Testing Axioms Against Human Reward Divisions in Cooperative Games

AAMAS 2020

Greg d'Eon[1], Kate Larson[2]

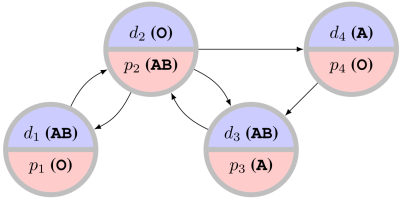
[1]University of British Columbia; [2]University of Waterloo

@gregdeon_
gregdeon.com
gregdeon@cs.ubc.ca

Algorithms with Human Values

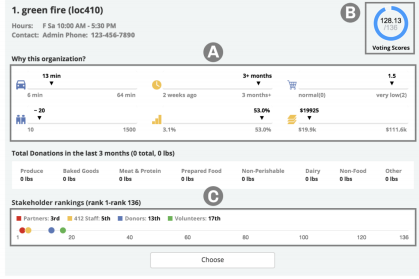
We are building algorithms that make **difficult moral decisions**.

Kidney exchanges



[Freedman et al., 2018]

Food rescue services



[Lee et al., 2019]

Algorithms with Human Values

Two approaches for designing values into algorithms

The **axiomatic** approach:

- Fix a set of axioms and derive the outcomes that satisfy them
- Conceptually simple with provable guarantees, but hard to capture social norms

The **empirical** approach:

- Elicit stakeholders' opinions and encode them into a model's behaviour
- Driven by data, but loses guarantees

Cooperative Game Theory

This tension appears in **cooperative game theory**:

- *Transferrable utility game*: set of players, and rewards for every coalition

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

Axioms: The Shapley Value

Major question: if everyone works together, how should they split the reward?

Most famous solution concept: **Shapley value** [Shapley 1953]

Unique reward division Sh satisfying 4 fairness axioms:

Axioms: The Shapley Value

Major question: if everyone works together, how should they split the reward?

Most famous solution concept: **Shapley value** [Shapley 1953]

Unique reward division Sh satisfying 4 fairness axioms:

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

Axioms: The Shapley Value

Major question: if everyone works together, how should they split the reward?

Most famous solution concept: **Shapley value** [Shapley 1953]

Unique reward division *Sh* satisfying 4 fairness axioms:

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

Axioms: The Shapley Value

Major question: if everyone works together, how should they split the reward?

Most famous solution concept: **Shapley value** [Shapley 1953]

Unique reward division *Sh* satisfying 4 fairness axioms:

1. **Efficiency**: all of the group'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

Axioms: The Shapley Value

Major question: if everyone works together, how should they split the reward?

Most famous solution concept: **Shapley value** [Shapley 1953]

Unique reward division Sh satisfying 4 fairness axioms:

1. **Efficiency**: all of the group'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 games f and g with the same players, $Sh(f+g) = Sh(f) + Sh(g)$

Alternatives and Empirical Studies

Do these axioms capture **fairness**?

Alternatives weaken the null player axiom:

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

Experiments on impartial decisions about reward divisions [De Clippel and Rozen, 2013]

- Rewards are convex combinations of equal split and Shapley value
- Satisfy efficiency, symmetry, and additivity, but not null player

Overview

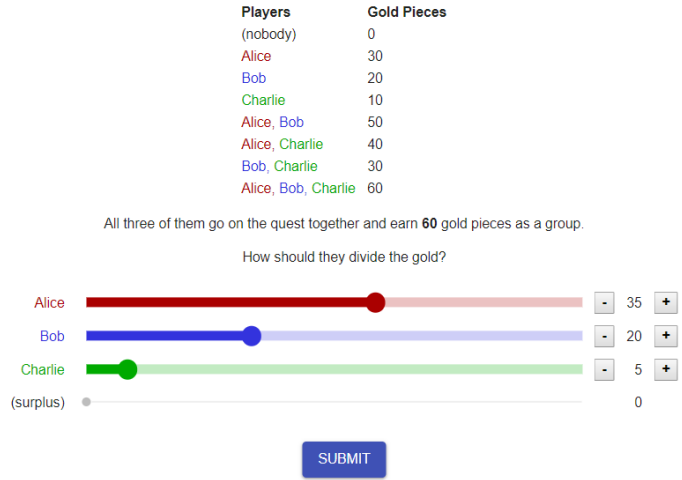
In this talk:

- Use crowdsourced experiments to study impartial reward divisions
- Find that people often pick rewards unrelated to the Shapley value
- Show that people violate additivity and null player axioms, but identify weaker axioms that align with their decisions

Experiments

Two crowdsourced MTurk experiments: divide rewards in fictional scenarios

- After filtering low-effort responses, $n = 74$ and 75



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?

Alice - 35 +

Bob - 20 +

Charlie - 5 +

(surplus) 0

SUBMIT

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	Solo	Pair
(nobody)		
Alice		
Bob		
Charlie		
Alice, Bob		
Alice, Charlie		
Bob, Charlie		
Alice, Bob, Charlie		

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	SOLO	PAIR
(nobody)	0	
Alice	40	
Bob	40	
Charlie	10	
Alice, Bob	60	
Alice, Charlie	60	
Bob, Charlie	60	
Alice, Bob, Charlie	60	

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	SOLO	PAIR
(nobody)	0	
Alice	**40**	
Bob	**40**	
Charlie	**10**	
Alice, Bob	60	
Alice, Charlie	60	
Bob, Charlie	60	
Alice, Bob, Charlie	60	

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	SOLO	PAIR
(nobody)	0	
Alice	40	
Bob	40	
Charlie	10	
Alice, Bob	**60**	
Alice, Charlie	**60**	
Bob, Charlie	**60**	
Alice, Bob, Charlie	60	

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	SOLO	PAIR
(nobody)	0	0
Alice	40	0
Bob	40	0
Charlie	10	0
Alice, Bob	60	45
Alice, Charlie	60	15
Bob, Charlie	60	15
Alice, Bob, Charlie	60	60

Experiment 1

Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	Solo	Pair
(nobody)	0	0
Alice	40	**0**
Bob	40	**0**
Charlie	10	**0**
Alice, Bob	60	45
Alice, Charlie	60	15
Bob, Charlie	60	15
Alice, Bob, Charlie	60	60

Experiment 1

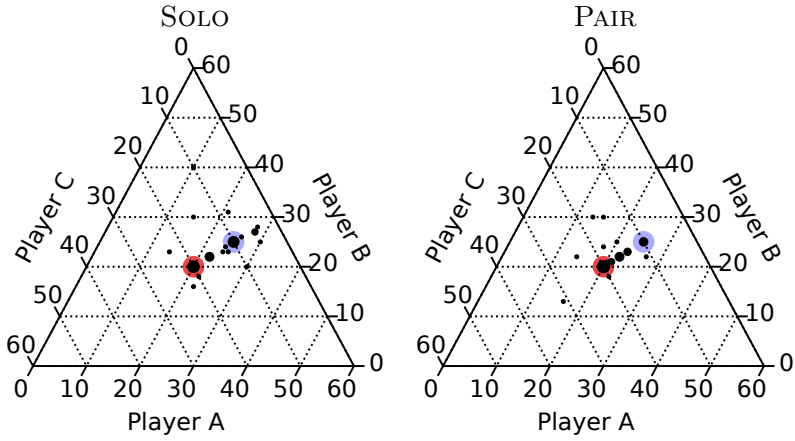
Experiment 1: design games to emphasize 1- or 2-player groups

- Example: Shapley value of $[25, 25, 10]$

Players	Solo	Pair
(nobody)	0	0
Alice	40	0
Bob	40	0
Charlie	10	0
Alice, Bob	60	**45**
Alice, Charlie	60	**15**
Bob, Charlie	60	**15**
Alice, Bob, Charlie	60	60

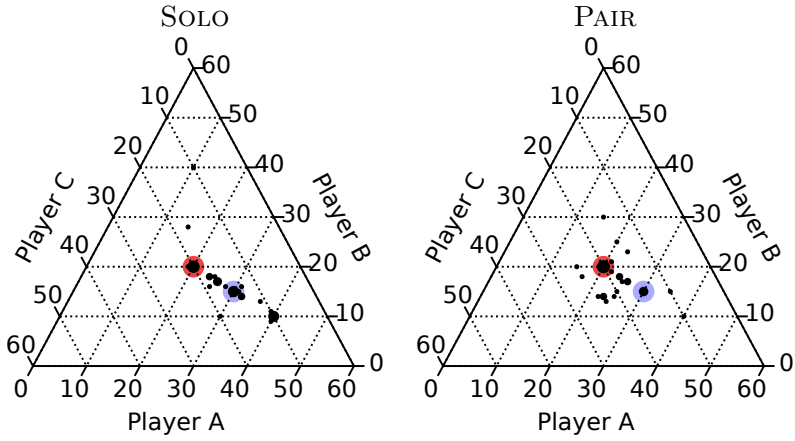
Experiment 1: Results

Shapley value $= [25, 25, 10]$:



Experiment 1: Results

Shapley value $= [30, 15, 15]$:



Experiment 2

Experiment 2: investigate impacts of 1-player groups

- Revisit example: Shapley value of $[25, 25, 10]$

Players	**Rewards**
(nobody)	0
Alice	25
Bob	10
Charlie	10
Alice, Bob	60
Alice, Charlie	30
Bob, Charlie	45
Alice, Bob, Charlie	60

Experiment 2

Experiment 2: investigate impacts of 1-player groups

- Revisit example: Shapley value of $[25, 25, 10]$

Players	**Rewards**
(nobody)	0
Alice	**25**
Bob	**10**
Charlie	10
Alice, Bob	60
Alice, Charlie	30
Bob, Charlie	45
Alice, Bob, Charlie	60

Experiment 2

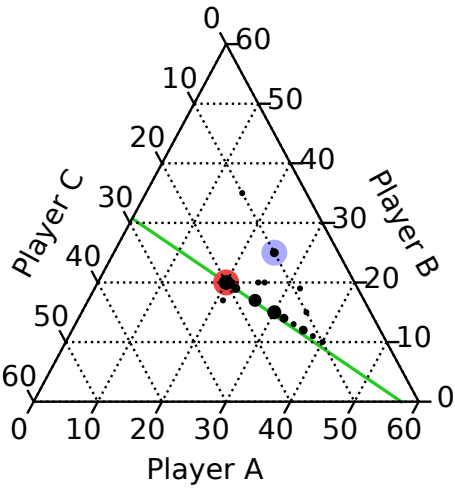
Experiment 2: investigate impacts of 1-player groups

- Revisit example: Shapley value of $[25, 25, 10]$

Players	**Rewards**
(nobody)	0
Alice	25
Bob	10
Charlie	10
Alice, Bob	60
Alice, Charlie	**30**
Bob, Charlie	**45**
Alice, Bob, Charlie	60

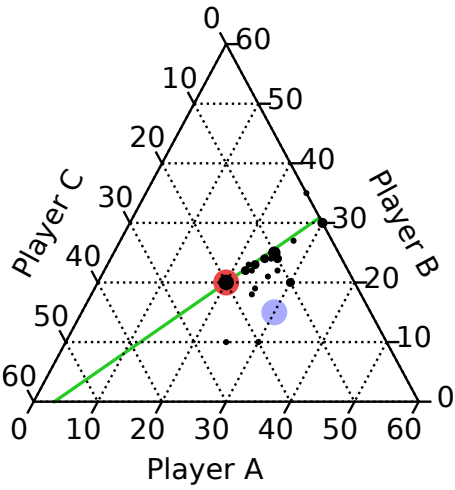
Experiment 2: Results

Shapley value $= [25, 25, 10]$:



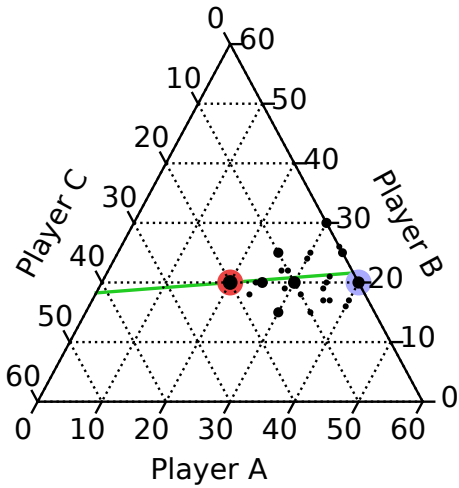
Experiment 2: Results

Shapley value $= [30, 15, 15]$:



Experiment 2: Results

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



Testing Axioms: Symmetry

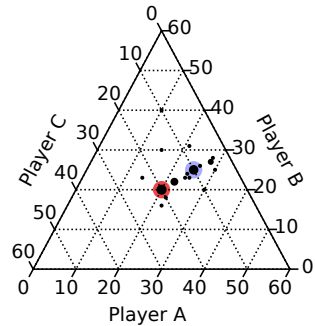
Which axioms did people violate?

- Efficiency: required by experiment interface

Symmetry: must give equal rewards to symmetric players

- Experiment 1 games had symmetric players
- 455/525 (86.7%) reward divisions obeyed symmetry
- No significant differences

Symmetry: ✓



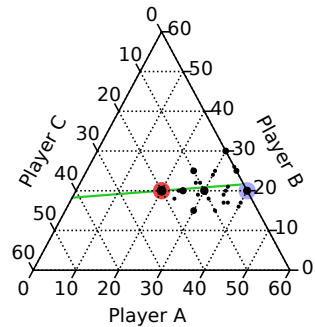
Testing Axioms: Null Player

Null player axiom: must give no reward to null players

- 4 games in Experiment 2 with null players
- Best case: 14/74 participants gave 0 reward

Null player: ✗

- Consistent with De Clippel and Rozen [2013]



Testing Axioms: Additivity

Additivity: test relationships between games

- Assuming efficiency and symmetry, must give **same rewards** in some games

Players	f	g	$f-g$
(nobody)			
Alice			
Bob			
Charlie			
Alice, Bob			
Alice, Charlie			
Bob, Charlie			
Alice, Bob, Charlie			

Testing Axioms: Additivity

Additivity: test relationships between games

- Assuming efficiency and symmetry, must give **same rewards** in some games

Players	f	g	$f - g$
(nobody)	0		
Alice	25		
Bob	10		
Charlie	10		
Alice, Bob	60		
Alice, Charlie	30		
Bob, Charlie	45		
Alice, Bob, Charlie	60		

Testing Axioms: Additivity

Additivity: test relationships between games

- Assuming efficiency and symmetry, must give **same rewards** in some games

Players	f	g	$f - g$
(nobody)	0	0	
Alice	25	20	
Bob	10	5	
Charlie	10	5	
Alice, Bob	60	60	
Alice, Charlie	30	30	
Bob, Charlie	45	45	
Alice, Bob, Charlie	60	60	

Testing Axioms: Additivity

Additivity: test relationships between games

- Assuming efficiency and symmetry, must give **same rewards** in some games

Players	f	g	$f-g$
(nobody)	0	0	
Alice	**25**	**20**	
Bob	**10**	**5**	
Charlie	**10**	**5**	
Alice, Bob	60	60	
Alice, Charlie	30	30	
Bob, Charlie	45	45	
Alice, Bob, Charlie	60	60	

Testing Axioms: Additivity

Additivity: test relationships between games

- Assuming efficiency and symmetry, must give **same rewards** in some games

Players	f	g	$f-g$
(nobody)	0	0	0
Alice	**25**	**20**	5
Bob	**10**	**5**	5
Charlie	**10**	**5**	5
Alice, Bob	60	60	0
Alice, Charlie	30	30	0
Bob, Charlie	45	45	0
Alice, Bob, Charlie	60	60	0

Testing Axioms: Additivity

Found that people gave inconsistent rewards to players 1 and 3

- Significant for games with $Sh = [25, 25, 10]$ ($p < 0.01$)
- Marginally significant for games with $Sh = [30, 15, 15]$ ($p = 0.07$ and $p = 0.08$)

Additivity: ✗

- Conflicts with De Clippel and Rozen [2013]

Alternative Axioms

Efficiency and symmetry put little structure on space of outcomes

Weaker axioms that align with reward divisions?

- **Local monotonicity:** if player i never has a smaller marginal contribution than player j, then player i should not have a smaller reward

Closer match to people's opinions:

- Experiment 1: 734/825 (89%)
- Experiment 2: 1203/1258 (95%)

Summary & Beyond Cooperative Games

Recap:

- People pick reward divisions that are often unrelated to Shapley value
- Reward divisions break null player and additivity axioms, but satisfy weaker axioms

These methods apply beyond cooperative game theory

- Fair division and voting rules: rich bodies of axiomatic literature
- Provides tools to direct and analyze experiments

Unsatisfying if axioms don't pin down a single outcome?

- Inevitable, since people don't agree on one definition of fairness
- "Algorithm-in-the-loop" systems: outcomes are starting point for human decisions