Applying Fair Reward Divisions to Collaborative Work

MMath Thesis Presentation

Greg d'Eon

University of Waterloo

June 26, 2019

Crowdsourcing and Mechanical Turk

Amazon Mechanical Turk: microtask crowdsourcing marketplace

- Requesters post Human Intelligence Tasks (HITs)
- Workers accept HITs, complete work, and submit for review
- Typically seconds or minutes of work for a few cents

Allows requesters to scale tasks to an enormous extent

- Worker population: 100,000 to 200,000 workers [Difallah et al., 2018]
- Instrumental in large datasets like ImageNet [Russakovsky et al., 2015]

Motivating Crowd Workers

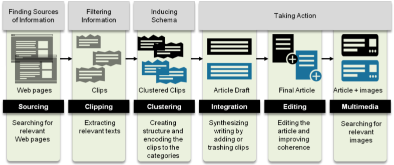
Important to keep workers motivated to ensure high quality work

Motivating workers on Mechanical Turk:

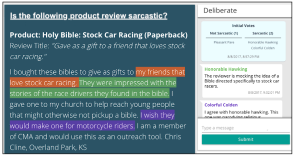
- Workers are primarily motivated by money [Kaufmann et al., 2011]
- Higher pay attracts more workers [Mason and Watts, 2009; Rogstadius et al., 2011]
- Performance-based pay can help in effort-responsive tasks [Ho et al., 2016]

Collaborative Crowdsourcing Tasks

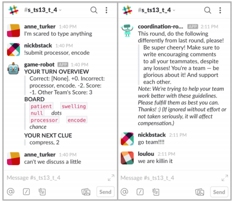
However, some crowdsourced tasks rely on collaboration between workers



[Hahn et al., 2016]



[Schaekermann et al., 2018]



[Zhou et al., 2018]

Worker motivation is not well understood in these collaborative tasks

Equity Theory

Collaboration changes the way that workers think about monetary rewards

Equity theory [Adams, 1965]: people think they are being treated fairly if

$$\frac{O_{self}}{I_{self}} = \frac{O_{other}}{I_{other}}$$

where:

- I: input (work quality, effort, time spent, ...)
- O: output (rewards or bonuses)

Related to motivation: underrewarded workers restore equity by putting in less work

This Talk

This Talk

Introduction

Collaborative Crowdsourcing Tasks

Fair Payments on Mechanical Turk

Human Perceptions of Fairness

Conclusion

Literature Review

Literature review of existing collaborative tasks

- Snowball sampling process
- Seeded literature review with Soylent [Bernstein et al., 2010]
- Found 114 papers describing collaborative crowdsourcing tasks

Informally coded types of collaboration based on task descriptions and interfaces

Features of Collaborative Tasks

Identified four distinguishing factors

Types of information that workers can have:

- **Aware of others**: Do they know that other workers are involved in the task?
- **See others' work**: Do they see other workers' output (*same* task or *other* task)?
- **Identify others' work**: Can they identify which worker did each part of the work?
- **Freely interact**: Can they have open, free-form conversations with other workers?

Found 13 combinations of these in existing tasks

Types of Collaborative Tasks

Anonymous shared interfaces

- ✓ Are aware of others
- ✓ Can see others' work
- ✗ Cannot identify others' work
- ✗ Cannot freely interact

Enables tasks that require some freedom to coordinate:

- Control arbitrary GUIs [Lasecki et al., 2011]
- Plan travel itineraries [Zhang et al., 2012]
- Write creative stories [Kim et al., 2017]

Types of Collaborative Tasks

Structured deliberation and shared interfaces

- ✓ Are aware of others
- ✓ Can see others' work
- ✓ Can identify others' work
- ✗ Cannot freely interact

Gives workers additional context about each others' work to:

- Create interface mockups [Lasecki et al., 2015]
- Power a chat bot [Huang et al., 2016]
- Reason about unclear instructions [Chang et al., 2017]

Types of Collaborative Tasks

Full collaboration

- ✓ Are aware of others
- ✓ Can see others' work
- ✓ Can identify others' work
- ✓ Can freely interact

Tightly coupled work through Google Documents, Etherpads, or Slack channels:

- Collectively brainstorm company slogans [Lykourentzou et al., 2017]
- Solve complex cognitive problems [Zhou et al., 2018]
- Deliberate about ambiguous questions [Schaekermann et al., 2018; Chen et al., 2019]

Existing Payment Systems

Existing tasks: most common to pay all workers equally

Paying for participation

- Example: pay bonuses for suggesting chat messages [Huang et al., 2016]
- Difficult to ensure these payments incentivize high effort

Paying for quality: requires measurement of work quality

- With ground truth, compare to correct answer
- Agreement with workers, influence on algorithm's output, or subjective judgements

Overall, payments are ad-hoc and not well motivated

This Talk

Fair Payments: Proportional

Two theoretically fair payments from equity theory and cooperative game theory

Equity theory: fair payments are proportional to inputs

$$O_i = c \cdot I_i$$

where c is amount of pay per unit of work

Subjective: input I_i could depend on work quality, work quantity, time spent

Fair Payments: Shapley Value

Cooperative game theory: *transferrable utility games* describe how a group can earn rewards by forming *coalitions*

Characteristic function: every coalition C could earn a reward $f(C)$ by working together

Players C	**Reward** $f(C)$
(nobody)	0
Alice	30
Bob	10
Charlie	0
Alice, Bob	60
Alice, Charlie	30
Bob, Charlie	10
Alice, Bob, Charlie	60

How to fairly divide the reward among them?

Fair Payments: Shapley Value

Shapley value [Shapley 1953]:

- Consider all possible orders of players joining the group
- Give players their *average marginal contribution* over these orders

Unique reward division satisfying 4 fairness axioms

Fair Payments: Shapley Value

Shapley value [Shapley 1953]:

- Consider all possible orders of players joining the group
- Give players their *average marginal contribution* over these orders

Unique reward division satisfying 4 fairness axioms

1. **Efficiency**: all of the grand coalition's reward is allocated

Fair Payments: Shapley Value

Shapley value [Shapley 1953]:

- Consider all possible orders of players joining the group
- Give players their *average marginal contribution* over these orders

Unique reward division satisfying 4 fairness axioms

1. **Efficiency**: all of the grand coalition's reward is allocated
2. **Symmetry**: players with *same marginal contributions* to all coalitions get same reward

Fair Payments: Shapley Value

Shapley value [Shapley 1953]:

- Consider all possible orders of players joining the group
- Give players their *average marginal contribution* over these orders

Unique reward division satisfying 4 fairness axioms

1. **Efficiency**: all of the grand coalition's reward is allocated
2. **Symmetry**: players with *same marginal contributions* to all coalitions get same reward
3. **Null Players**: players with *no marginal contribution* to any coalition get no reward

Fair Payments: Shapley Value

Shapley value [Shapley 1953]:

- Consider all possible orders of players joining the group
- Give players their *average marginal contribution* over these orders

Unique reward division satisfying 4 fairness axioms

1. **Efficiency**: all of the grand coalition's reward is allocated
2. **Symmetry**: players with *same marginal contributions* to all coalitions get same reward
3. **Null Players**: players with *no marginal contribution* to any coalition get no reward
4. **Additivity**: for all games f and g, $Sh(f+g) = Sh(f) + Sh(g)$

Research Questions

How do these theoretically fair payments affect crowd workers?

Specific questions:

1. Do workers think proportional pay and Shapley values are fairer than equal pay?
2. Are workers' fairness perceptions biased toward themselves?
3. Do workers put in more effort when they are paid fairly?

Study 1: HITs

Hired 132 workers

- 25 minute time estimate
- Offered base payment of $1.75 and typical bonus of $1

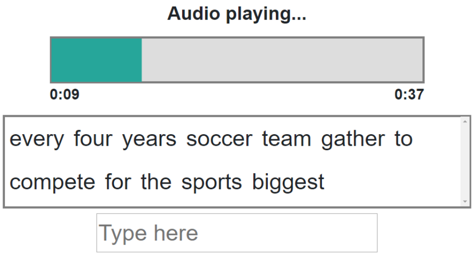
Placed workers into virtual groups

- Picked 2 prior workers as virtual teammates
- Placed group into one of four conditions

Study 1: Task

Experiment using audio transcription task based on Scribe [Lasecki 2012]

- Real time transcription: no pausing or rewinding
- 14 audio clips (21 – 31 s each)



Study 1: Teams and Payments

After each audio clip, paid group performance-based bonus

Worker 3 (you): words typed: 28/72 (38%), **correct:** 25/28 (89%)

every four years soccer teams from across the globe *team* gather to compete for the sports biggest trophy the world cup historically the americans have been brilliant winning three of the past seven world cups never finishing worse than third the american women that is the mens national team not so hot the us has *team* never finished higher than eighth except for 1930 the very first world cup when we finished third *eight*

Your team earned **$0.30** for typing **61** correct words (5c per 10 words).

Individual payments:

P1: 11c	P2: 12c	P3: 5c

Given you and your teamates' performance, how fair do you think your team's payments are?

− UNFAIR   ○ NEUTRAL   + FAIR

Study 1: Conditions

Split group's bonuses in one of four ways:

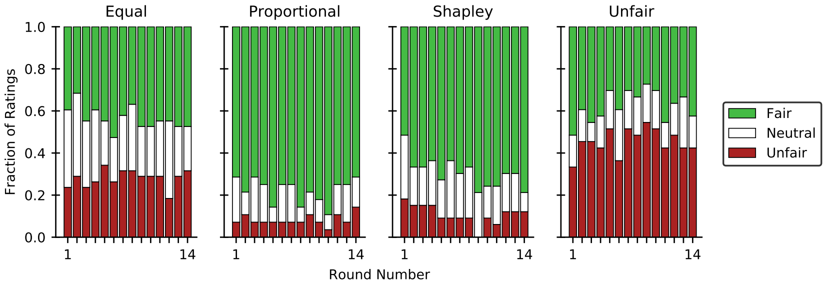
- Equal: pay each worker one third of the group's bonus
- Proportional: pay bonuses in proportion to number of correct words
- Shapley: compute rewards that each subset of workers would earn; pay Shapley valued bonuses based on these rewards
- Unfair: give 50% of bonus to worst worker and 25% to other two workers

Study 1: Fairness Ratings

Theoretically fair payments considered more fair than equal pay

Proportional odds model:

- PROPORTIONAL ($p < 0.001$) and SHAPLEY ($p < 0.01$) more fair than EQUAL
- UNFAIR not significantly different

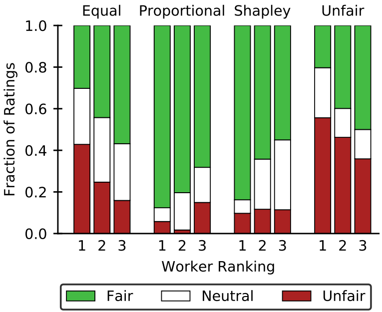


Study 1: Worker Biases

Best and worst workers in each group have different fairness perceptions

Including skill differentials in model, more skilled workers think:

- EQUAL ($p < 0.01$) and UNFAIR ($p < 0.001$) bonuses are less fair
- SHAPLEY ($p < 0.001$) bonuses are more fair

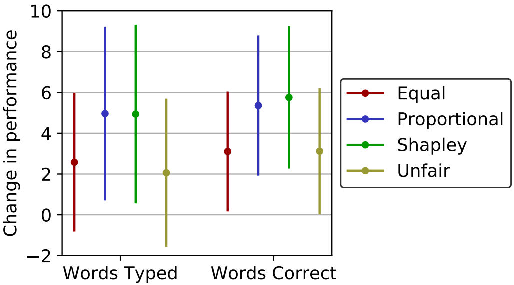


Study 1: Effort

To measure changes in effort, compared words typed in first and last rounds

Found no significant differences between conditions

- Noisy measurement of effort
- Other tasks better suited for analyzing effort



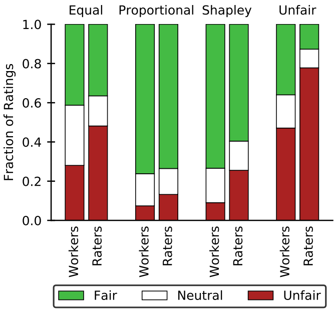
Study 2: External Raters

Follow-up: ask unbiased workers to rate bonuses (79 workers; \$1.50 for 12 minutes)

- Picked 4 rounds for each payment type: 1 fixed and 3 random

Raters were more critical of bonuses than original workers

- More negative: EQUAL ($p < 0.01$), SHAPLEY ($p < 0.001$), UNFAIR ($p < 0.001$)



This Talk

Fair Reward Divisions

Shapley value satisfies 4 axioms: efficiency, symmetry, null players, and additivity

Do these axioms represent fairness?

Weakening null player axiom results in more “human” alternatives:

- Solidarity value [Nowak and Radzik 1994]
- Egalitarian Shapley values [Joosten 1996, Casajus and Huettner 2013]
- Procedural values [Malawski 2013, Radzik and Driessen 2013]

Empirical Studies

Most experimental work: bargaining [Kalisch et al. 1954, Kahan and Rapoport 1984, Maschler 1992]

Impartial decisions about reward divisions [De Clippel et al. 2013]

- Rewards are convex combinations of equal split and Shapley value
- Rewards satisfy efficiency, symmetry, and additivity, but not null player
- Limitation: only studies *zero-normalized games*

Experiments

Question: How do single-player coalitions affect people's impartial reward divisions?

Answer this question through two experiments

- **Experiment 1:** Do people put more weight on 1- or 2-player coalitions' values?
- **Experiment 2:** How do people reason about 1-player coalitions?

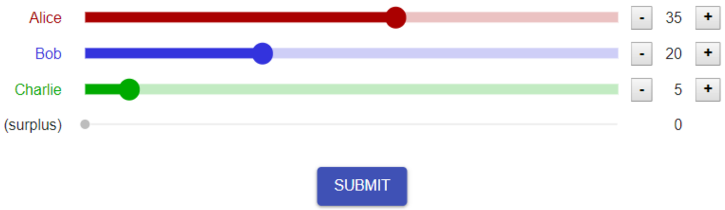
Experiment Interface

Experiment: divide rewards in fictional scenario

Players	**Gold Pieces**
(nobody)	0
Alice	30
Bob	20
Charlie	10
Alice, Bob	50
Alice, Charlie	40
Bob, Charlie	30
Alice, Bob, Charlie	60

All three of them go on the quest together and earn **60** gold pieces as a group.

How should they divide the gold?



Procedure

Within-subjects experiments

- Participants selected rewards for 11 or 17 games
- Hired 100 workers from Mechanical Turk for each experiment

Filtered out low-quality workers

- Spending under 5 seconds on any screen
- Submitting blatantly non-sensical answers

Experiment 1

Experiment 1: designed games to emphasize values of 1- or 2-player coalitions

	Game								Shapley value		
Condition	$\emptyset$	1	2	3	12	13	23	123	1	2	3

Experiment 1

Experiment 1: designed games to emphasize values of 1- or 2-player coalitions

- Choose target Shapley value

	Game								Shapley value		
Condition	$\emptyset$	1	2	3	12	13	23	123	1	2	3
									25	25	10

Experiment 1

Experiment 1: designed games to emphasize values of 1- or 2-player coalitions

- Choose target Shapley value
- Design game where only 1-player values differ

	Game								Shapley value		
Condition	$\emptyset$	1	2	3	12	13	23	123	1	2	3
Solo	0	40	40	10	60	60	60	60	25	25	10

Experiment 1

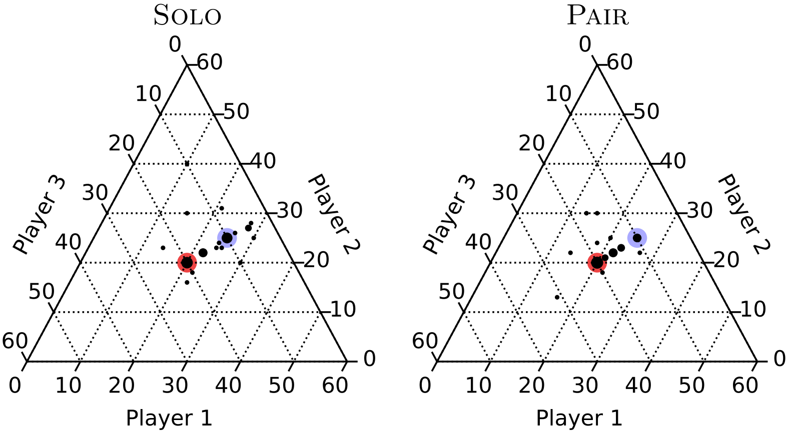
Experiment 1: designed games to emphasize values of 1- or 2-player coalitions

- Choose target Shapley value
- Design game where only 1-player values differ
- Design game where only 2-player values differ

	Game								Shapley value		
Condition	$\emptyset$	1	2	3	12	13	23	123	1	2	3
Solo	0	40	40	10	60	60	60	60	25	25	10
Pair	0	0	0	0	45	15	15	60			

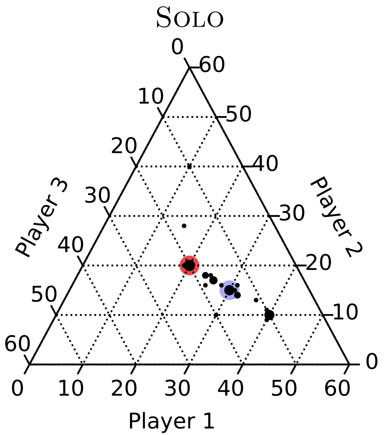
Experiment 1

Shapley value $= [25, 25, 10]$ (1-Worse)



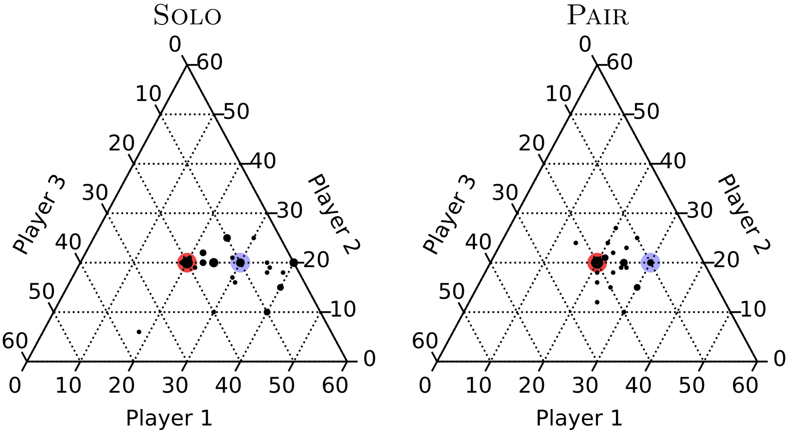
Experiment 1

Shapley value $= [30, 15, 15]$ (1-BETTER)



Experiment 1

Shapley value $= [30, 20, 10]$ (DISTINCT)



Experiment 2

Experiment 1: 1-player coalition values have larger effect on people's reward divisions

Goal of Experiment 2: understand how people reason about these values

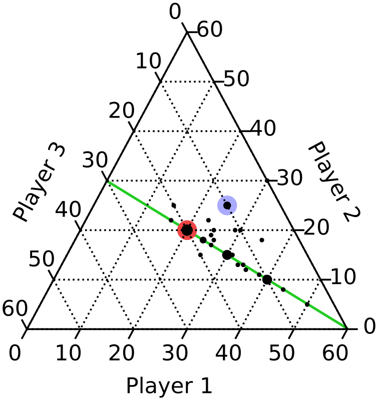
Focus on three features:

- 1-player values not a multiple of the Shapley value
- Varying sum of 1-player values
- Games with null players

Experiment 2

Shapley value $= [25, 25, 10]$, with 1-player values $[20, 5, 5]$:

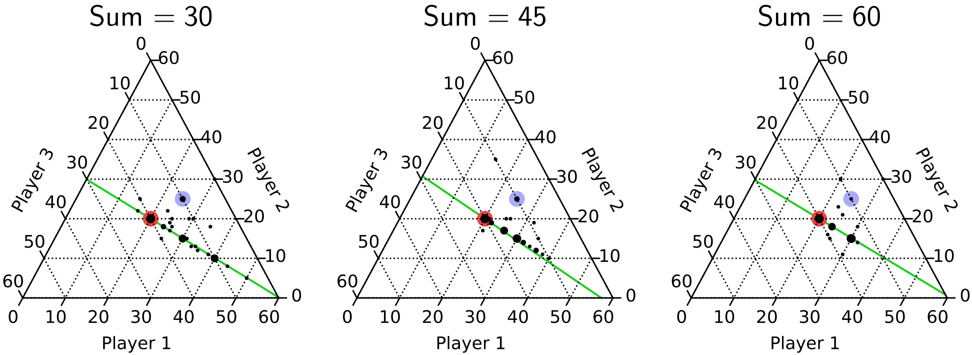
Game								Shapley value		
$\emptyset$	1	2	3	12	13	23	123	1	2	3
0	20	5	5	60	30	45	60	25	25	10



Experiment 2

Shapley value $= [25, 25, 10]$, with 1-player values summing to 30, 45, or 60:

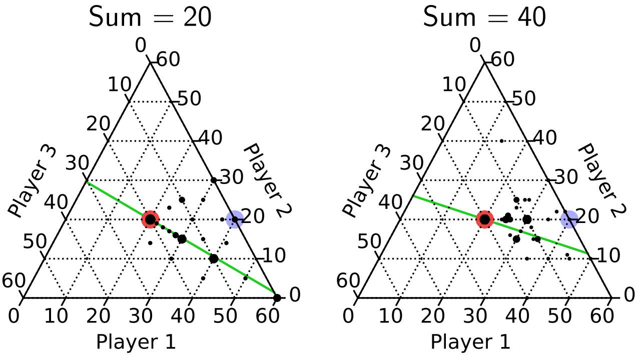
	Game								Shapley value		
Sum	$\emptyset$	1	2	3	12	13	23	123	1	2	3
30	0	20	5	5	60	30	45	60	25	25	10
45		25	10	10							
60		30	15	15							



Experiment 2

Shapley value $= [40, 20, 0]$, with player 3 null

	Game								Shapley value		
Sum	$\emptyset$	1	2	3	12	13	23	123	1	2	3
20	0	20	0	0	60	20	0	60	40	20	0
40	0	30	10	0	60	30	10	60			



Testing Axioms

Experiment 2: reward divisions are quite consistent, but unrelated to the Shapley value

Which axioms did people violate?

- Efficiency: required by experiment interface
- Symmetry ✓: always gave similar rewards to symmetric players in experiment 1
- Null player ✗: rarely gave 0 reward to null players in experiment 2
- Additivity ✗: gave inconsistent reward divisions in three games ($p < 0.01$):

	Game								Shapley value		
Sum	∅	1	2	3	12	13	23	123	1	2	3
30	0	20	5	5	60	30	45	60	25	25	10
45		25	10	10							
60		30	15	15							

This Talk

Summary

Performed literature review of existing collaborative tasks

- Identified different types of collaboration
- Found tasks that are only possible with close interaction

Tested effects of fair group payments on Mechanical Turk

- Designed two payment methods with theoretical motivation
- Workers are biased, but are perceptive of fair and unfair payments

Studied human reward divisions in cooperative games

- Reward divisions violate two of Shapley's fairness axioms

Broader Impacts

Ethical issues in crowdsourcing research

- Median wage on Mechanical Turk is under \$2/hour [Hara et al., 2018]
- Improving crowdsourced work can attract more low-paying requesters

Fair, transparent payments are beneficial to both workers and requesters

- Improve trust and reputation with workers
- Help workers understand how to do high quality work
- Collaborative tasks: get these benefits without relying on the platform

Future Work: Perceptions of Fairness

Models for people's fair reward divisions

- Had little success fitting procedural values
- Shapley value after applying non-linear utility function to coalition values
- Shapley value with weaker additivity axiom
- Ideas from bargaining: heuristics [Selten 1987] or stability concerns

Other factors affecting rewards

Future Work: Group Tasks

Worker motivation and fair pay in other group tasks

Tasks with no correct answer

- Use workers' subjective opinions about each other
- Theoretical mechanisms from "divide the dollar" game [De Clippel et al., 2008]
- Practical systems inspired by PageRank [Vaish et al., 2017]

Human-AI teams

- Collaborative tasks including chatbots [Huang et al., 2016; Zhou et al., 2018]
- Could impact worker motivation if AI takes easy jobs

Applying Fair Reward Divisions to Collaborative Work

MMath Thesis Presentation

Greg d'Eon

University of Waterloo

June 26, 2019