Takeaways

- Human factors are essential in crowdsourcing
- They need to be observed during task completion and leveraged in task assignment
- Their evolving nature requires to optimize crowdsourcing processes holistically

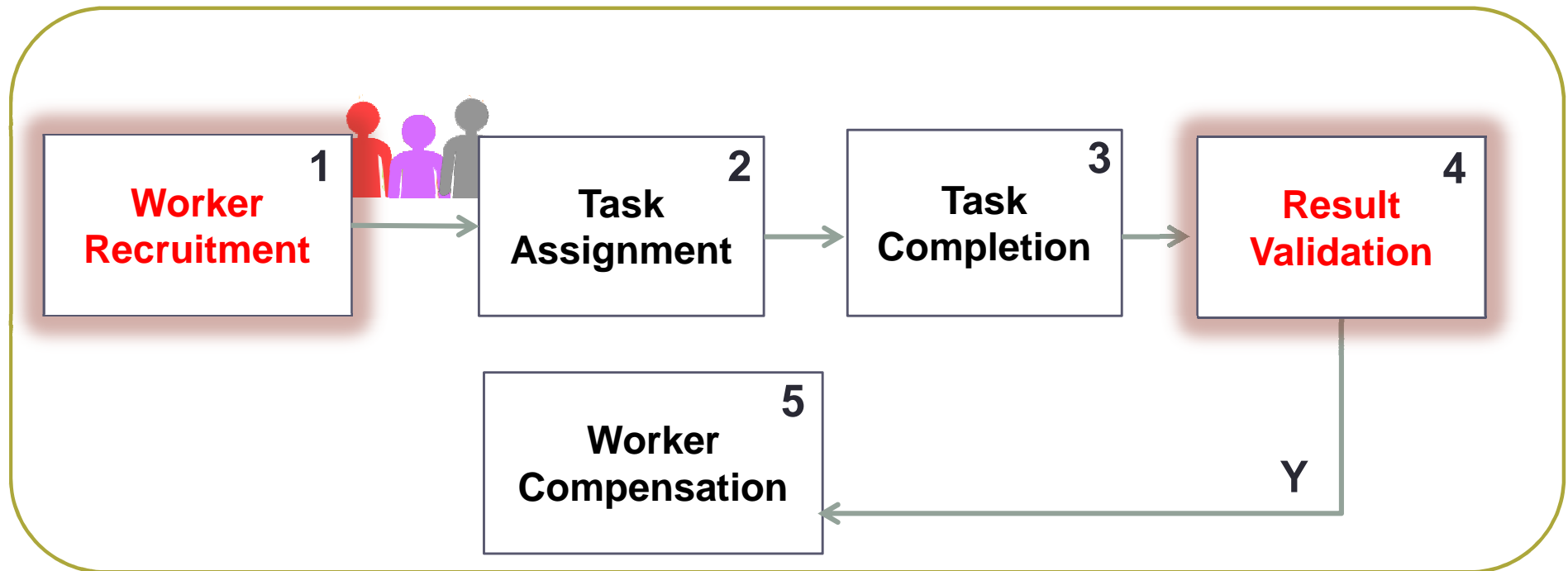
Human Factors in Task Assignment



Adaptive Task Assignment

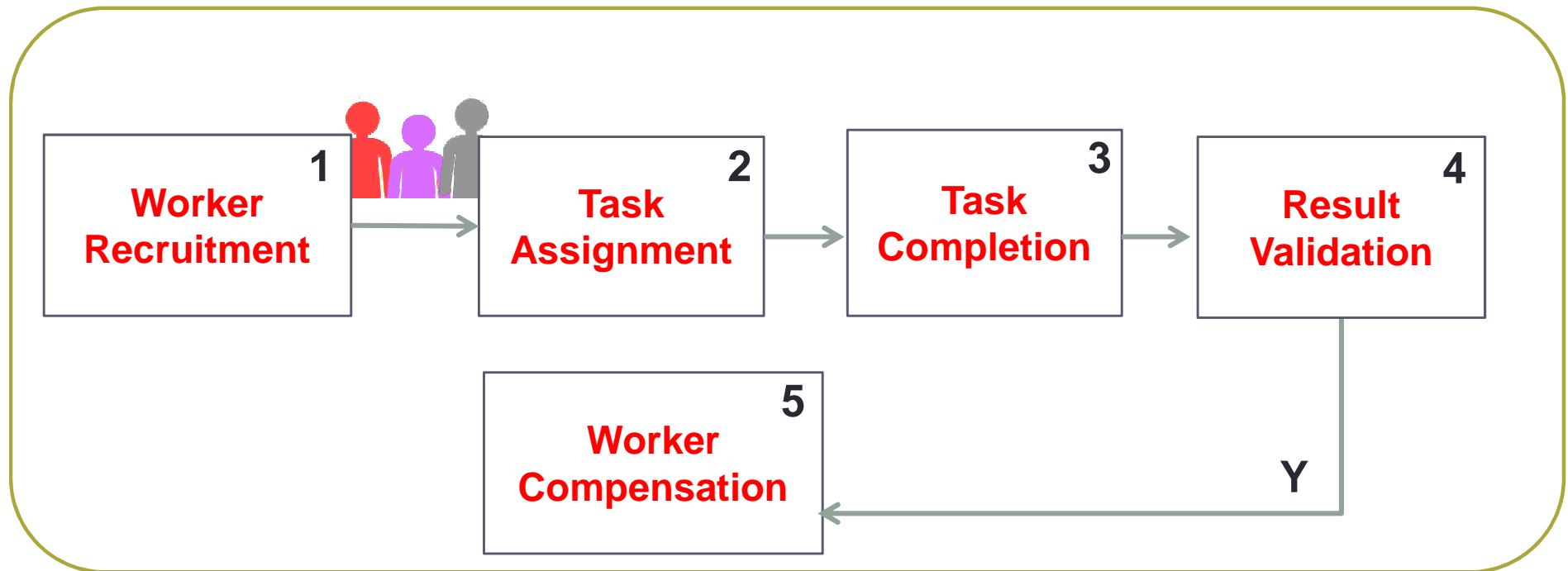


Learning Human Factors



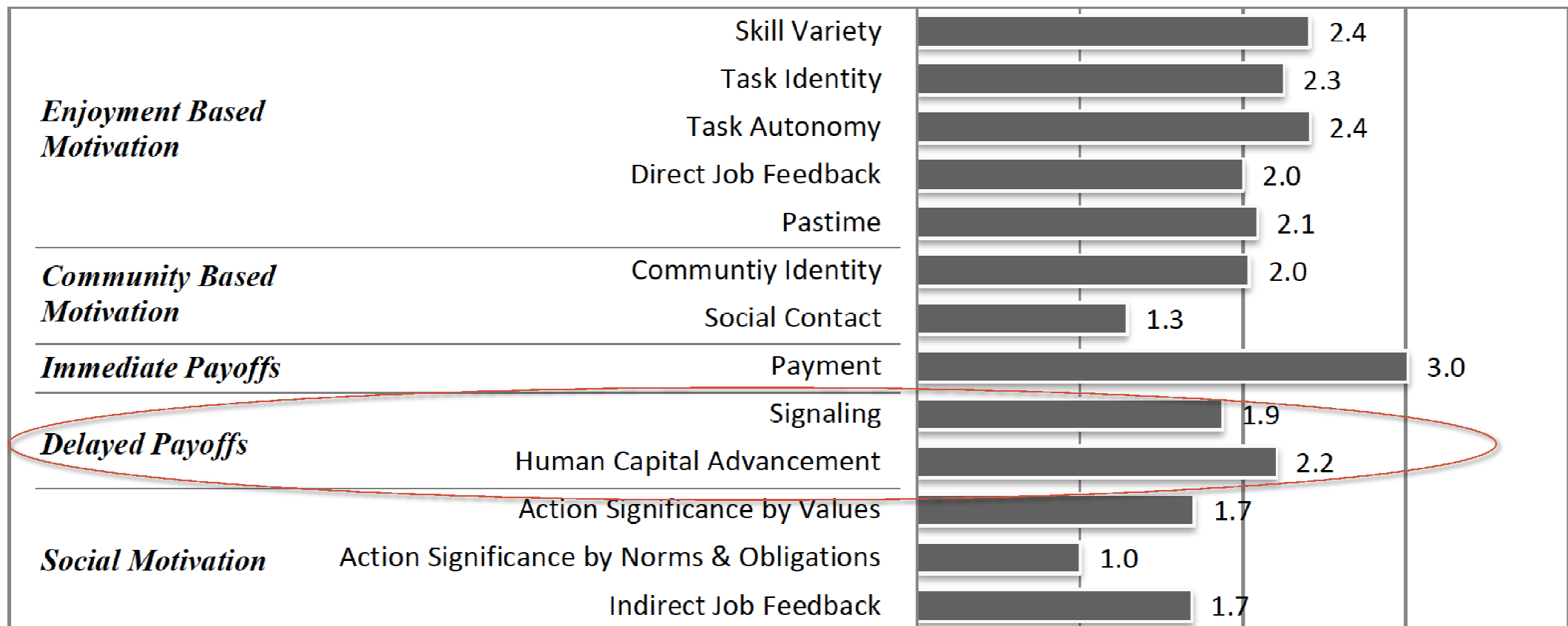
Looking Beyond

Toward Holistic Optimization



Why does a holistic view matter that much?

More than fun and money. worker motivation in crowdsourcing-a study on mechanical turk. N. Kaufmann, T. Schulze, and D. Veit. AMCIS 2011



Crowdsourcing platforms as a learning destination

- Ability for workers to express:
 1. I have x amount of time, which tasks should I complete?
 2. I want to make x dollars, which tasks should I complete?
 3. *I want to improve some skill, which tasks are best suited to me?*
- Requires to capture human factor end-to-end
 - to model team formation (who to team up with to learn faster?)
 - to model feedback in result validation and include it in task assignment