
POTSDAM INSTITUTE FOR CLIMATE IMPACT RESEARCH

Berlin Mathematics Research Center
MATH+

# Responsibility Functions for Explaining Deviations in Decision Behaviour

## - CHANGES+ Colloquium -

**Sarah Hiller | Anna-Katharina Kothe**

**April 2020**

---

Leibniz-Gemeinschaft

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

GaNe
FutureLab on Game Theory
& Networks of Interacting Agents

# Outline



# Introduction

- Motivation:
  - Responsibility decision-making nexus
  - Assign responsibility: Assign call for actions
- Approach:
  - Formalized Responsibility Function
  - Game and according experiment
- Responsibility Functions based on Heiztig & Hiller (submitted)
- Decision dilemma in game and according experiment based on Kline et al. (2018)

# Framework

Ingredients:

- Agents $I$
- Directed tree $\langle V, E \rangle$
- Possible actions $A_v$, consequences $c_v : A_v \to S_v$
-
-
-
-

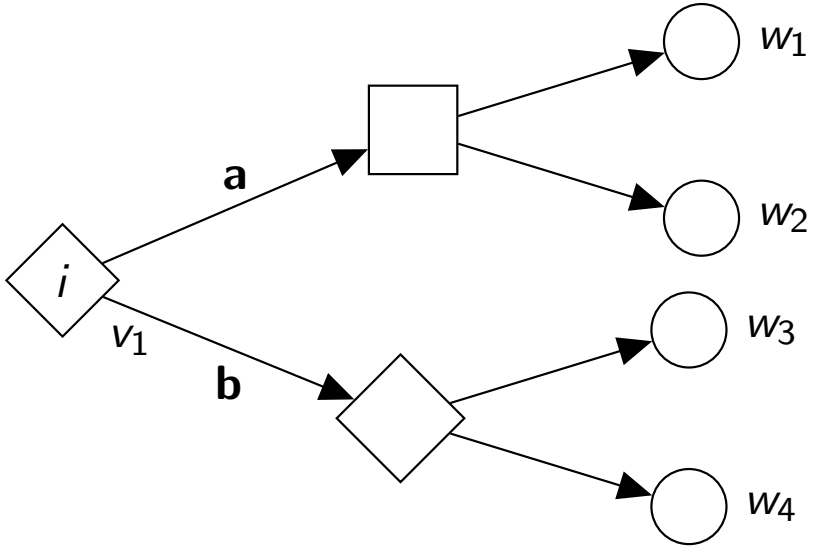



Figure: Graphical depiction of a morally evaluated multi-agent decision tree with uncertainty.

# Framework

Ingredients:

- Agents $I$
- Directed tree $\langle V, E \rangle$
- Possible actions $A_v$, consequences $c_v : A_v \rightarrow S_v$
- Set of ethically undesired outcomes $\epsilon$
-
-
-

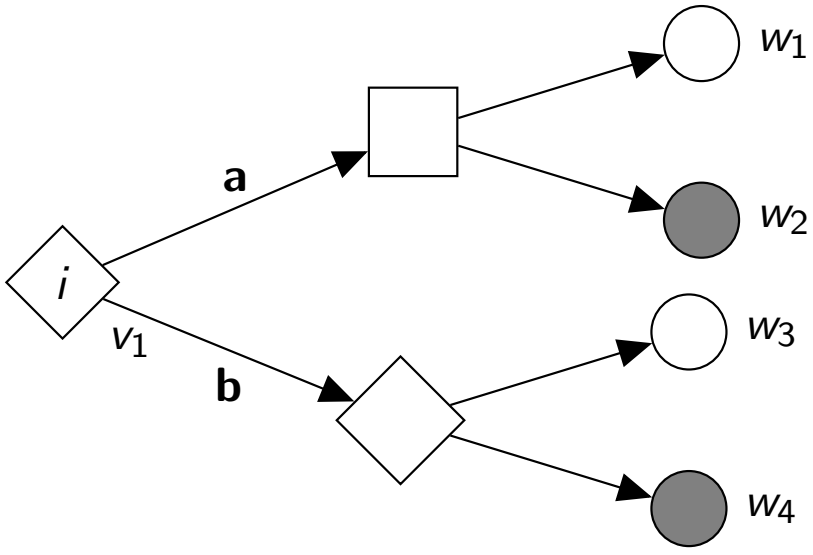



Figure: Graphical depiction of a morally evaluated multi-agent decision tree with uncertainty.

# Framework

Ingredients:

- Agents $I$
- Directed tree $\langle V, E \rangle$
- Possible actions $A_v$, consequences $c_v : A_v \rightarrow S_v$
- Set of ethically undesired outcomes $\epsilon$
- Ambiguity nodes $V_a$
- Probabilistic uncertainty $V_p$
-

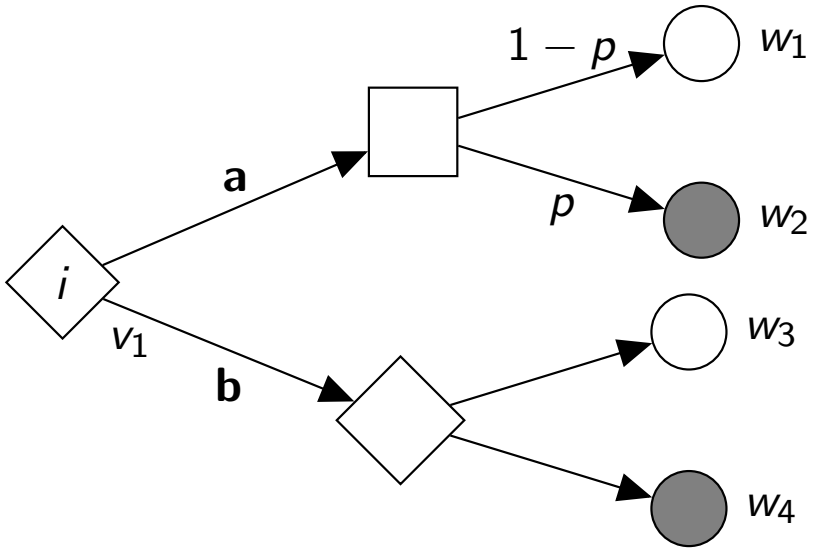



Figure: Graphical depiction of a morally evaluated multi-agent decision tree with uncertainty.

# Framework

Ingredients:

- Agents $I$
- Directed tree $\langle V, E \rangle$
- Possible actions $A_v$, consequences $c_v : A_v \to S_v$
- Set of ethically undesired outcomes $\epsilon$
- Ambiguity nodes $V_a$
- Probabilistic uncertainty $V_p$
- Information sets $\sim$

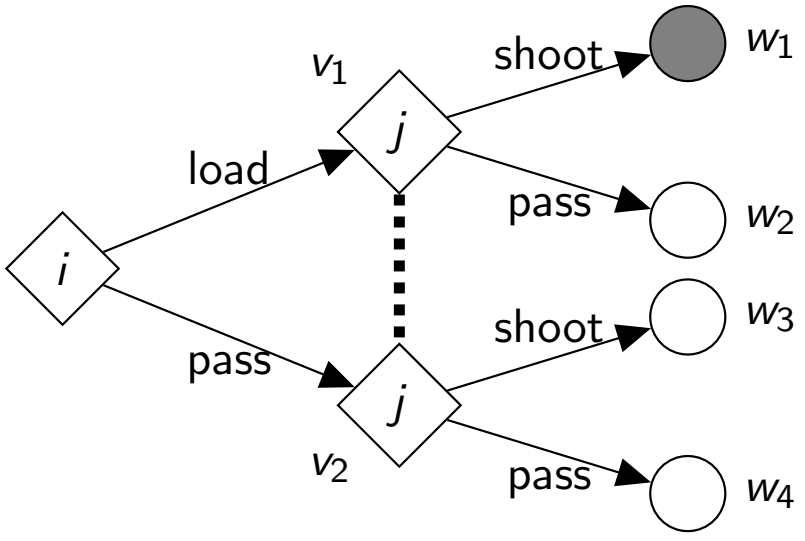



Figure: Graphical depiction of a morally evaluated multi-agent decision tree with uncertainty.

# Responsibility Function

## Scenario, strategy

A *scenario* $\zeta \in Z^{\sim}$ resolves all ambiguity and information uncertainty A *strategy* $\sigma \in \Sigma$ of a group $G \subseteq I$ selects actions for all future decision nodes.

# Responsibility Function

## Scenario, strategy

A *scenario* $\zeta \in Z^{\sim}$ resolves all ambiguity and information uncertainty A *strategy* $\sigma \in \Sigma$ of a group $G \subseteq I$ selects actions for all future decision nodes.

## Hypothetical shortfall

Given a scenario $\zeta$, the *shortfall* of playing $\mathbf{a}$ in node $v$ is

$$\Delta\omega(v, \mathbf{a}) := \min_{\sigma} \ell(\epsilon \mid c_v(\mathbf{a}), \sigma, \zeta) - \min_{\sigma} \ell(\epsilon \mid v, \sigma, \zeta)$$

# Responsibility Function

## Scenario, strategy

A *scenario* $\zeta \in Z^{\sim}$ resolves all ambiguity and information uncertainty A *strategy* $\sigma \in \Sigma$ of a group $G \subseteq I$ selects actions for all future decision nodes.

## Hypothetical shortfall

Given a scenario $\zeta$, the *shortfall* of playing $\mathbf{a}$ in node $v$ is

$$\Delta\omega(v, \mathbf{a}) := \min_{\sigma} \ell(\epsilon \mid c_v(\mathbf{a}), \sigma, \zeta) - \min_{\sigma} \ell(\epsilon \mid v, \sigma, \zeta)$$

## Responsibility

$$\mathcal{R}(v, \mathbf{a}) := \max_{\zeta \in Z^{\sim}(v)} \Delta\omega(v, \zeta, \mathbf{a})$$

# Criteria

- Differentiated control groups
- Uncertainty
- Ethically (un)desired outcomes
- Non-linearity

# Game specification

nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-018-0418-0

## Differentiated responsibilities and prosocial behaviour in climate change mitigation

Reuben Kline[1,2*], Nicholas Seltzer[3], Evgeniya Lukinova[4,5] and Autumn Bynum[6]

**A characteristic feature of the global climate change dilemma is interdependence between the underlying economic development that drives anthropogenic climate change—typically modelled as a common pool resource dilemma[1,2]—and the subsequent dilemma arising from the need to mitigate the negative effects of climate change, often modelled as a public goods dilemma[3,4]. In other words, in a carbon-based economy, causal responsibility for climate change is a byproduct of economic development, and is therefore endogenous to it. To capture this endogeneity, we combine these two dilemmas into a 'compound climate dilemma' and conduct a series of incentivized experiments in the United States and China to test its implications for cooperation and prosocial behaviour. Here we show that, in a differentiated development condition, even while the advantaged parties increase their prosociality relative to an endogenous but homogeneous baseline condition, the accompanying decrease in cooperative behaviour by the disadvantaged par-**

and many others, the details are open to interpretation. The decision of who to hold accountable and how to differentiate their obligations is, has been and will likely continue to be a source of conflict. Precisely how much more should the more materially advantaged pay and based on what metric[10,13]? Should current generations be liable for the emissions produced by their ancestors[14]? How do we appropriately factor in production versus consumption, and the related issues of international trade and the outsourcing of emissions[15]? Once acknowledged, such endogeneity creates an additional dimension of conflict: in light of this endogeneity on what basis should obligations to mitigate climate change be differentiated?

This endogeneity suggests that to the extent that economic development is a function of greenhouse gas emissions, wealthier parties bear more causal responsibility for the severity of the climate change problem. Because of their wealth, they have a greater capacity to shoulder the burden of preventing or mitigating climate change. In this study, we argue that this endogeneity is so fundamental to the

# Game specification

| **Phase 1**: 10 rounds *appropriation* | **Phase 2**: 10 rounds *mitigation* |
|---|---|
| Appropriate $0, \ldots, 4$ of the common resource.<br>Differentiated case: half of the agents only start in round 6. | Mitigation goal: 0.53 of total appropriation (phase 1).<br>Contribute $0, \ldots, 4$ to mitigation effort.<br>If the mitigation effort is not met, everyone loses everything with a certain probability $p$, which increases step-wise from $\frac{2}{12}$ to $\frac{6}{12}$ to $\frac{9}{12}$ to $\frac{11}{12}$ with rising total appropriation. |

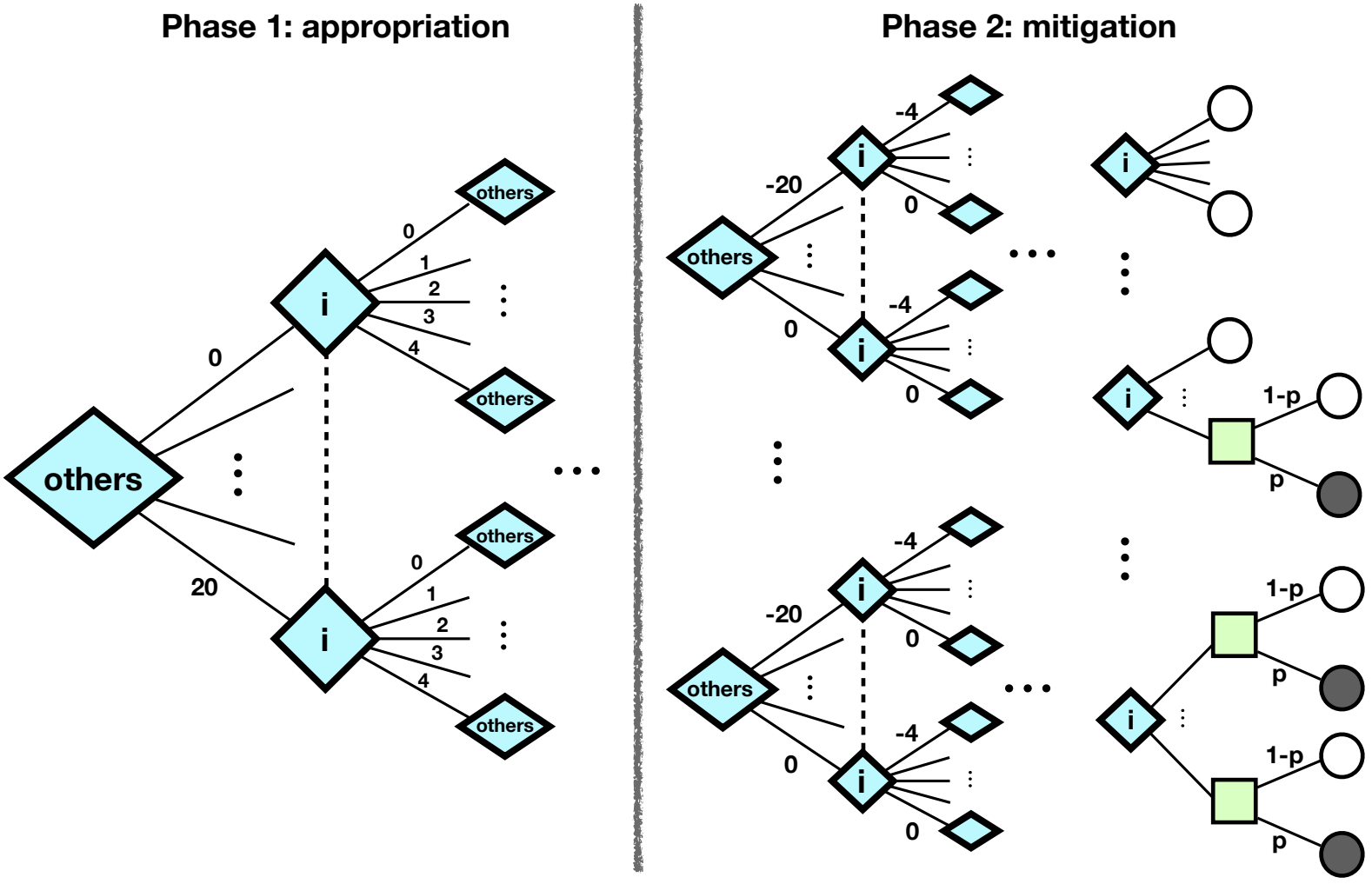
Everyone's choices are made public after each round.

# Game specification

- Two between-subject treatments
  - Baseline development
  - Endogeneous differentiated development

- Players in the US and China

# Computing responsibility



# Computing responsibility

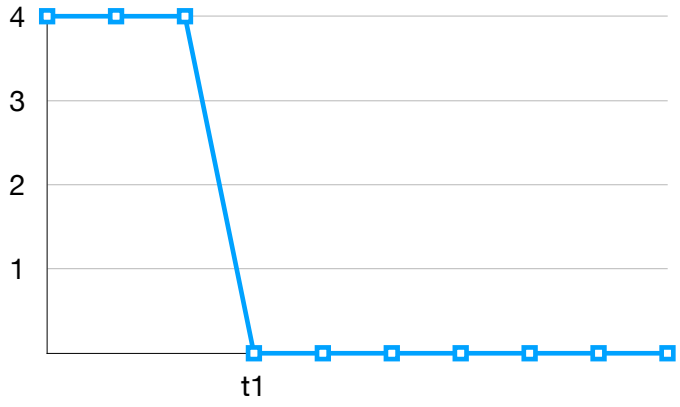
Except for limit cases (that do not occur in the observed situations), responsibility in phase one is as follows:

- If we are not in reach of any of the thresholds: 0
- When the first appropriation threshold might be crossed: $\frac{1}{3}$
- When the second appropriation threshold might be crossed: $\frac{1}{4}$
- When the last appropriation threshold might be crossed: $\frac{1}{6}$

Unless agents choose 0 appropriation, in which case the responsibility is also 0

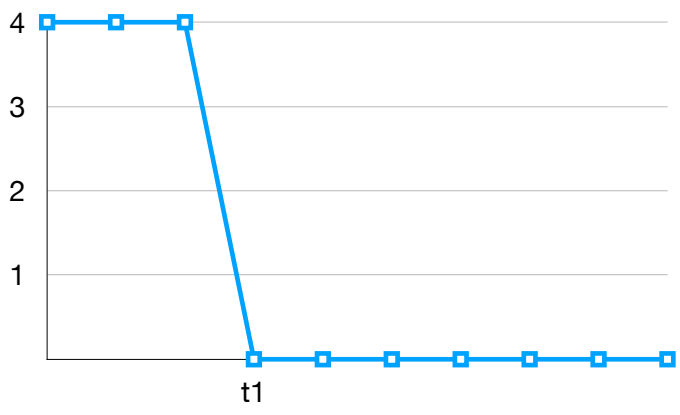
# Expected behaviour change

Always ensure $\mathcal{R} = 0$



# Expected behaviour change

Always ensure $\mathcal{R} = 0$




Instead:

$$\mathbf{a}_{i,t} = \begin{cases} 0 & \text{with probability} \\ & p = \lambda \mathcal{R}(v, \mathbf{nd}_t) \\ \mathbf{nd}_t & \text{else} \end{cases}$$

where $\mathbf{nd}_t$ is the mean of what agents selected in the experiments in the *non-differentiated* case in round $t$.

# Expected behaviour change

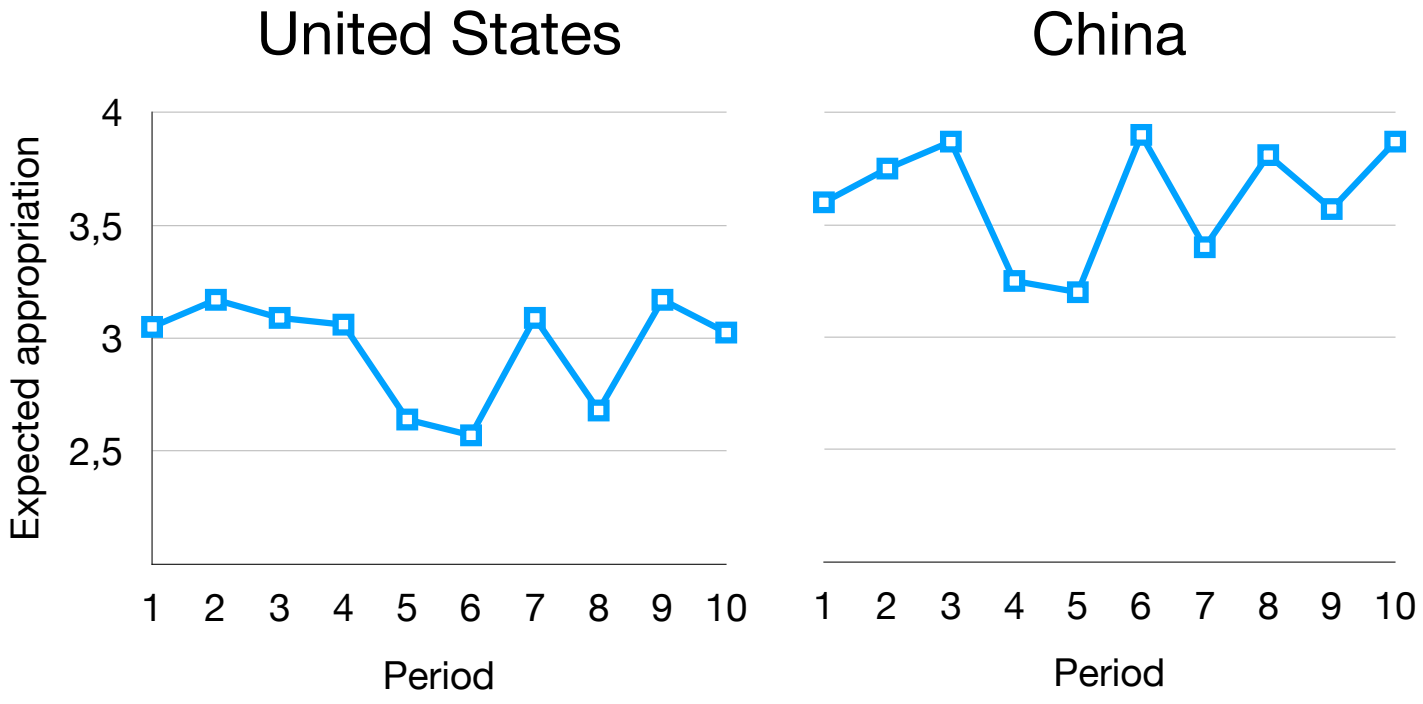
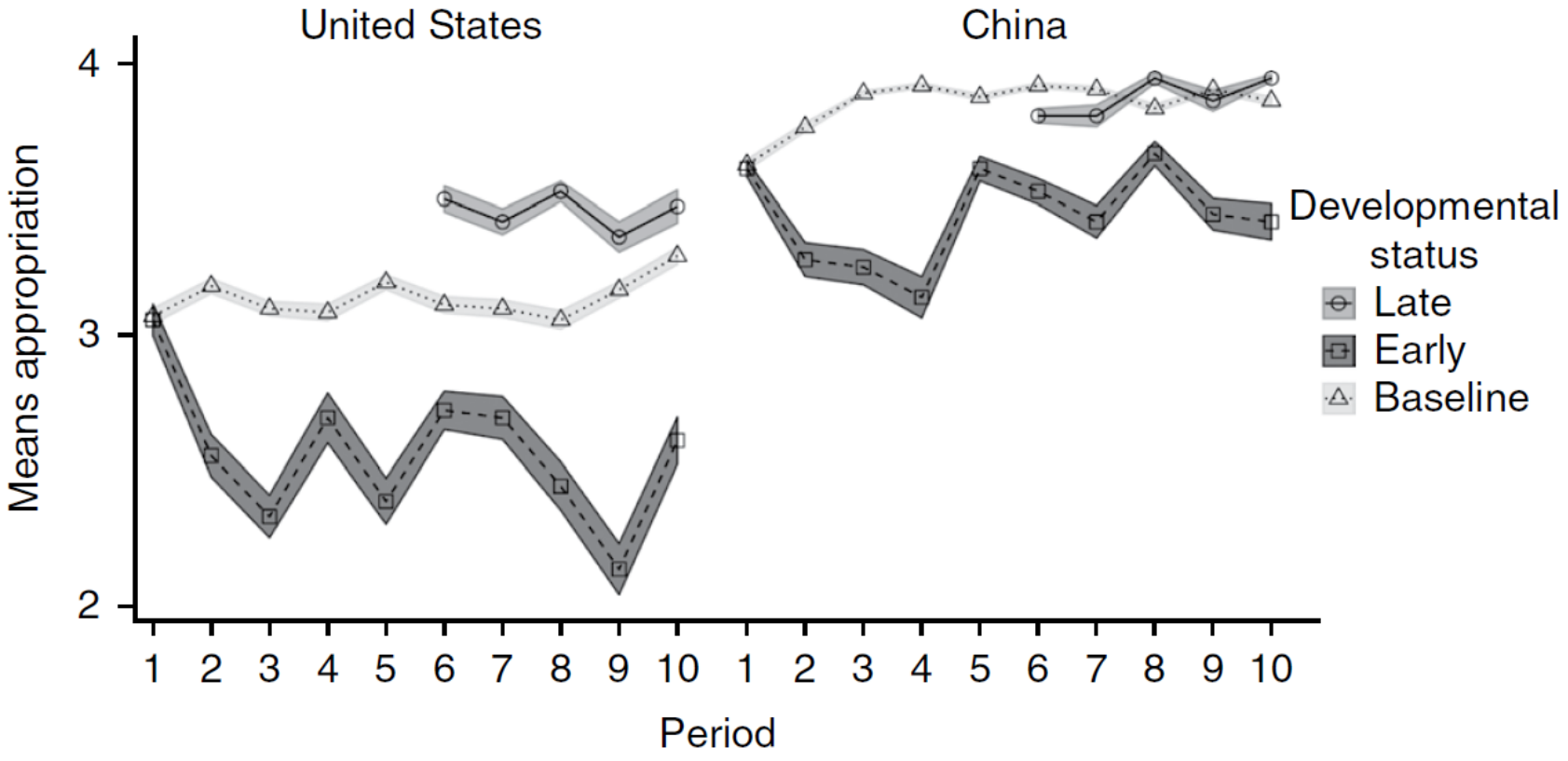



Figure: Expected value of the appropriation of the early developer group, $E[\mathbf{a}_{i,t} \mid \lambda = 0.5]$.

# Experimental Results




Results for mean appropriation per period in both treatment groups, taken from Kline et al. (2018)

# Discussion and Future Work

## Discussion

- Curves are shifted between experimental results and computed expectation - possibly due to agents acting according to *expectations*

  $\Rightarrow$ We will not consider this, for normative reasons
- No account of *partial contribution* in our framework

  $\Rightarrow$ Could include in future variant of a responsibility function

## Future work

- Application with other games
- Extend responsibility function accordingly