The Effect of Incentives & Leadership in Non-Routine Analytical Team Tasks

Florian Englmaier (LMU Munich)

with Stefan Grimm (LMU Munich), *Dominik Grothe (LMU Munich)*, David Schindler (Tilburg University) and Simeon Schudy (LMU Munich)

Keynote: *IGL Global Conference*

Berlin, May 21, 2019

Jobs in the industrial age...

Jobs today...

Motivation

Four broad categories of tasks (Autor et al., 2003)

- Routine manual tasks
- Routine cognitive tasks
- Non-routine manual tasks
- Non-routine analytical and interpersonal tasks

Motivation

Four broad categories of tasks (Autor et al., 2003)

- Routine manual tasks
- Routine cognitive tasks
- Non-routine manual tasks
- Non-routine analytical and interpersonal tasks

Main finding

- Strong and persistent increase in prominence of non-routine analytical and interpersonal tasks since the 60ies (Autor et al., 2003; Autor and Price, 2013)

▸ Non-routine tasks in the US

Ongoing work

What will kids do in the future?

An amazing John Oliver Video
(https://youtu.be/_h1ooyyFkF0)

Motivation

These non-routine analytical team tasks are often performed in **teams with flat hierarchies**

Think of teams of engineers, software developers, doctors, or lawyers.

Motivation

These non-routine analytical team tasks are often performed in **teams with flat hierarchies**

Think of teams of engineers, software developers, doctors, or lawyers.

However, causal evidence on the effectiveness of incentives and leadership in non-routine tasks is scarce.

Motivation

These non-routine analytical team tasks are often performed in **teams with flat hierarchies**

Think of teams of engineers, software developers, doctors, or lawyers.

However, causal evidence on the effectiveness of incentives and leadership in non-routine tasks is scarce.

The larger agenda
Understand the importance of ...

- incentives (bonuses, status tournaments, ...)
- leadership

for team performance and team organization in non-routine tasks.

Lack of evidence

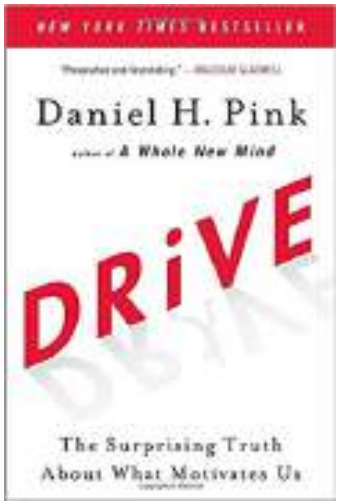
We know little about how to motivate workers in non-routine analytical team tasks.
Evidence mainly on routine individual tasks, often literally from the field:
Erev et al. (1993): Fruit pickers, Shearer (2004): Tree planters, Bandiera et al. (2005, 2007, 2009, 2013), Englmaier et al. (2016): Farm workers, Hossain & List (2012): Manufacturing workers, Delgaauw et al.(2015), Friebel et al. (2017): Retail chains

▸ Lab experiments on creativity, incentives for teachers and employees' ideas

Lack of evidence

We know little about how to motivate workers in non-routine analytical team tasks.
Evidence mainly on routine individual tasks, often literally from the field:
Erev et al. (1993): Fruit pickers, Shearer (2004): Tree planters, Bandiera et al. (2005, 2007, 2009, 2013), Englmaier et al. (2016): Farm workers, Hossain & List (2012): Manufacturing workers, Delgaauw et al.(2015), Friebel et al. (2017): Retail chains

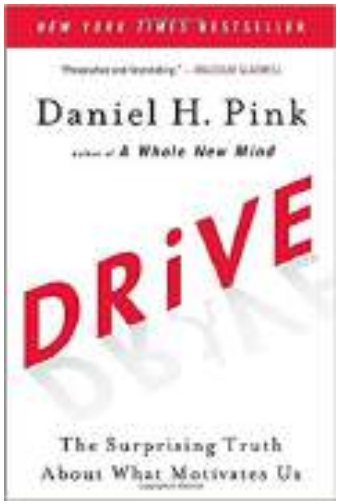
▸ Lab experiments on creativity, incentives for teachers and employees' ideas



Lack of evidence

We know little about how to motivate workers in non-routine analytical team tasks.
Evidence mainly on routine individual tasks, often literally from the field:
Erev et al. (1993): Fruit pickers, Shearer (2004): Tree planters, Bandiera et al. (2005, 2007, 2009, 2013), Englmaier et al. (2016): Farm workers, Hossain & List (2012): Manufacturing workers, Delgaauw et al.(2015), Friebel et al. (2017): Retail chains

▸ Lab experiments on creativity, incentives for teachers and employees' ideas



Popular claim:
Incentives do not work

NYT # 1 bestseller
TED talk with >20mio views

Our contribution

A series of **large scale** (in total $> 6,000$ participants!) **field experiments** in a setting that

- resembles **non-routine analytical team tasks** but
- allows for an **objective, quantitative measurement** of performance.

Today, a glimpse at 3 papers on the topic:

1. Effects of bonus incentives
2. Effects of leadership
3. Why and how does leadership work

Our cooperation partner

Our partner is *ExitTheRoom* (ETR), a provider of *real-life escape games* with locations in Germany, Austria, Hungary, and Saudi-Arabia.

- In real-life escape games, participants try to complete a challenge within a given time limit.
- Teams try to 'escape' from a room (e.g. by finding a door code) or solve some other task (e.g. defuse a fictitious bomb).
- To achieve their goal, they have to find clues, collect information, and re-combine those in innovative ways.

Experimental design and procedures

The setting

We conduct the field experiments with customers of *ETR* at their Munich location.

- Participants pay €79-119 for a group of usually 2-6 people.
- Participants have up to 60 minutes to escape one of three distinct rooms.
- If participants get stuck, they can request up to five free hints from ETR staff.
- Teams are highly motivated to solve the task quickly:
  They proudly write their remaining times at the wall in the entrance area.

Experimental design and procedures

The setting

Study 1: Effects of bonus incentives

Study 1: Effects of bonus incentives

Main treatments

Two main treatments:

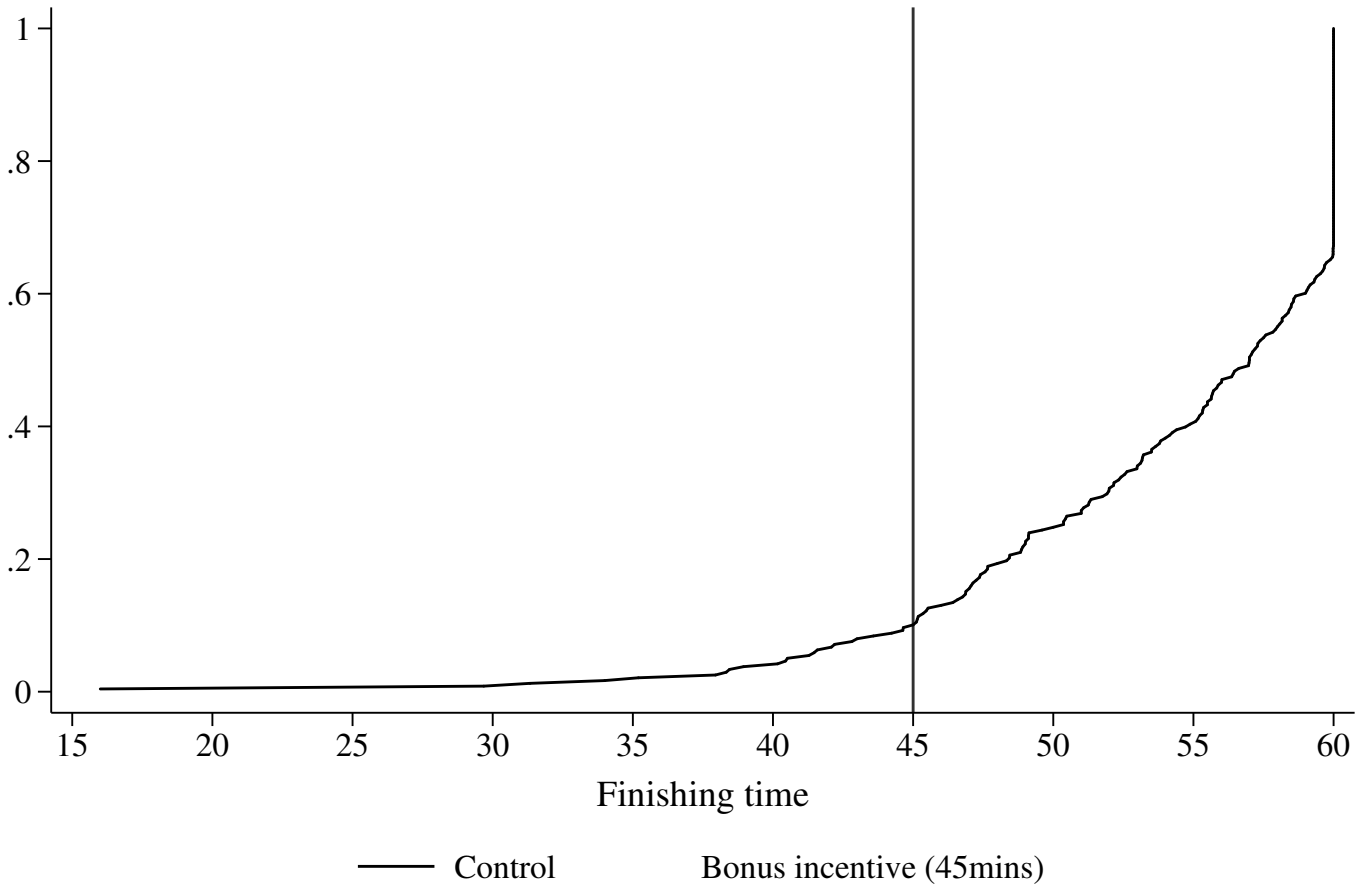
- *Control*: No bonus paid to subjects. (238 groups)
- *Bonus incentive (45mins)*: If subjects manage to escape the room in less than 45 minutes, they receive a €50 bonus (for the team). (249 groups)

- 2250 Regular ETR customers (487 groups)
- Randomization on daily basis
- Data collection from Nov to Dec ’15 and Jan to May ’17

▸ Task perception ▸ Sample Balance

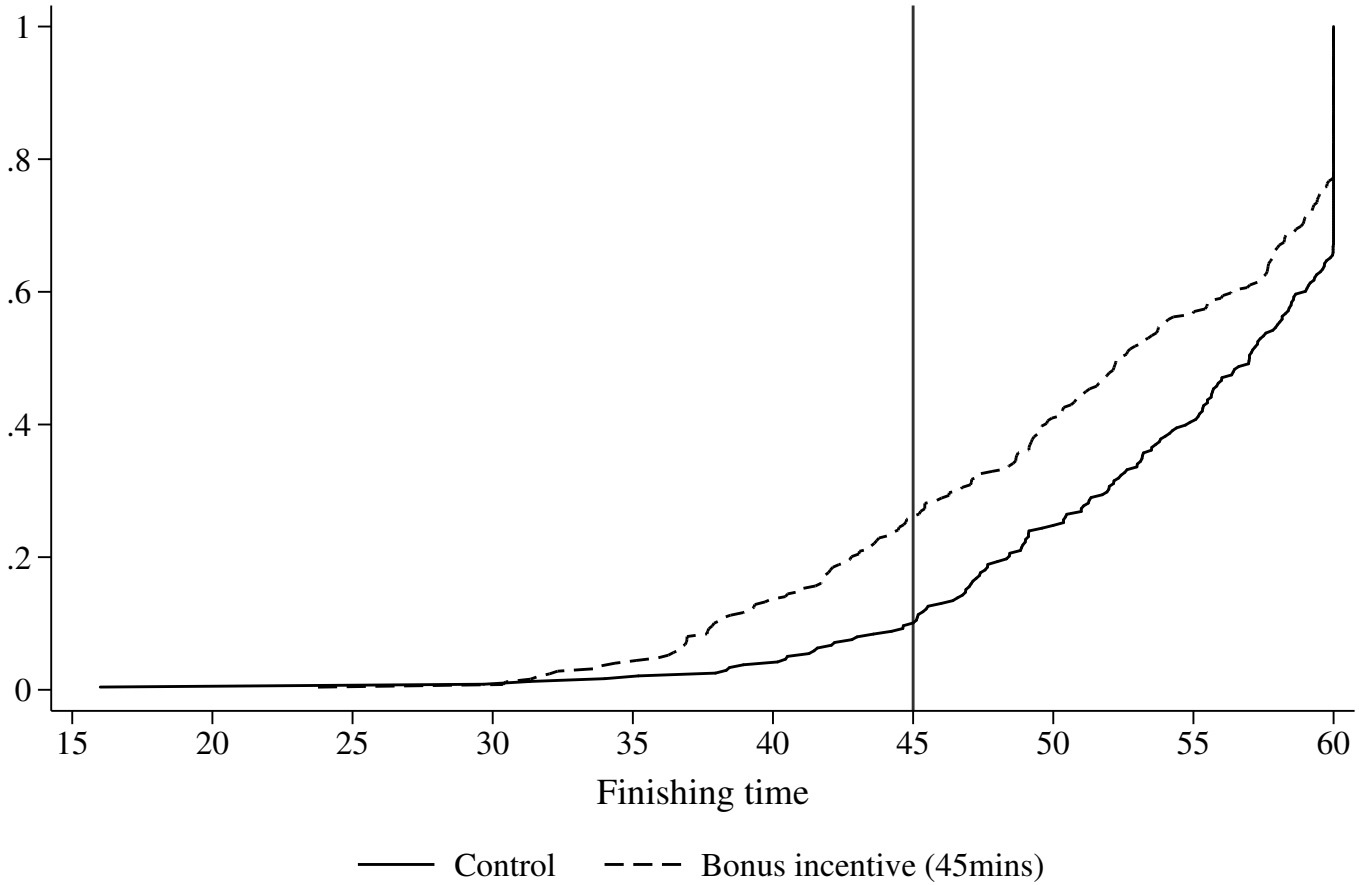
Bonus Incentives: Main result

CDF of finishing time



Bonus Incentives: Main result

CDF of finishing time



▸ Test statistics

Bonus Incentives: Main result

Regression analyses: Game solved in less than 45 minutes

	Probit: Solved in less than 45 minutes			
	(1)	(2)	(3)	(4)
Bonus incentives (45 minutes)	0.661***	0.715***	0.841***	0.709***
	(0.141)	(0.151)	(0.170)	(0.268)
Constant	-1.301***	-3.267***	-3.121***	-7.662***
	(0.112)	(0.542)	(0.637)	(0.948)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	487	487	487	487

Coefficients from Probit regressions of whether a team solved the game within 45 minutes on our treatment indicator (with *Control* as base category). Control variables added from column (2) onwards include group size (number of team members), share of males in a team, a dummy whether someone in the team has been to an Escape Game before, median age of the team, a dummy whether all group members speak German and a dummy for private teams (opposed to company team building events). Staff fixed effects control for the employees of *ExitTheRoom* present onsite. Robust standard errors reported in parentheses, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Note: Incentive effect does not significantly interact with observable group characteristics.

▸ Remaining time ▸ Proportional Hazard Model

Bonus Incentives: Student sample

Do bonus incentives increase performance also for teams who are exogenously formed and confronted with the task?

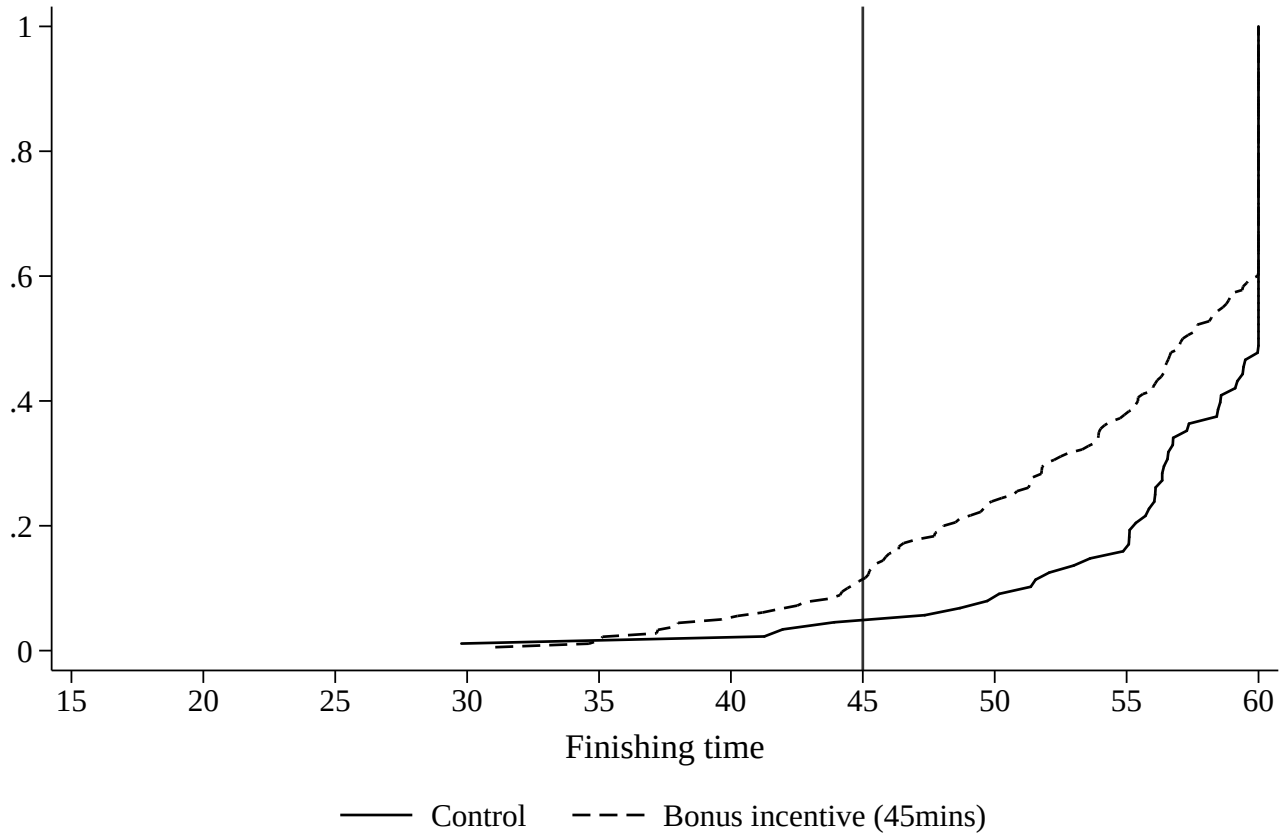
Replication of main treatments (*Gain45*, *Loss45* and *Control*) with student participants (268 groups, i.e. 804 individuals)

1. Participants are invited to the local econ lab
   (invitation does not contain information about escape game)
2. In the lab, we elicit personal characteristics and socioeconomic background using surveys and experimental tasks (30mins)
3. Participants are brought to ETR (6mins walk) and randomly allocated into groups of three to perform the task at ETR
4. After the task, participants answer a survey on behavior and team organization and receive payments

▸ Sample Balance

Bonus Incentives: Student sample results

CDF of finishing time



- Bonus incentives "work" also for teams who are exogenously formed and confronted with the task.

▸ Compare ▸ Test statistics ▸ Regression

Bonus Incentives: Exploration

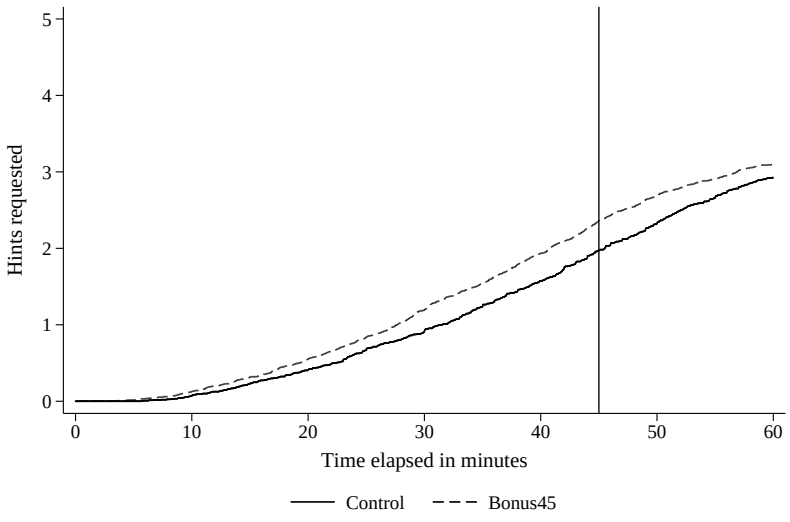
Do bonus incentives reduce teams' willingness to explore original solutions?

We use hints as an indication of teams' unwillingness to explore and compare reactions to incentives by

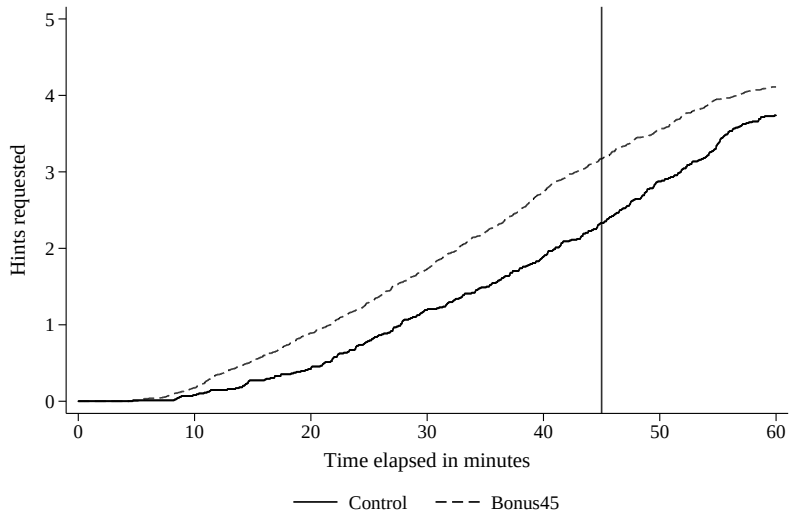
- endogenously formed teams who chose to perform the non-routine task
- exogenously formed teams who were confronted with the task

Bonus Incentives: Exploration

Customer sample



Student sample



- Exploration behavior of self-selected teams is not strongly affected.
- Teams we confronted with the task explore far less!

▸ Regression analysis

Bonus Incentives: Channel

- This is a task where incentives are unlikely to just trigger more sweat and toil...
- Survey evidence from ex-post questionnaires suggests that incentives may lead to an **endogenous emergence of leadership** and hierarchies.

▸ Survey 1 ▸ Survey 2

Study 2: Effects of Leadership

The experiment

Wording of treatments

"One tip before you start: a good team needs a good leader. Past experience has shown that less successful teams often wanted to have been better led. Thus, decide on someone of you, who takes over the leading role and …

- Sub-treatment *Motivation*: …consistently motivates the team."
- Sub-treatment *Coordination*: …consistently coordinates the team."

▸ Leadership Functions

The experiment

Wording of treatments

"One tip before you start: a good team needs a good **leader**. Past experience has shown that less successful teams often wanted to have been better **led**. Thus, decide on someone of you, who takes over the **leading** role and …

- Sub-treatment *Motivation*: …consistently *motivates* the team."
- Sub-treatment *Coordination*: …consistently *coordinates* the team."

Experimental design - sample

Sample

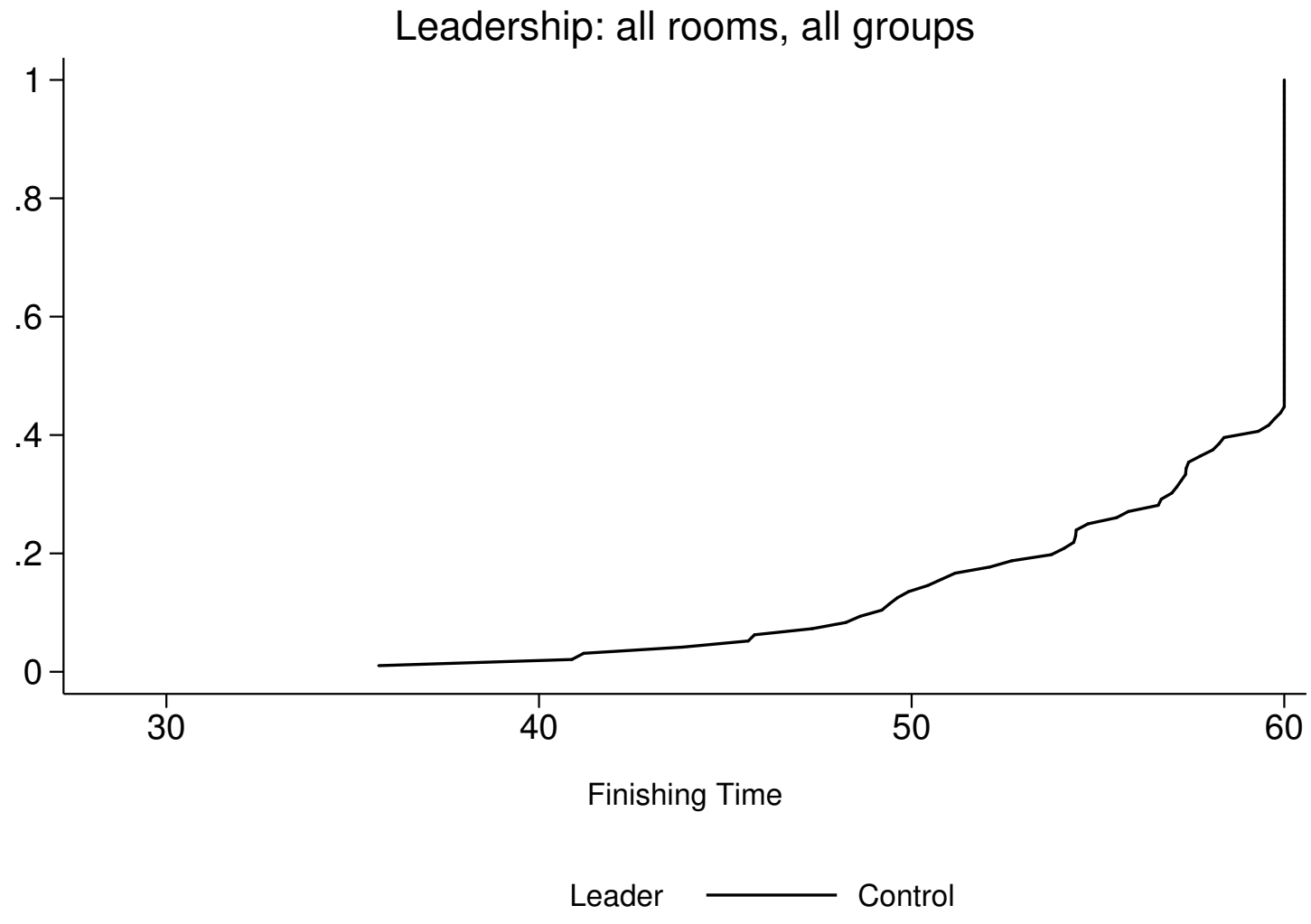
- 1273 Regular ETR customers (281 groups)
- Data collection from January 2018 to March 2018
- Randomization on daily basis
- Study was preregistered at AEA registry

Treatments

- *Control*: 95 groups
- *Motivation*: 95 groups
- *Coordination*: 91 groups

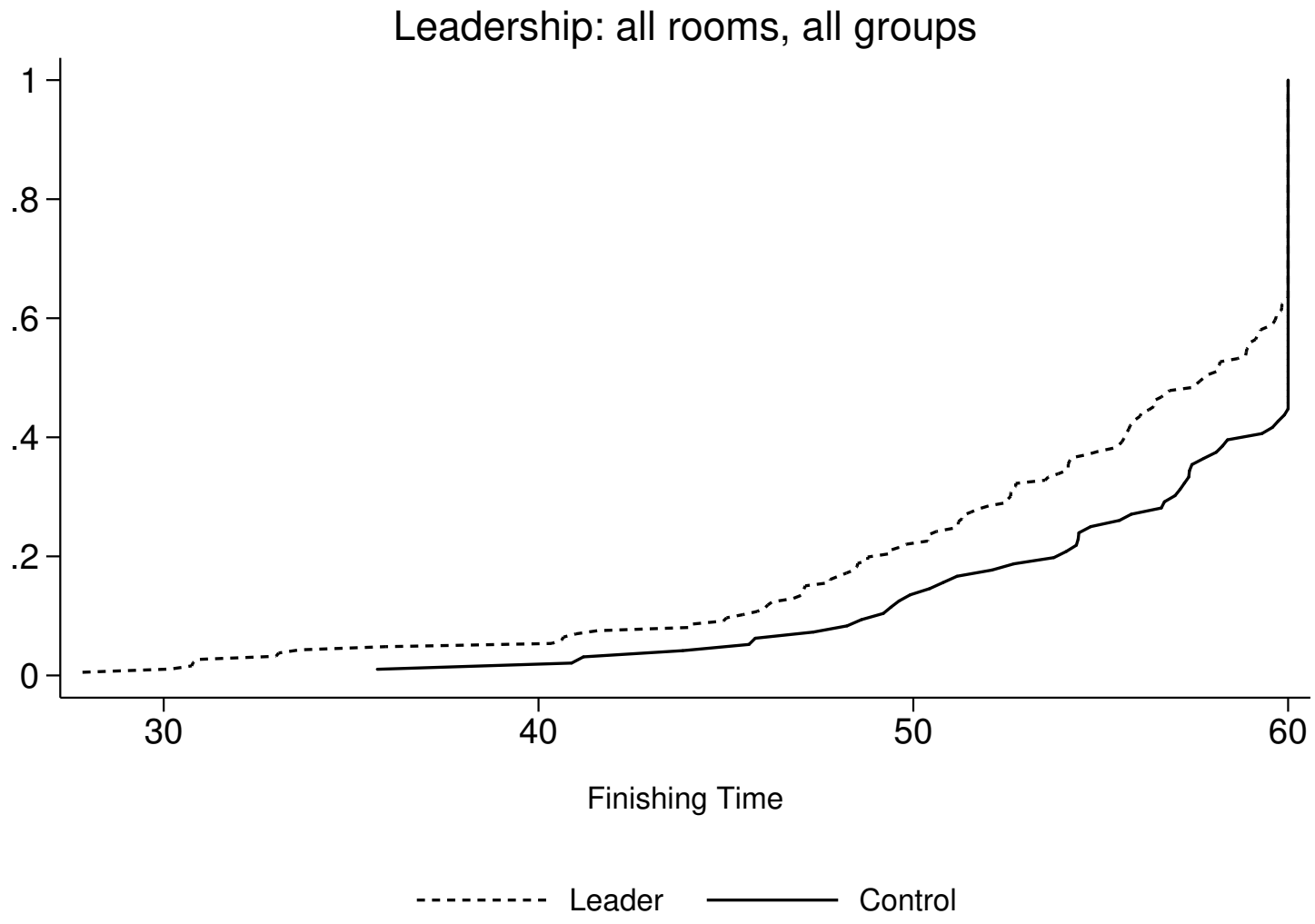
▸ Balance Table

Control



Non-parametric test:

Leadership (pooled treatments)



Non-parametric test: Mann-Whitney test (Prob $> |z| = 0.0038$)

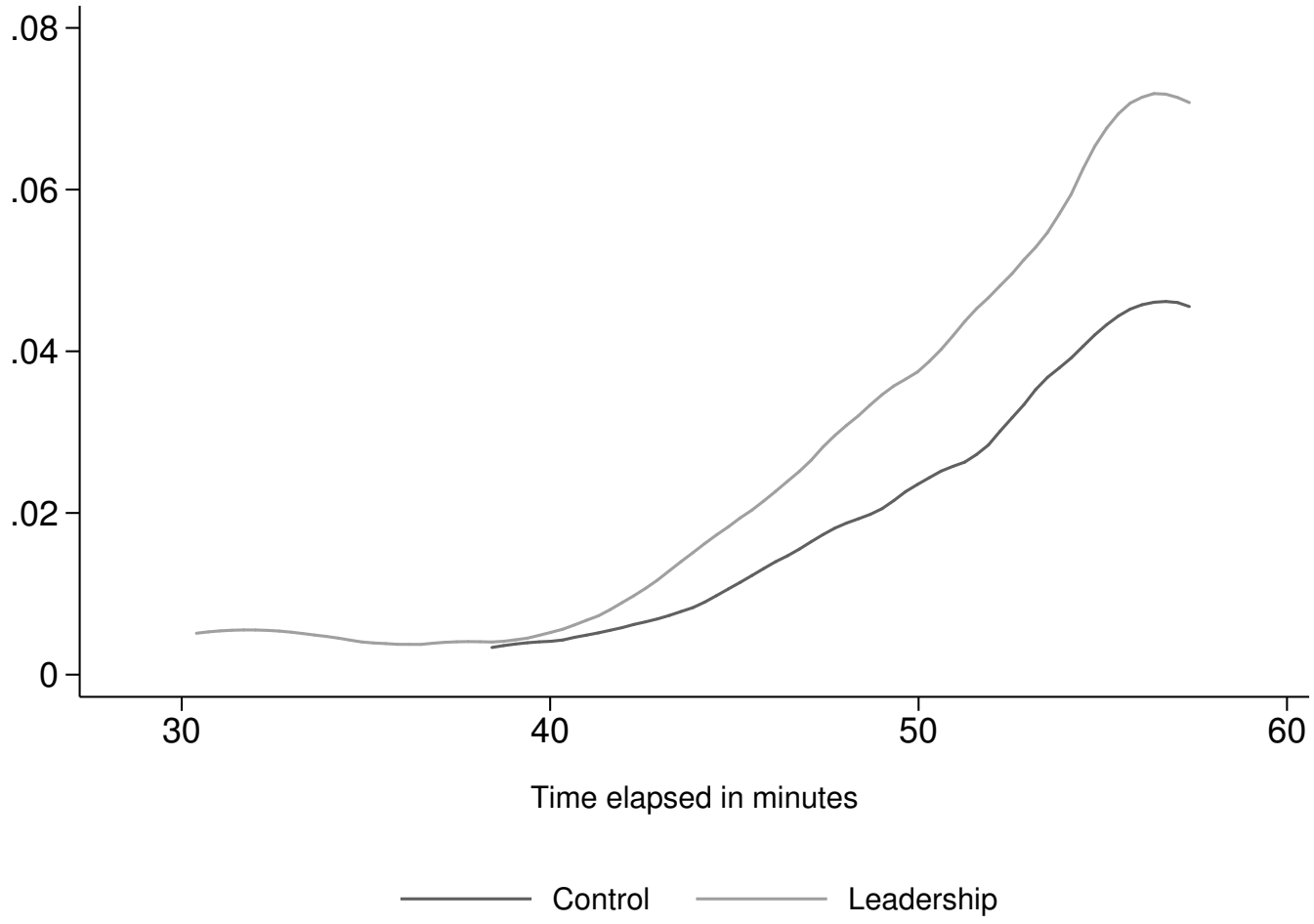
Solved within 60 minutes

	Probit (ME): Solved within 60 minutes			
	(1)	(2)	(3)	(4)
Leadership	0.177***	0.180***	0.222***	0.225***
	(0.052)	(0.054)	(0.064)	(0.051)
Fraction of control teams solving the task	0.442	0.442	0.442	0.442
Observations	281	281	281	281
Team Controls	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes
Room, Day and Week FE	No	No	No	Yes

Notes: The table displays average marginal effects from Probit regressions of whether a team solved the game within 60 minutes on our treatment indicator (with Control as base category). Team controls (group size, share of males, experience, median age, language, private, natural leader, walkie talkie), staff, room, day and week fixed effects are step-wise included. Standard errors in parentheses are clustered on the date level, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

▸ Show controls ▸ Linear Model ▸ Definition of Variables

Dynamics of Completion



The figure displays smoothed hazard rates (probability that a team solves the task in the next instant) conditional on the fact that a team did not solve the task up to time t.

Responsiveness (?) to treatment

Choosing a leader

- Roughly 50 percent of teams chose a leader immediately (i.e. before working on the task)
- Are the results driven by teams who chose the leader immediately?

Responsiveness (?) to treatment

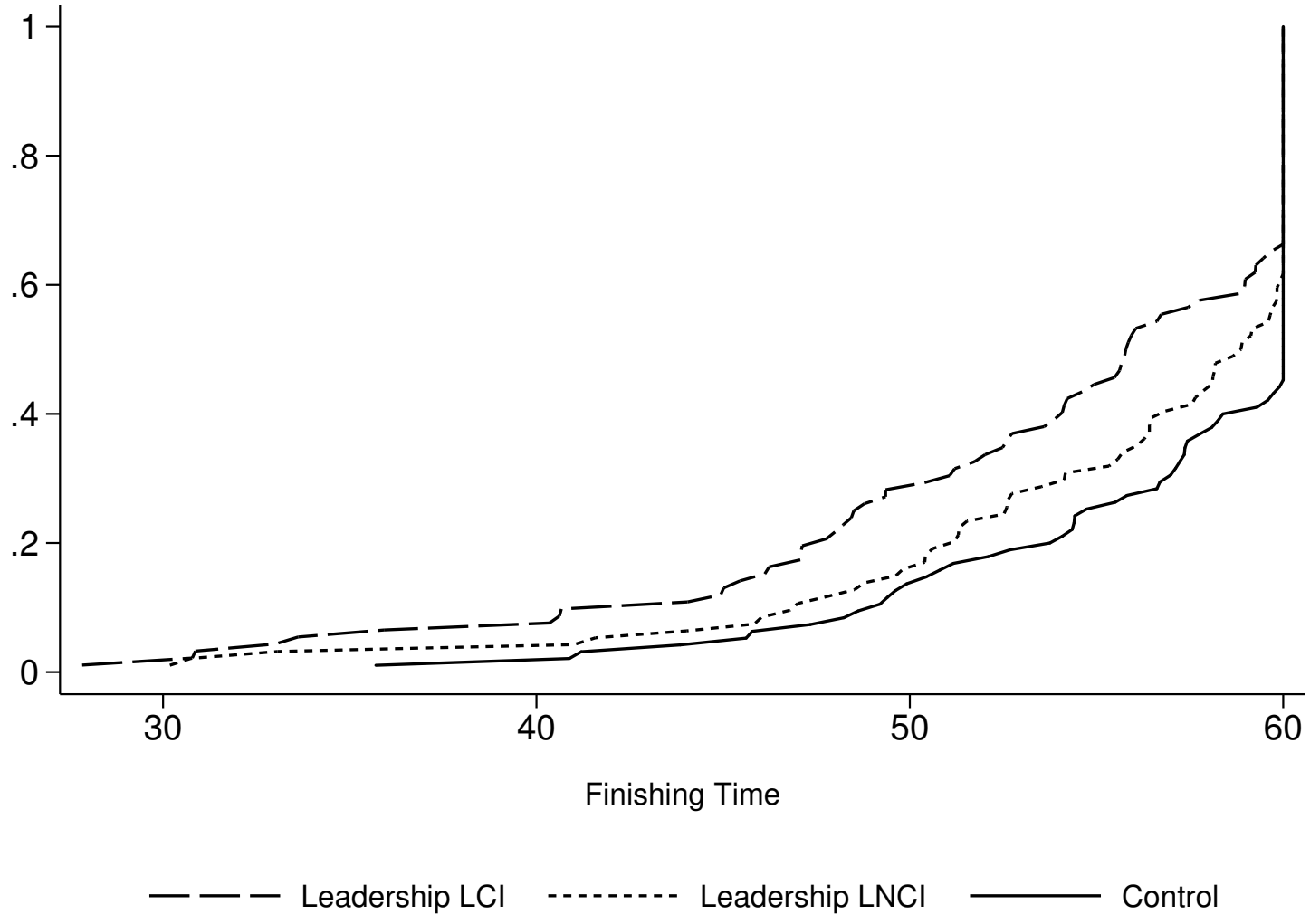
Choosing a leader

- Roughly 50 percent of teams chose a leader immediately (i.e. before working on the task)
- Are the results driven by teams who chose the leader immediately?

Results separately for...

- Leader chosen immediately (LCI)
- Leader not chosen immediately (LNCI)

Leadership (LCI and LNCI)

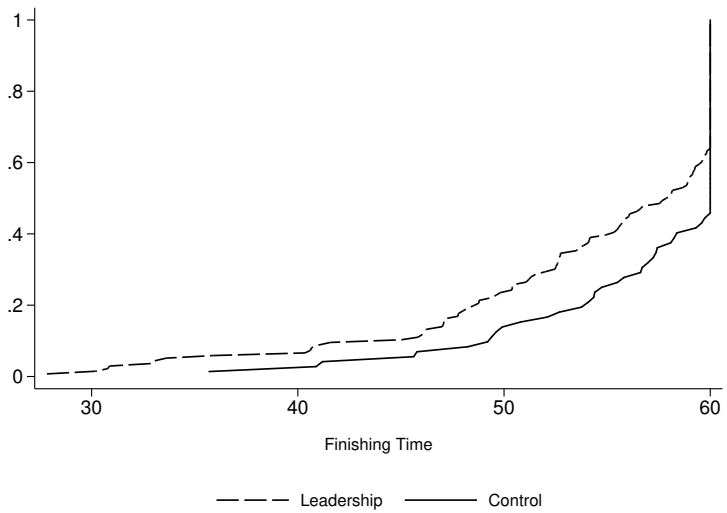


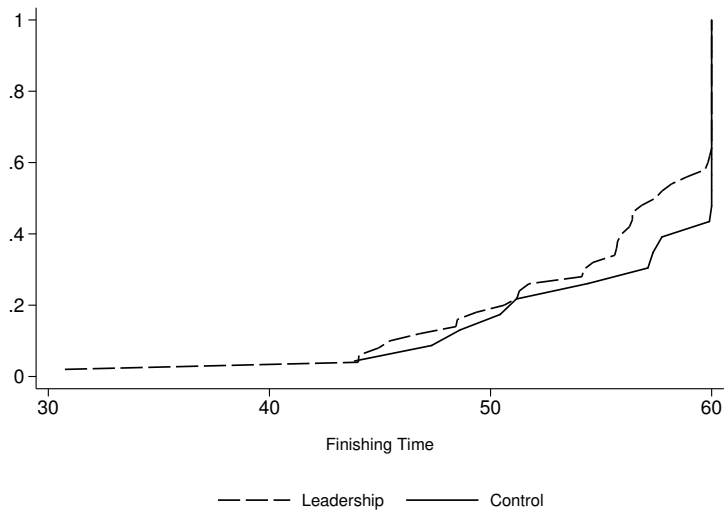
Notes: LCI = Leader chosen immediately, LNCI = Leader not chosen immediately

▸ Finishing Time

Private vs. Corporate Teams

The location we work with is popular both with private customers as well as with corporate customers that use it for teambuilding events





Originality

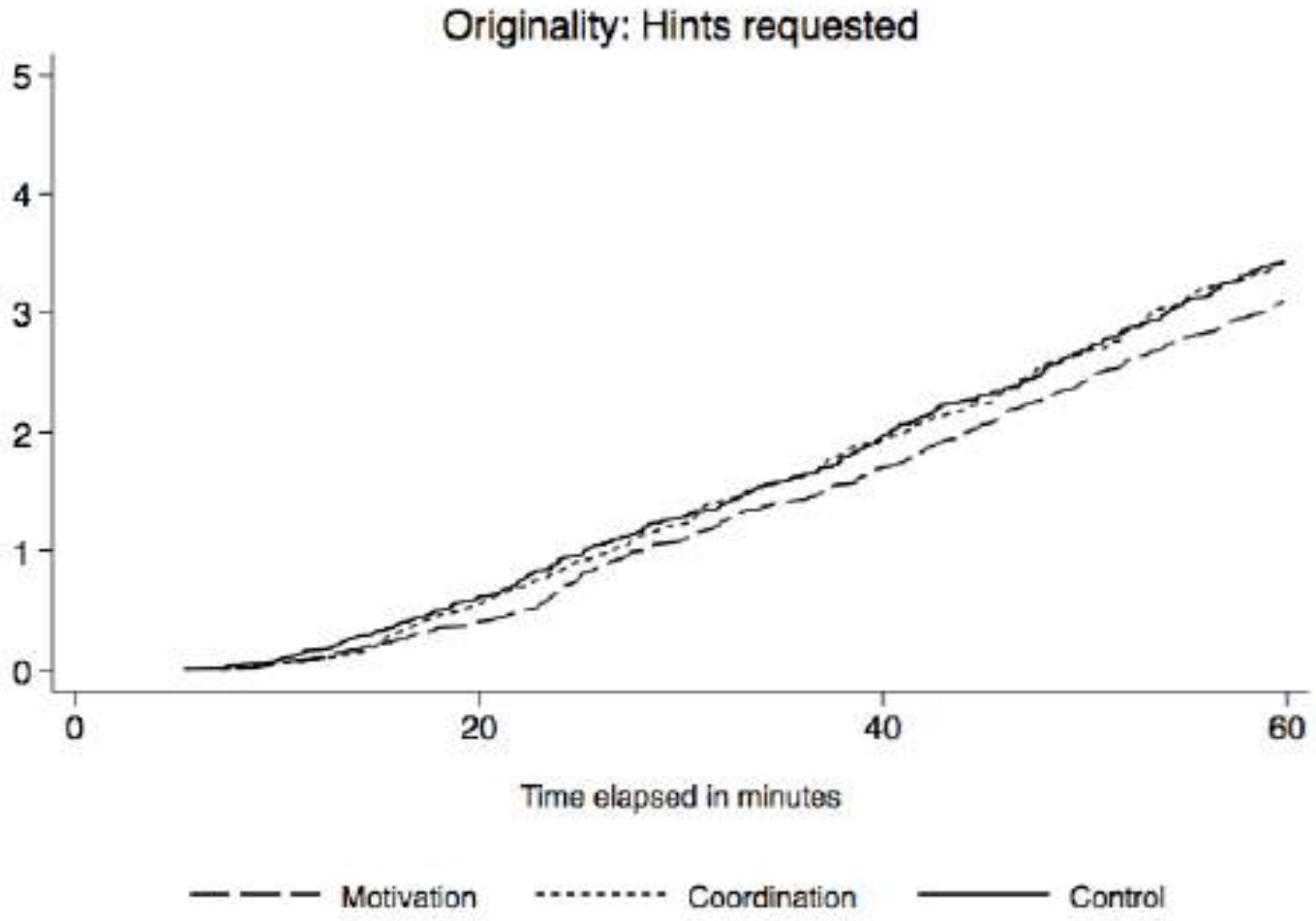
Leadership and exploration

We have seen that priming leadership increases performance

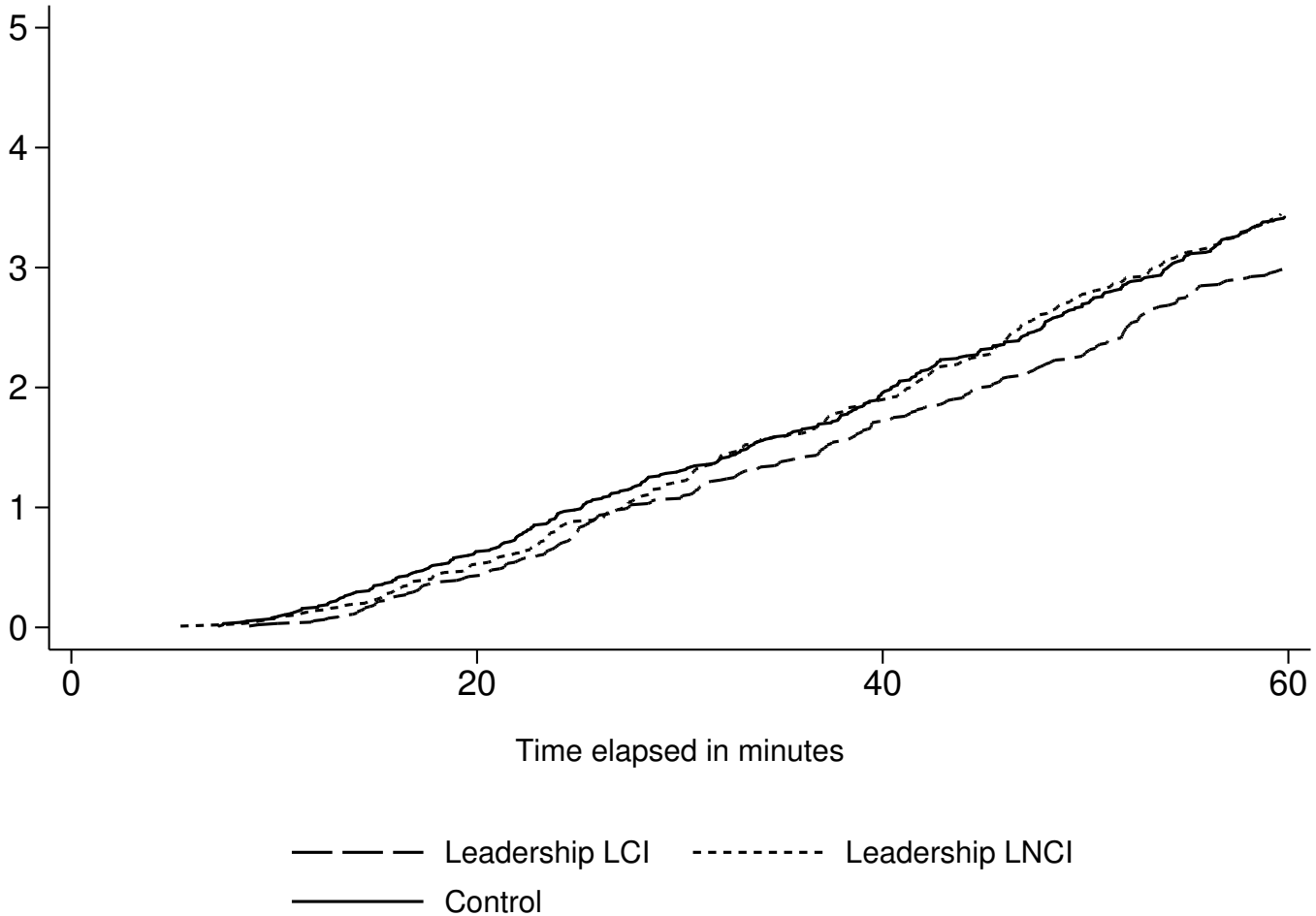
- Does our treatment affect teams' willingness to explore "innovative" solutions on their own?
- Do teams with leaders (i.e., in treatment) ask for more / less hints?

Originality

Avg. number of hints over time by treatment

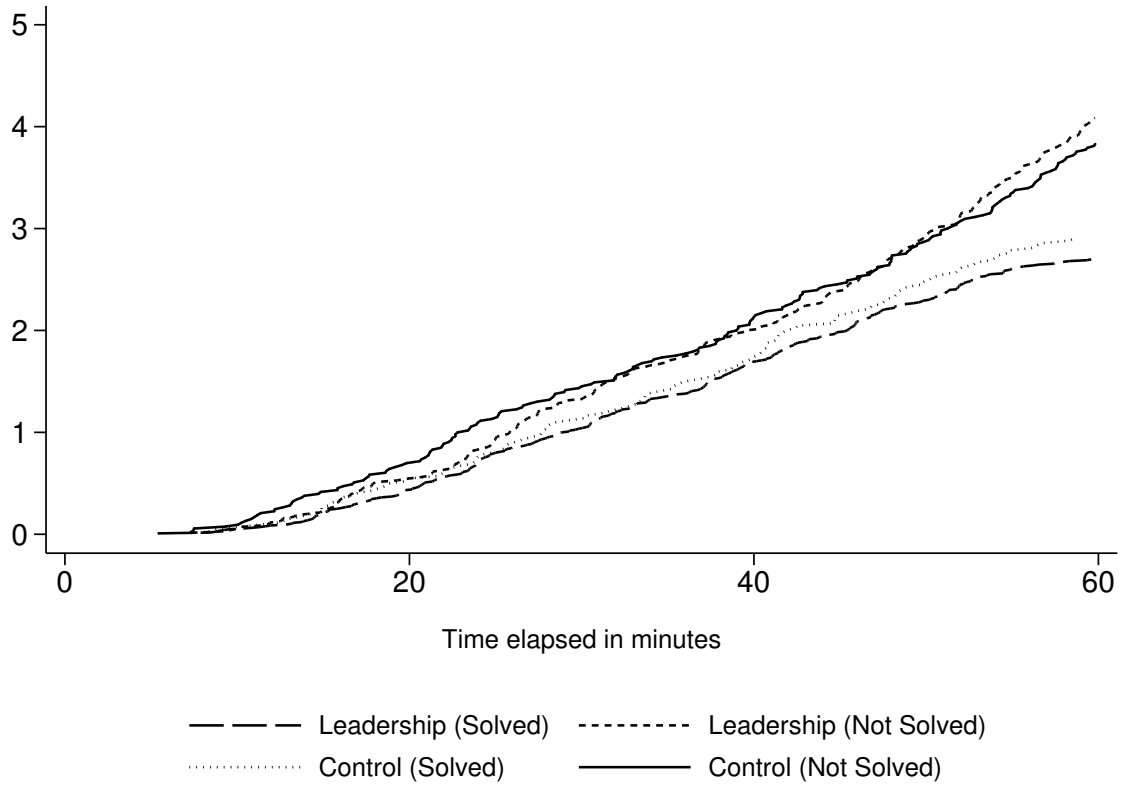


Avg. number of hints over time by LCI vs. LNCI



Originality

Avg. ♯ of hints (successful vs. non-successful teams)



Failing teams in treatment take the most hints.
Going out on a limb: Read this as …

- …leadership is good as they at least try to make it by all means given a bad team
- …leadership is bad as they fail and use up a lot of hints

Study 3: Why and how does leadership work

▸ Oops, I am way over time

Ongoing work

Endogenous vs. exogenous leadership

- New experiments with student sample

  (including rich information on a variety of background characteristics e.g. socioeconomic, creativity, risk,...)

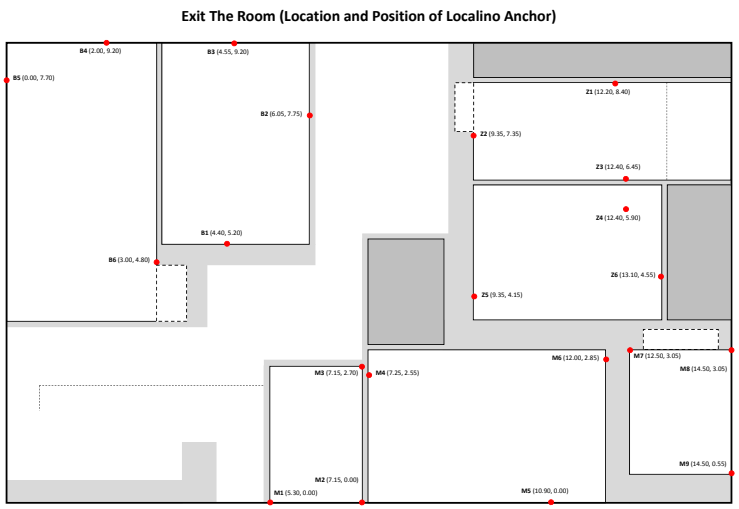
Ongoing work

Endogenous vs. exogenous leadership

- New experiments with student sample

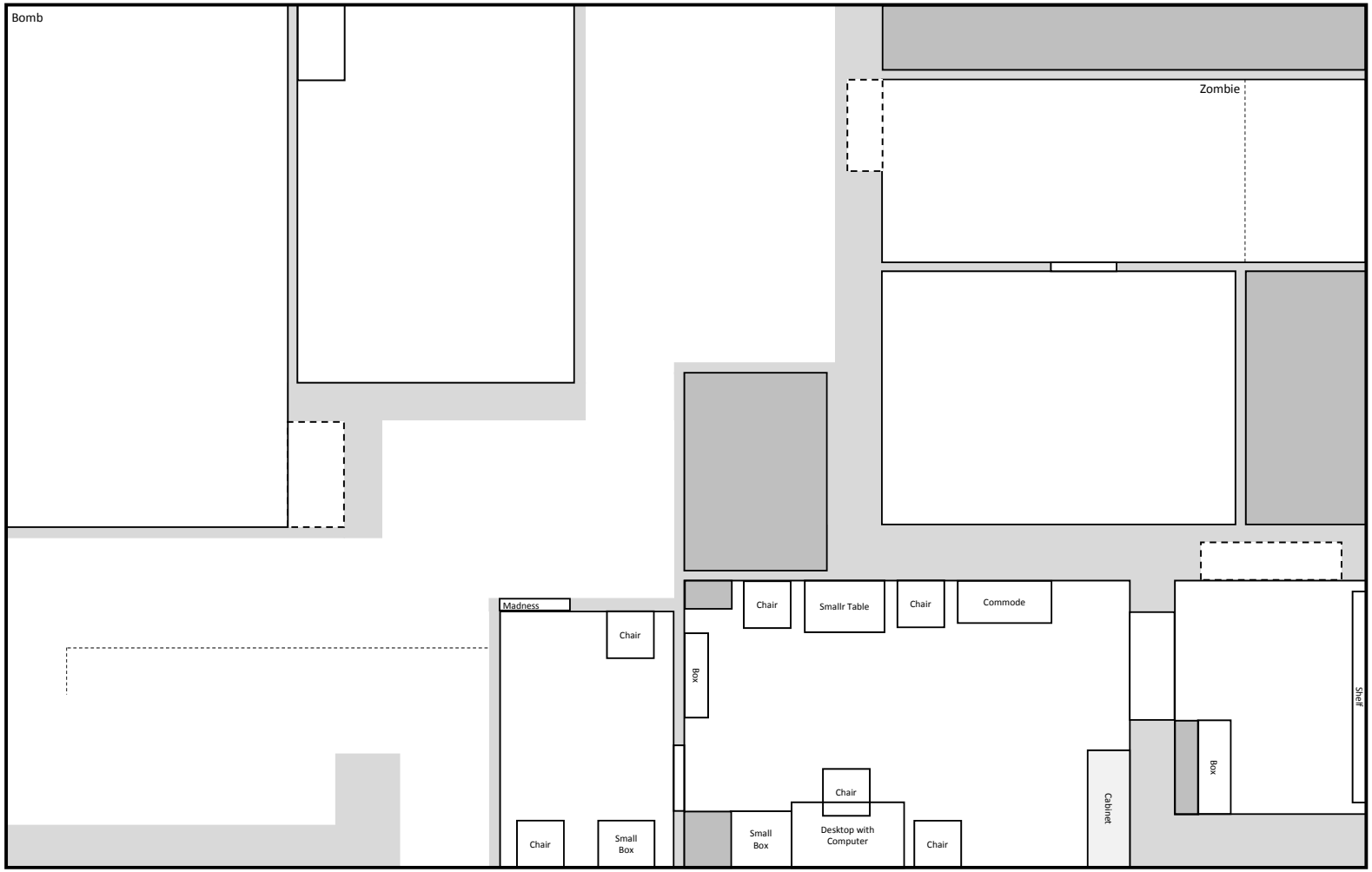
  (including rich information on a variety of background characteristics e.g. socioeconomic, creativity, risk,...)

- Detailed data on team organization (audio and location)
  - Objective measures for team organization
  - Communication structures, task specialization and leader's actions



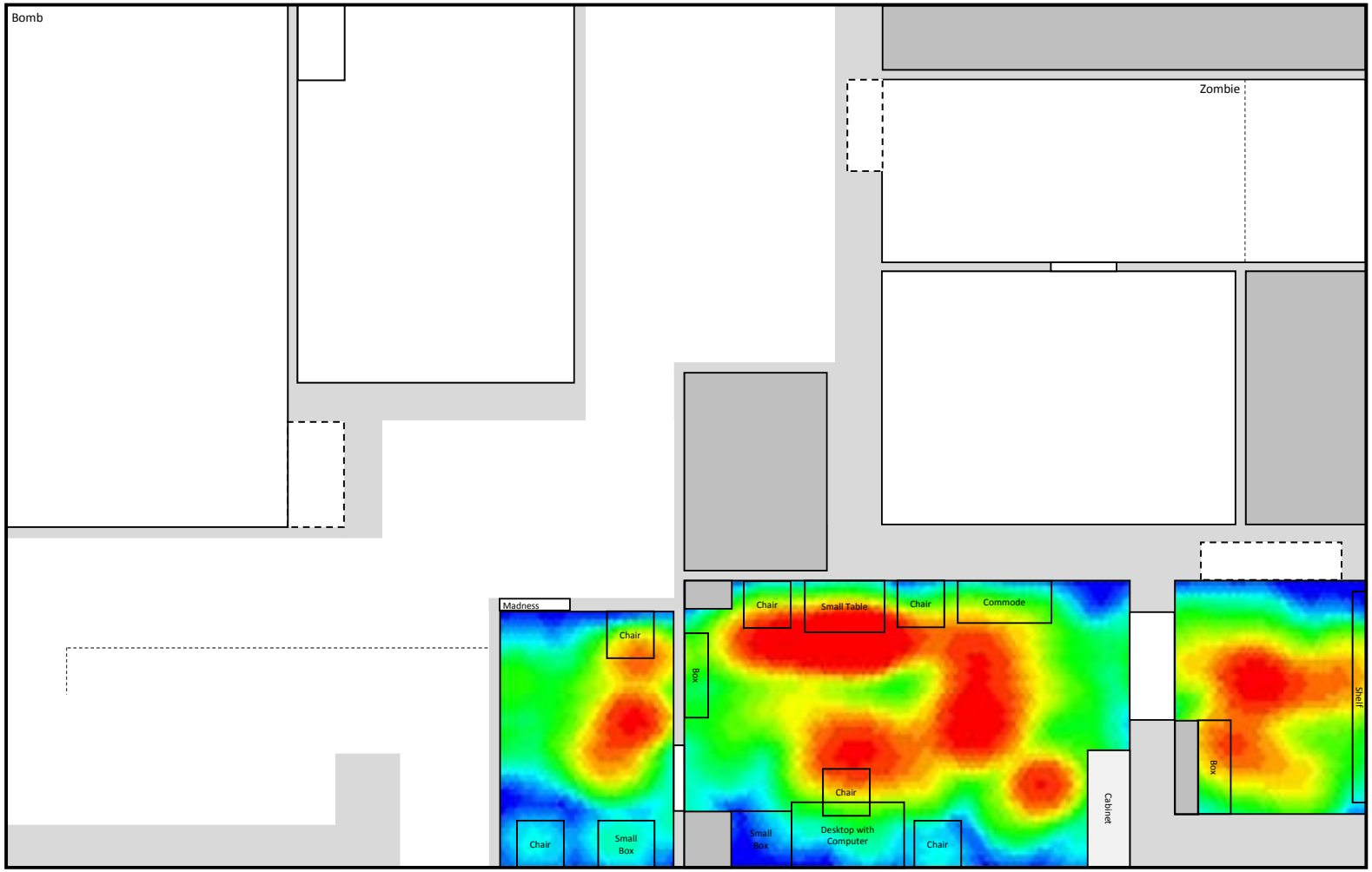
Endogenous vs. exogenous leadership

Exit The Room (Location)



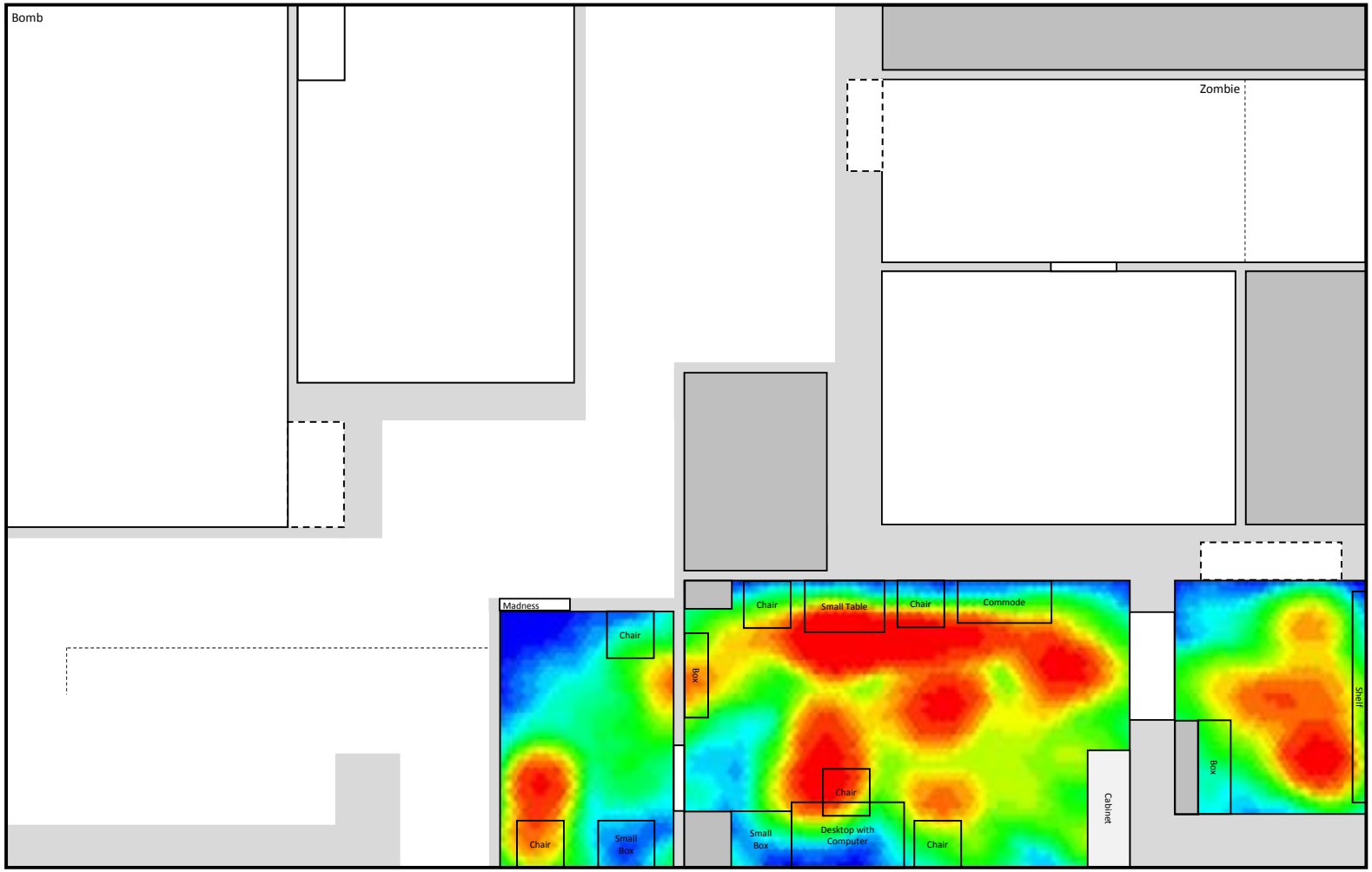
Endogenous vs. exogenous leadership

Exit The Room (Location and Heat Map) – Subject 1



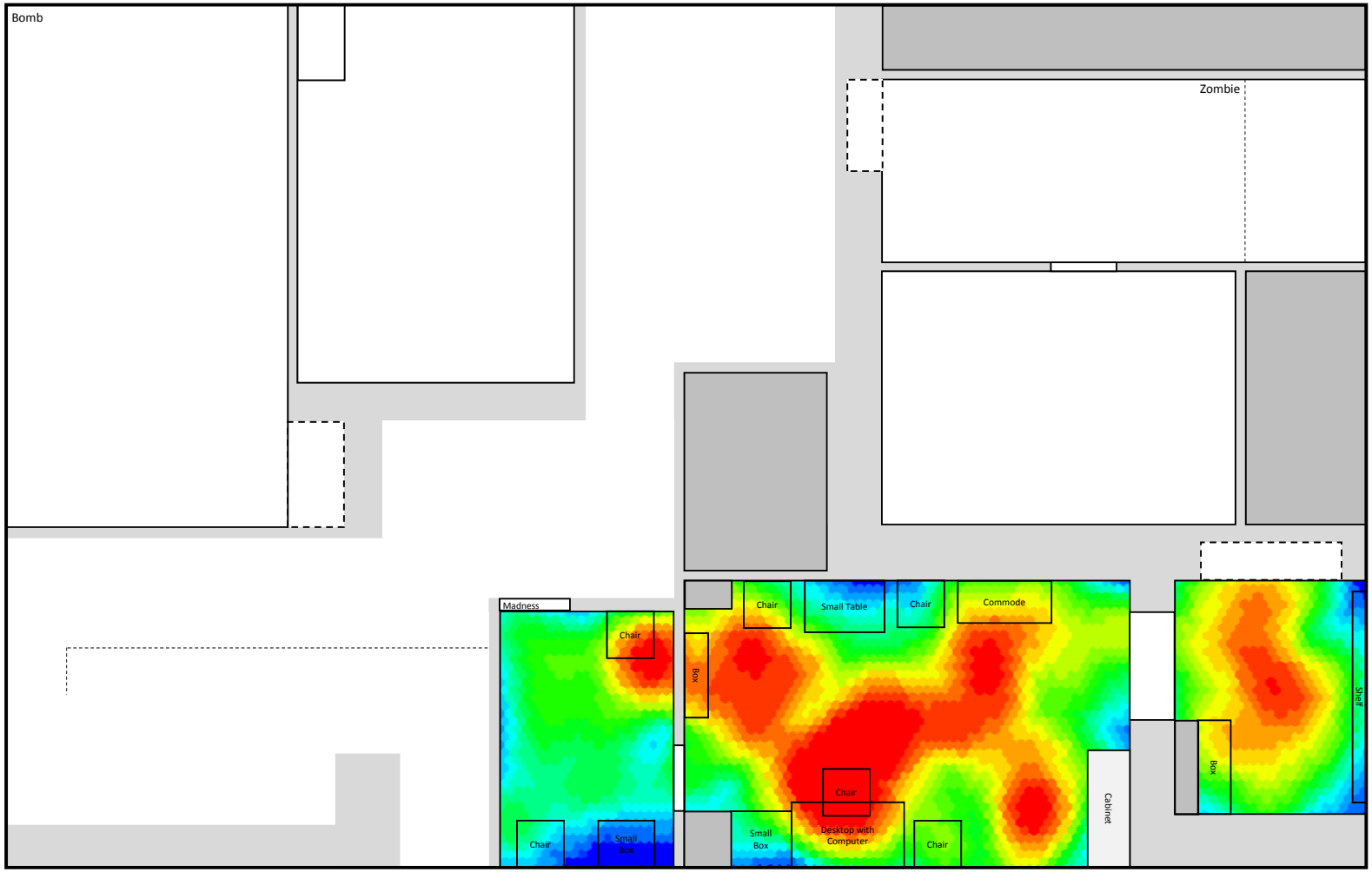
Endogenous vs. exogenous leadership

Exit The Room (Location and Heat Map) – Subject 2



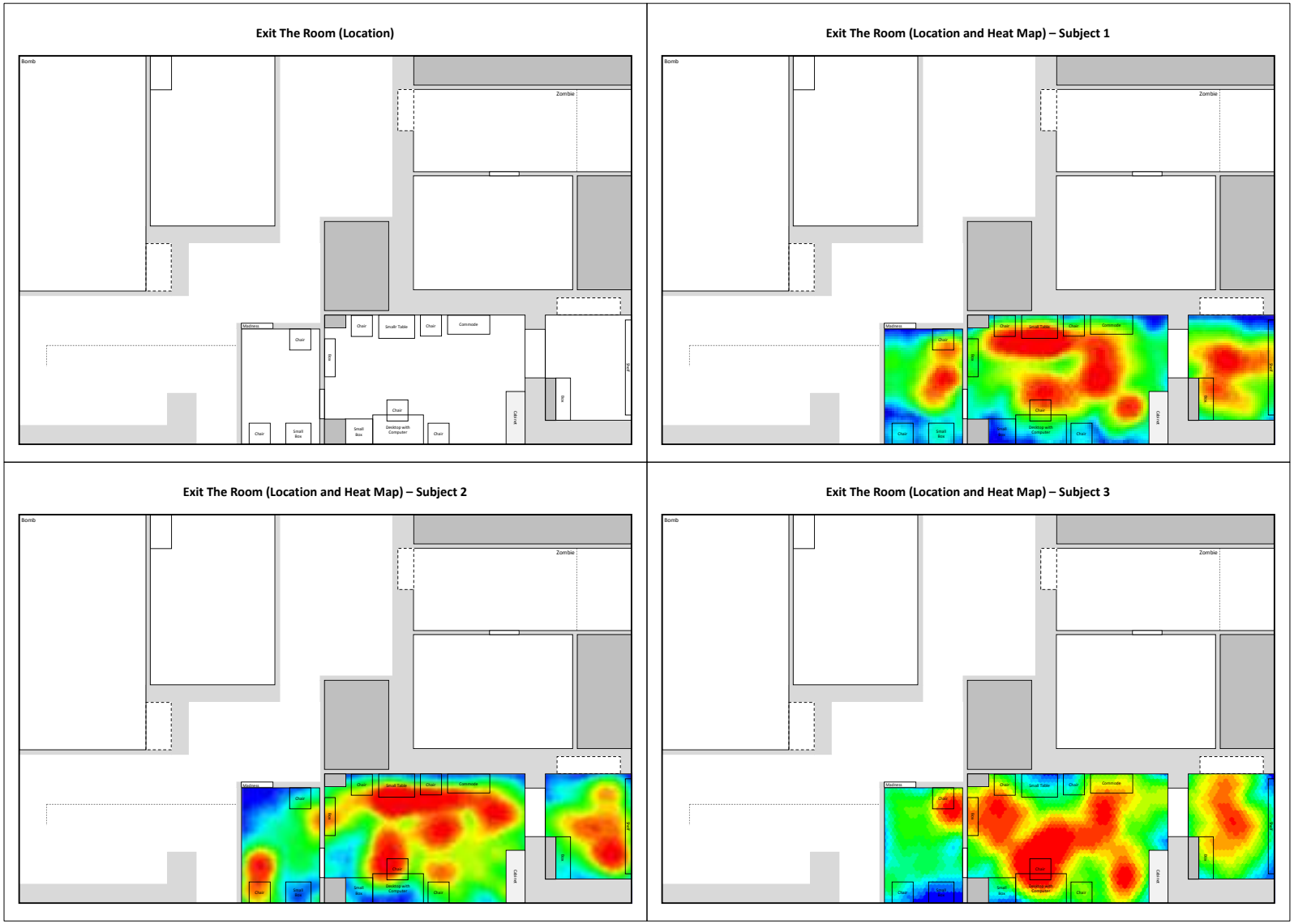
Endogenous vs. exogenous leadership

Exit The Room (Location and Heat Map) – Subject 3



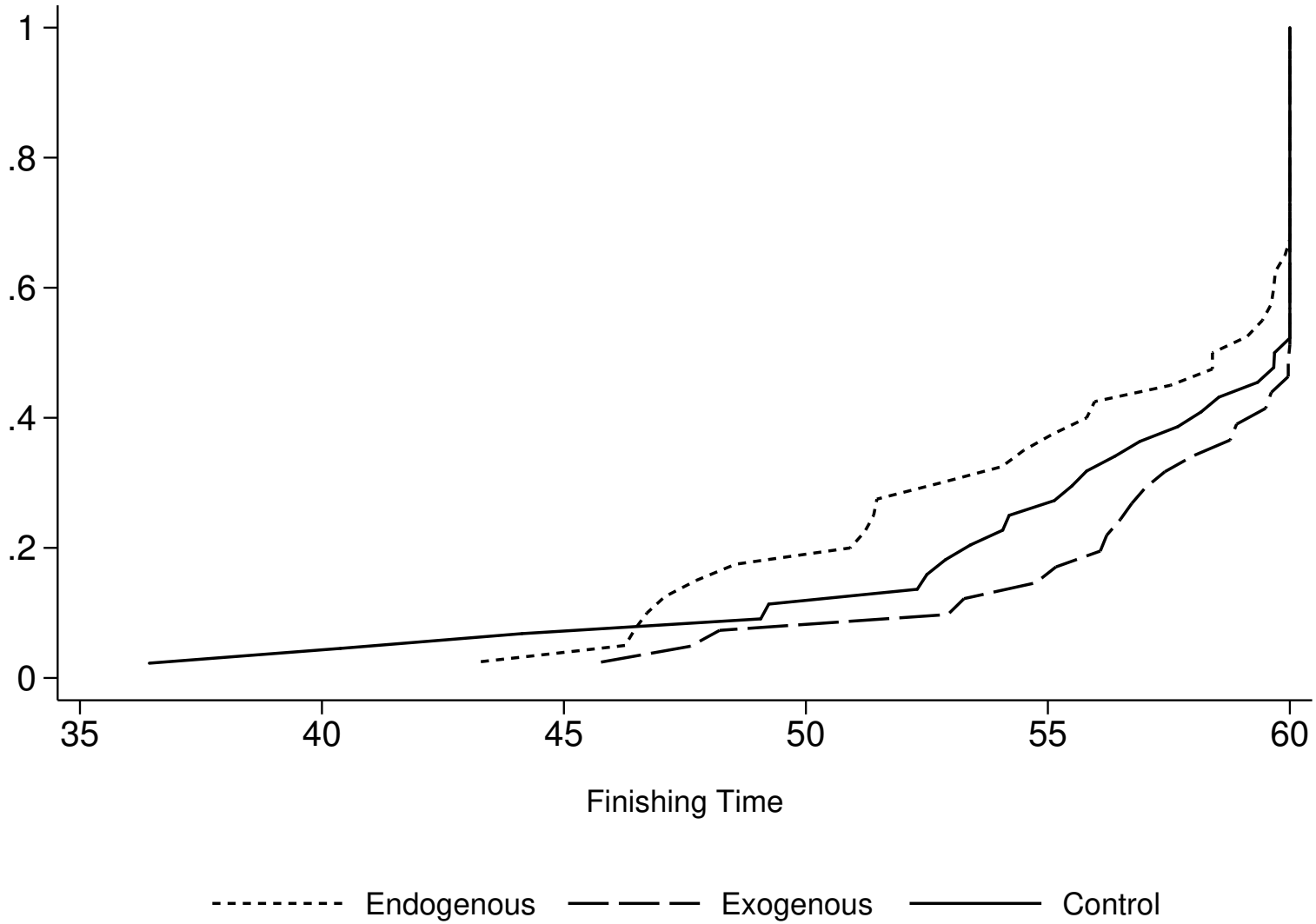
Ongoing work

Endogenous vs. exogenous leadership



A glimpse at interim results

Endogenous vs. exogenous leadership



Conclusion

Motivating performance in **non-routine analytical team tasks** is understudied

Today, a glimpse at 3 papers on the topic:

- Effects of bonus incentives
  - Bonus incentives increase performance
  - Self selected teams do not explore less; confronted teams do ...
  - Likely channel: endogenous emergence of leadership

Conclusion

Motivating performance in **non-routine analytical team tasks** is understudied

Today, a glimpse at 3 papers on the topic:

- Effects of bonus incentives
  - Bonus incentives increase performance
  - Self selected teams do not explore less; confronted teams do ...
  - Likely channel: endogenous emergence of leadership
- Effects of leadership
  - Nudging leadership increases performance
  - No signs of decreased exploration

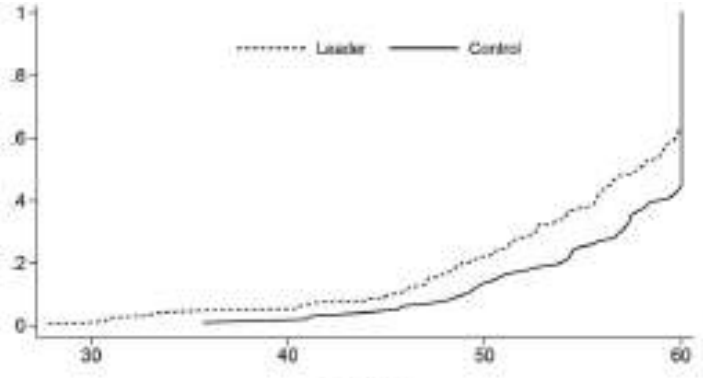
Conclusion

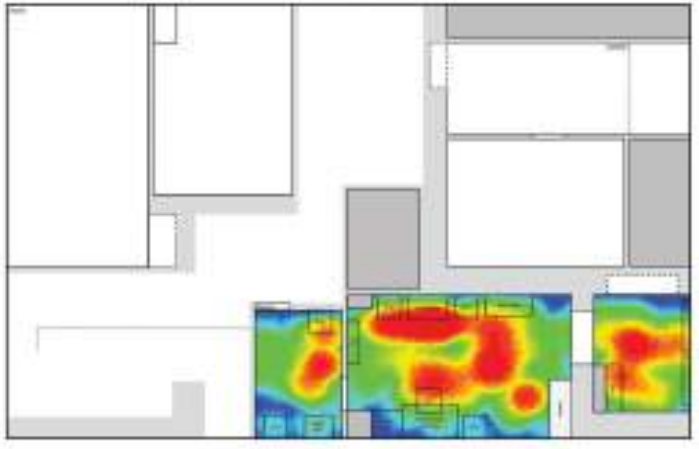
Motivating performance in **non-routine analytical team tasks** is understudied

Today, a glimpse at 3 papers on the topic:

- Effects of bonus incentives
  - Bonus incentives increase performance
  - Self selected teams do not explore less; confronted teams do ...
  - Likely channel: endogenous emergence of leadership
- Effects of leadership
  - Nudging leadership increases performance
  - No signs of decreased exploration
- Why and how does leadership work
  - ongoing work
  - movement and communication patterns are tracked
  - opening of the black box

Thank you





Thank you!

florian.englmaier@econ.lmu.de

Literature

Work that (at least partially) relates to non-routine team tasks is scarce

Laboratory experiments on creativity

- Ramm et al. (2013); Bradler et al. (2014); Charness and Grieco (2014); Laske and Schroeder (2016); Erat and Gneezy (2016).

Field experiments on teacher performance

- Muralidharan and Sundararaman (2011); Fryer et al. (2012)

Field experiment on creation of ideas

- Gibbs et al. (2014)

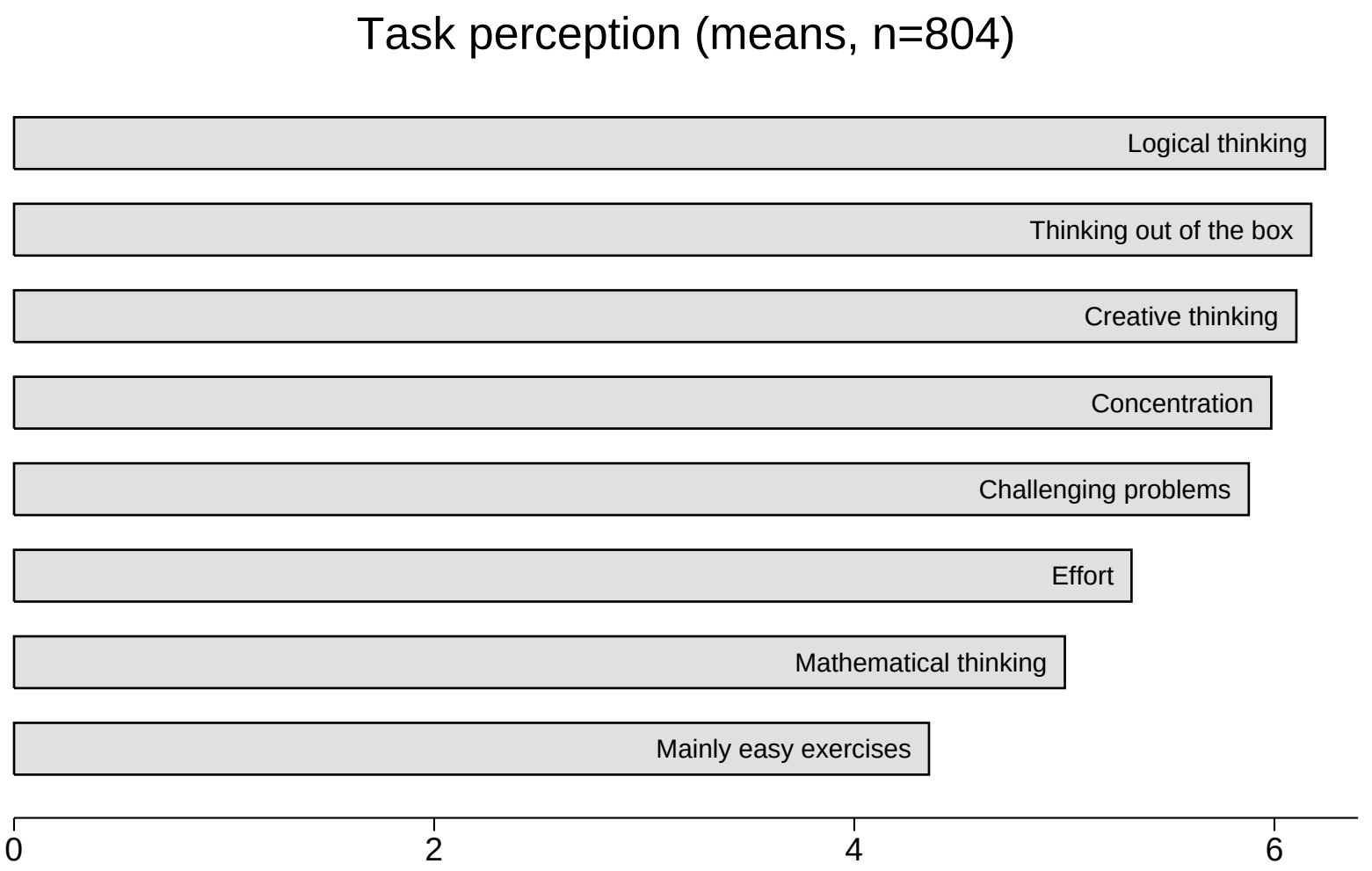
back

Real-life escape games

Source: http://boredinvancouver.com/listing/escape-game-room-experience-vancouver/

The Setting

Perception of task (measured on 7-point Likert scale)



Back

Main Result

Randomization check - Field

	Control (n=238)	*Bonus (45mins) (n=249)*
Share males	.52 (.29) [0,1]	.51 (.29) [0,1]
Group size	4.52 (1.18) [2,7]	4.71 (1.05) [2,8]
Experience	.48 (.50) [0,1]	.48 (.50) [0,1]
Private	.69 (.46) [0,1]	.63 (.48) [0,1]
English speaking	.12 (.32) [0,1]	.08 (.28) [0,1]
Age category $\in \{$18-25;26-35;36-50;51+$\}$	$\{0.29;0.45;0.21;0.05\}$	$\{0.18;0.42;0.33;0.07\}^{***}$

All variables except age category on group level. Standard deviations and minimum and maximum values for group variables in parentheses; (std.err.)[min, max]. Age category displays fractions of participants in the respective age category on the individual level. Stars indicate significant differences to *Control* (using χ^2-tests (for frequencies) and Mann-Whitney tests (for distributions), with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back

Bonus Incentives: Test Statistics

Table: Task performance for main treatments

	Control	*Bonus45*
Fraction of teams solving task in 45 min	0.10	0.26***
Fraction of teams solving task in 60 min	0.67	0.77**
Mean remaining time (in sec)	345	530***
Mean remaining time (in sec) if solved	515	688***

Notes: This table summarizes key variables and their differences across treatment *Control* and the pooled bonus incentive treatment (*Bonus45*). Stars indicate significant differences from *Control* (using χ^2 tests for frequencies and Mann–Whitney tests for distributions), and *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

▸ Back

Bonus Incentives: Further results

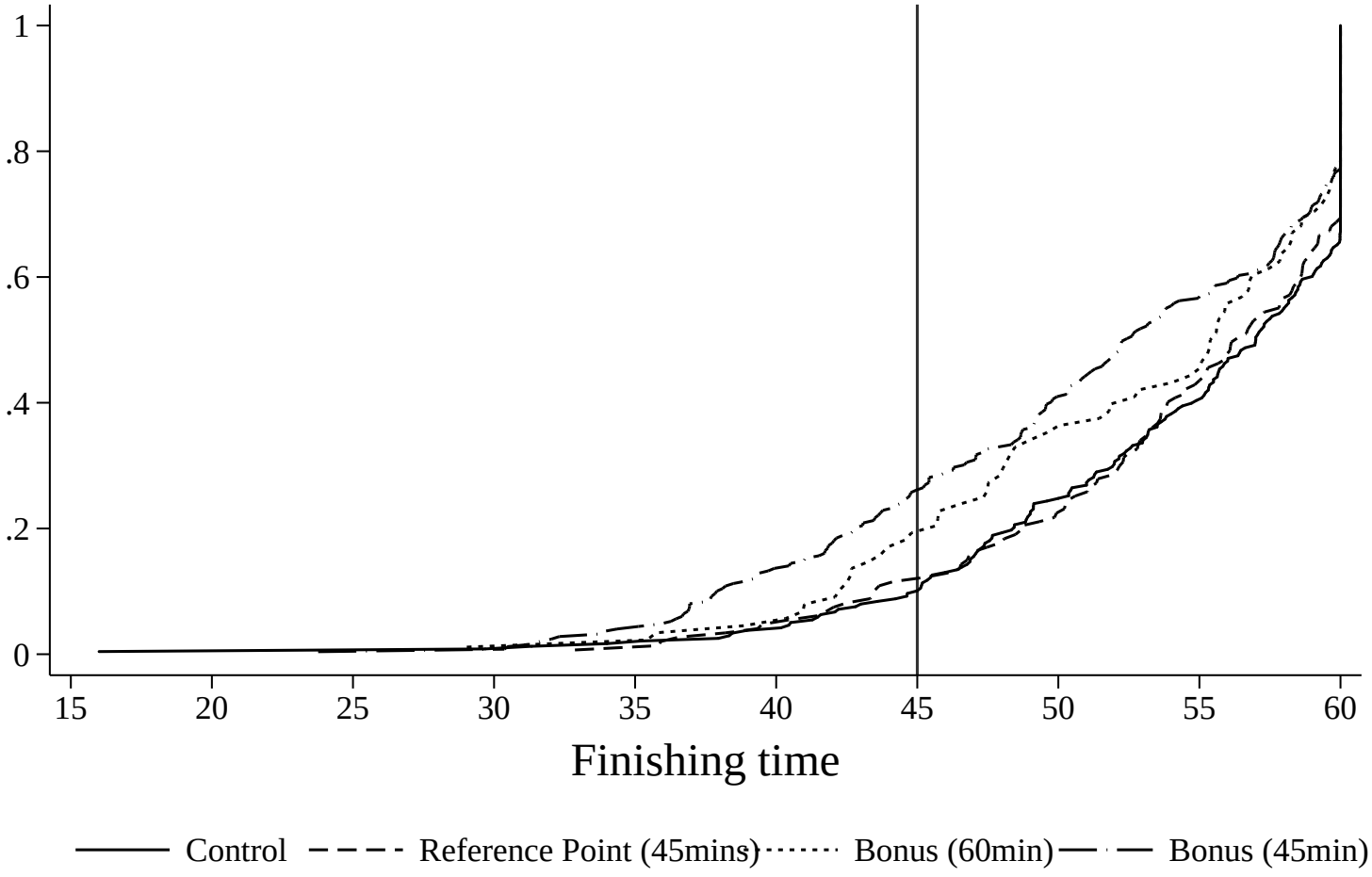
Reference points (thresholds) vs. money

From Jan to May ’17 we additionally ran treatments to study whether bonuses work due to thresholds (45mins) or money (50 euros)

- *Reference Point* (n=147): ”In order for you to judge what constitutes a good performance in terms of remaining time: If you make it in 45 minutes or less, this is a very good result!”
- *Incentive60* (n=88): Bonus (again framed as gain or loss) for solving task in 60mins

Bonus Incentives: Further results

Reference points (thresholds) vs. money



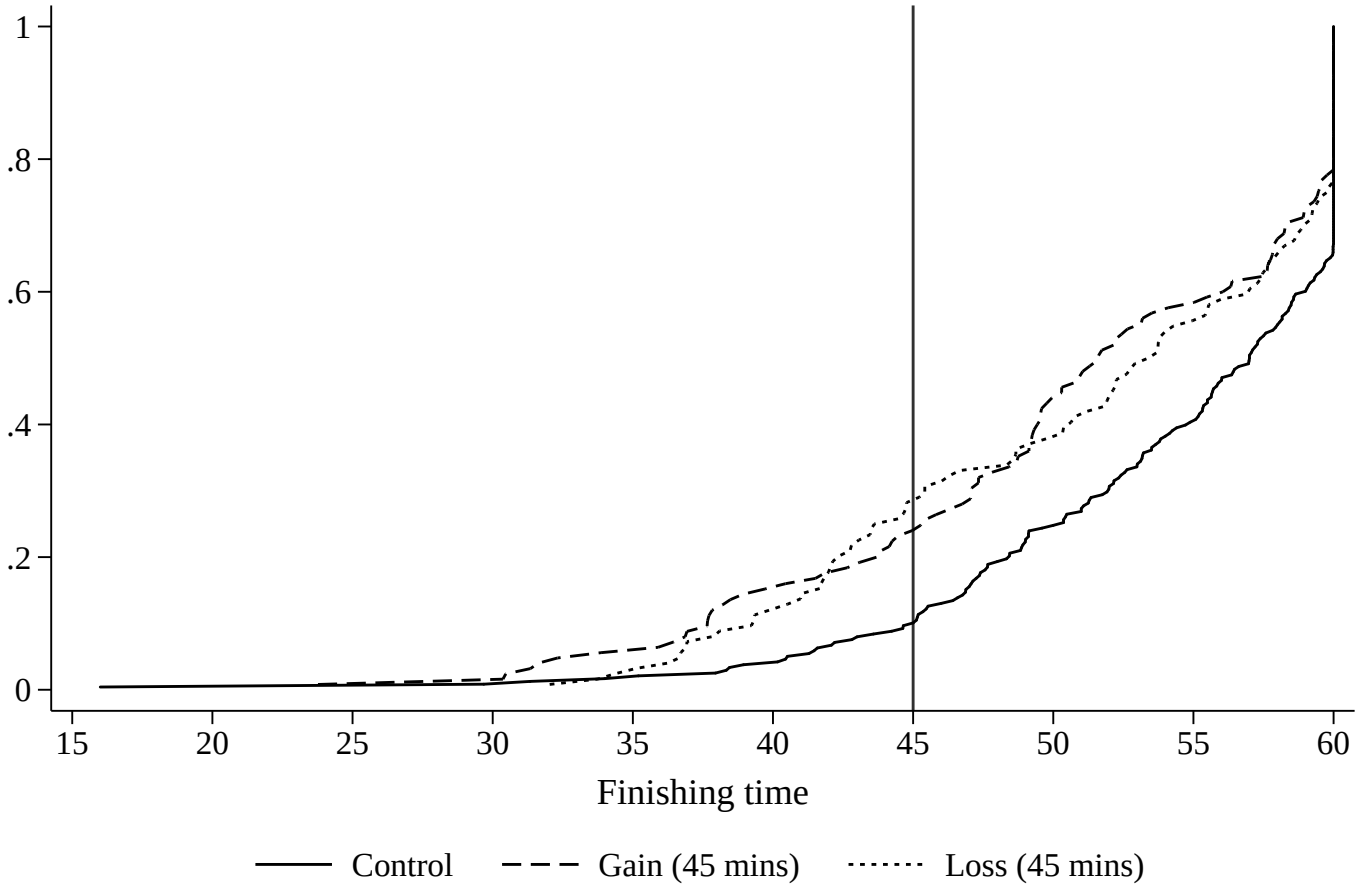
▸ Solved in 45mins regression

▸ Remaining time regression

Bonus Incentives: Further Results

Framing: Gains vs. Losses

Framing the bonus as a gain or a loss does not matter here.



▸ Test statistics

Framing: Gains vs. Losses

Main outcome variables

	Control	*Bonus45 (pooled)*	*Gain45*	*Loss45*
fraction of teams solving task in 45 mins	.10	.26***	.24***	.28***
fraction of teams solving task in 60 mins	.67	.77**	.78**	.77*
mean remaining time (in sec)	345	530***	548***	512***
mean remaining time (in sec) if solved	515	688***	707***	669***

This table summarizes key variables and their differences across our three treatments *Control*, *Gain45*, and *Loss45* and the pooled bonus incentive treatments. Stars indicate significant differences from *Control* (using Fisher's exact test for frequencies and Mann-Whitney tests for distributions), with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back ▸ BackGL

Main Result

Regression analyses: Remaining time in seconds

	GLM: Remaining time in seconds			
	(1)	(2)	(3)	(4)
Bonus incentives (45 minutes)	0.432***	0.447***	0.406***	0.257
	(0.100)	(0.098)	(0.109)	(0.174)
Constant	5.842***	4.041***	4.251***	3.803***
	(0.079)	(0.355)	(0.404)	(0.482)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	487	487	487	487

Coefficients from a generalized linear model regression with a log link of the remaining time on our treatment indicator (with *Control* as base category). Control variables added from column (2) onwards include group size (number of team members), share of males in a team, a dummy whether someone in the team has been to an Escape Game before, median age of the team, a dummy whether all group members speak German and a dummy for private teams (opposed to company team building events). Staff fixed effects control for the employees of *ExitTheRoom* present onsite. Robust standard errors reported in parentheses, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back

Main Result

Proportional hazard model

	Cox Proportional Hazard Model: Finishing the Game					
	First 45 minutes (1)-(3)			Last 15 minutes (4)-(6)		
	(1)	(2)	(3)	(4)	(5)	(6)
Bonus45 (pooled)	2.853***	2.947***	2.914***	1.178	1.250*	0.841
	(0.680)	(0.718)	(1.371)	(0.145)	(0.165)	(0.214)
χ^2 prop. haz. assumption	0.11	8.56	44.59	0.04	7.94	42.23
Degrees of freedom	1	9	45	1	9	46
Control Variables	No	Yes	Yes	No	Yes	Yes
Staff Fixed Effects	No	No	Yes	No	No	Yes
Week Fixed Effects	No	No	Yes	No	No	Yes
Observations	487	487	487	487	487	487

Hazard ratios from a Cox proportional hazard regression of time elapsed until a team has completed the task on our treatment indicator *Bonus45*. Control variables, as before. Robust standard errors reported in parentheses, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

back

Results: All treatments

Regression analyses: Game solved in less than 45 minutes

Probit: Solved in less than 45 minutes	(1)	(2)	(3)	(4)
Bonus (pooled, 45min)	0.661***	0.708***	0.746***	0.731***
	(0.141)	(0.150)	(0.162)	(0.163)
Bonus (pooled, 60min)	0.435**	0.458**	0.475**	0.481**
	(0.190)	(0.199)	(0.211)	(0.212)
Reference Point (45min)	0.104	0.102	0.051	0.107
	(0.176)	(0.183)	(0.203)	(0.204)
Constant	-1.301***	-3.242***	-3.080***	531.187
	(0.112)	(0.468)	(0.546)	(458.493)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	722	722	722	722

Coefficients from Probit regressions of whether a team solved the task within 45 minutes on our treatment indicators *Bonus (pooled, 45min)*, *Bonus (pooled, 60min)* and *Reference Point (45min)* with *Control* being the base category). Control variables added from column (2) onwards include group size (number of team members), share of males in a team, a dummy whether someone in the team has been to an Escape Game before, dummies for median age category of the team, a dummy whether all group members speak German and a dummy for private teams (opposed to company team building events). Staff fixed effects in column (3) and (4) control for the employees of *ExitTheRoom* present onsite. Column (4) includes week fixed effects (all models include the full sample, including weeks that perfectly predict failure to receive the bonus). Robust standard errors reported in parentheses, and $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Back

Results: All treatments

Regression analyses: Remaining time

GLM: Remaining time	(1)	(2)	(3)	(4)
Bonus (pooled, 45min)	0.432***	0.436***	0.376***	0.244
	(0.100)	(0.097)	(0.106)	(0.150)
Bonus (pooled, 60min)	0.233*	0.267**	0.392***	0.449**
	(0.135)	(0.120)	(0.127)	(0.185)
Reference Point (45min)	0.002	-0.001	0.102	0.131
	(0.123)	(0.118)	(0.128)	(0.149)
Constant	5.842***	4.044***	4.225***	3.713***
	(0.079)	(0.296)	(0.342)	(0.417)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	722	722	722	722

Coefficients from a generalized linear model regression with a log link of the remaining time on our treatment indicators (with *Control* being the base category). Control variables added from column (2) onwards include group size (number of team members), share of males in a team, a dummy whether someone in the team has been to an Escape Game before, dummies for median age category of the team, a dummy whether all group members speak German and a dummy for private teams (opposed to company team building events). Staff fixed effects in column (3) and (4) control for the employees of *ExitTheRoom* present onsite. Column (4) includes week fixed effects. Robust standard errors reported in parentheses, and $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back

Further results

Randomization Check - Student Sample

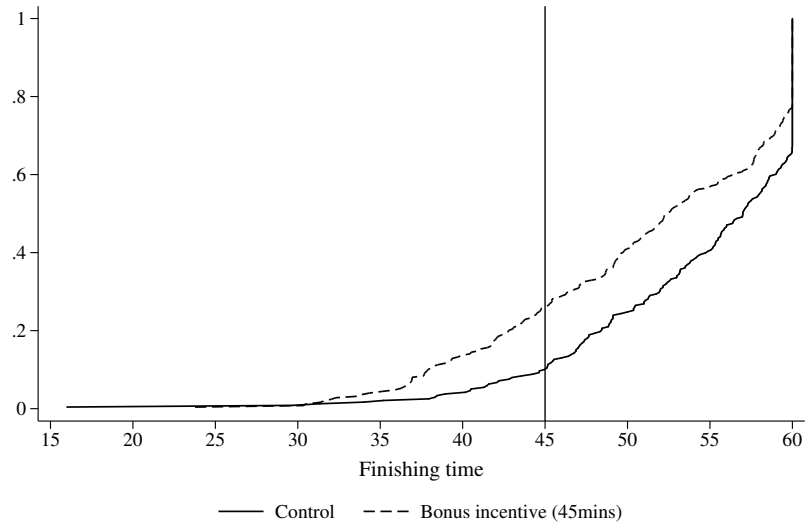
	average (a)	min	max	a in Control	a in Gain	a in Loss	Diff
Observations	268	-	-	88	90	90	-
Age	23.6	18.3	37	23.2	22.9	23	-
Share male(s)	0.44	0	1	0.46	0.39	0.47	GL*
Intrinsic motivation	6.8	2	14.3	6.75	6.70	6.93	-
Loss aversion	1.68	0.91	2.75	1.70	1.68	1.65	-
Risk taking	3.92	1.66	5.66	3.89	3.94	3.96	-
Total creativity points	222	102	487	226	216	223	-
Experience ETR	0.05	0	0.67	0.05	0.04	0.05	-
Experience EG	0.14	0	1	0.14	0.11	0.17	GL*
School GPA	1.97	1.13	2.90	2.01	2.00	1.92	-
Political Attitude	3.49	2	6	3.50	3.45	3.53	-
Available Income	425	167	1833	407	407	461	CL*, GL*

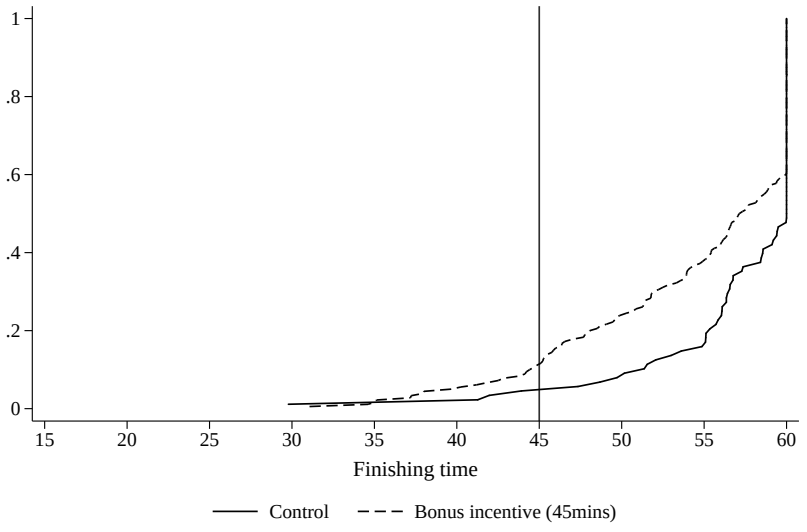
Stars indicate significant differences between control and gain (CG), control and loss (CL) and gain and loss (GL) using Mann-Whitney-Tests for distributions, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Back

Results (Lab & Field Subjects)

CDF - Comparison (left panel: Field participants, right panel: Lab subjects)

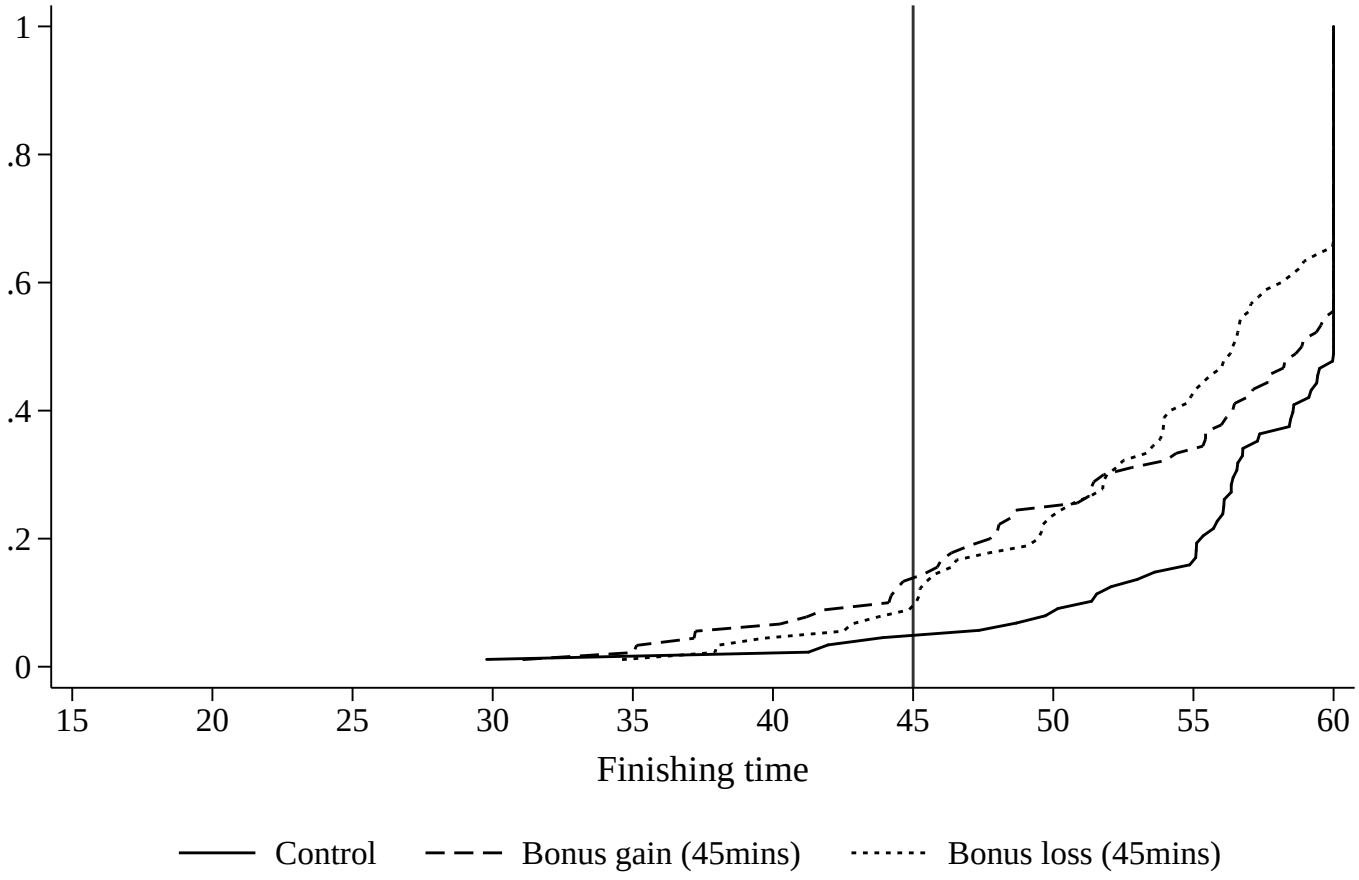




Back

Results (Student participants)

CDF of finishing time



Back

Results (Student participants)

Main outcome variables

	Control	*Bonus incentives (pooled)*	*Gain45*	*Loss45*
fraction of teams solving task in 45 mins	0.05	0.11^{*}	0.13^{**}	0.09
fraction of teams solving task in 60 mins	0.48	0.60^{*}	0.54	0.65^{**}
mean remaining time (in sec)	169.90	327.97^{***}	334.67^{*}	321.28^{***}
mean remaining time (in sec) if solved	355.98	546.62^{***}	590.10^{**}	510.50^{***}

This table summarizes key variables and their differences across our three treatments *Control*, *Gain45*, *Loss45*, and *Bonus incentives (pooled)* treatments. Stars indicate significant differences from *Control* (using χ^2-test for frequencies and Mann-Whitney tests for distributions), with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$. P-values of non-parametric comparisons between *Gain45* and *Loss45* are larger than 0.10 for all four performance measures.

▸ Back

Results (Student participants)

Regression analyses (pooled): Remaining time

GLM: Remaining time	(1)	(2)	(3)	(4)
Bonus incentives (45 minutes)	0.894*	0.877*	0.830	0.981*
	(0.533)	(0.532)	(0.542)	(0.550)
Constant	-3.091***	-3.258	-2.584	-18.737***
	(0.489)	(2.006)	(2.336)	(2.300)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	268	268	268	268

Coefficients from a generalized linear model regression with a log link of the remaining time on our treatment indicators (with *Control* being the base category).Control variables added from column (2) onwards include share of males in a team, a dummy whether someone in the team has been to an Escape Game before and average age of the team. Staff fixed effects control for the employees of *ExitTheRoom* present onsite. Robust standard errors reported in parentheses, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back

Results (Student participants)

Regression analyses (OLS): Number of hints taken (in customer and student sample)

	OLS: Number of hints requested							
	Field experiment (1)-(4)				Framed Field Experiment (5)-(8)			
	within 60 minutes		within 45 minutes		within 60 minutes		within 45 minutes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bonus45 (pooled)	0.172	0.098	0.387***	0.172	0.372***	0.361***	0.843***	0.817***
	(0.132)	(0.221)	(0.107)	(0.192)	(0.133)	(0.134)	(0.126)	(0.124)
Constant	2.924***	0.506	1.971***	-0.596	3.739***	5.449***	2.330***	3.698***
	(0.100)	(1.228)	(0.079)	(1.080)	(0.111)	(1.032)	(0.099)	(1.027)
Control Variables	No	Yes	No	Yes	No	Yes	No	Yes
Staff Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Week Fixed Effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	487	487	487	487	268	268	268	268

Coefficients from OLS regressions of the number of hints requested within 60 minutes or 45 minutes regressed on our treatment indicator *Bonus45*. Controls and fixed effects identical to previous tables. Robust standard errors reported in parentheses, and $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

▸ Back

Main Result

Regression analyses (all treatments): Game solved in less than 45 minutes

	Probit: Solved in less than 45 minutes			
	(1)	(2)	(3)	(4)
Bonus45 (pooled)	0.661***	0.708***	0.746***	0.512**
	(0.141)	(0.150)	(0.162)	(0.224)
Bonus60 (pooled)	0.435**	0.458**	0.475**	0.602**
	(0.190)	(0.199)	(0.211)	(0.282)
Reference Point	0.104	0.102	0.051	0.096
	(0.176)	(0.183)	(0.203)	(0.246)
Constant	-1.301***	-3.242***	-3.080***	-3.294***
	(0.112)	(0.468)	(0.546)	(0.934)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	722	722	722	722

Coefficients from Probit regressions of whether a team solved the task within 45 minutes on our treatment indicators *Bonus45*, *Bonus60* and *Reference Point* with *Control* being the base category. Control variables added from column (2) onwards include group size (number of team members), share of males in a team, a dummy whether someone in the team has been to an Escape Game before, median age of the team, a dummy whether all group members speak German and a dummy for private teams (opposed to company team building events). Robust standard errors reported in parentheses, and $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Back

Results (Student participants)

Regression analyses (pooled): Solved in less than 45 minutes

Probit: Solved in less than 45 minutes	(1)	(2)	(3)	(4)
Bonus incentives (45 minutes)	0.470*	0.463*	0.516*	0.605**
	(0.264)	(0.266)	(0.272)	(0.299)
Constant	-1.691***	-1.766	-1.540	-6.160***
	(0.233)	(1.094)	(1.305)	(1.342)
Control Variables	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Week Fixed Effects	No	No	No	Yes
Observations	268	268	268	268

Coefficients from Probit regressions of whether a team solved the game within 45 minutes on our treatment indicator (with *Control* as base category). Control variables added from column (2) onwards include share of males in a team, a dummy whether someone in the team has been to an Escape Game before and average age of the team. Staff fixed effects control for the employees of *ExitTheRoom* present onsite. Robust standard errors reported in parentheses, and $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Back

Questionnaire: Team organization and perceptions

First set of questions ($n = 804$)

	Control	*Incentives*	*p-val*
"The team was very stressed."	3.57	4.13***	<0.01
"One person was dominant in leading the team."	2.60	2.86**	0.03
"We wrote down all numbers we found."	5.64	5.50**	0.04
"I was dominant in leading the team."	2.64	2.87**	0.05
"We first searched for clues before combining them."	4.58	4.39	0.11
"We exchanged many ideas in the team."	5.87	5.74	0.12
"When we got stuck we let as many team members try as possible."	5.43	5.28	0.14
"The team was very motivated."	6.14	6.26	0.22
"We communicated a lot."	5.78	5.88	0.23
"All team members exerted effort."	6.23	6.37	0.24
"Our notes were helpful in finding the solution."	5.50	5.43	0.41
"I was able to present all my ideas to the group."	5.95	5.93	0.41
"We were well coordinated in the group."	5.73	5.80	0.61
"I was to concentrated on my own part."	2.88	2.83	0.76
"We made our decisions collectively."	5.51	5.58	0.87
"I would like to do a similar task again."	6.30	6.28	0.88
"Our individual skills complemented well."	5.65	5.68	0.89
"The mood in our team was good."	6.30	6.36	0.93
"All team members contributed equally."	5.97	6.00	0.96

Back

Questionnaire: Team organization and perceptions

Second set of questions (based on Högl and Gemünden, 2001) ($n = 375$)

	Control	*Incentives*	*p-val*
"How much did you wish somebody would take leadership?"	2.67	3.32***	<0.01
"How well lead was the team?"	3.85	4.21**	0.04
"How much did you think about the problems?"	6.00	5.79	0.11
"How much did you follow ideas ' that were not promising?"	5.02	4.79	0.17
"How much team spirit evolved?"	5.54	5.80	0.17
"How much coordination of individual" tasks and joint strategy?"	3.28	3.51	0.18
"How much exploitation of individual potential?"	5.14	4.94	0.22
"How much helping when somebody stuck?"	5.70	5.58	0.22
"How much did you search the room for solutions?"	6.31	6.22	0.51
"How much exertion of effort by all members?"	5.98	5.96	0.60
"How much communication about procedures?"	5.30	5.35	0.88
"How much accepting help of others?"	5.80	5.85	0.89

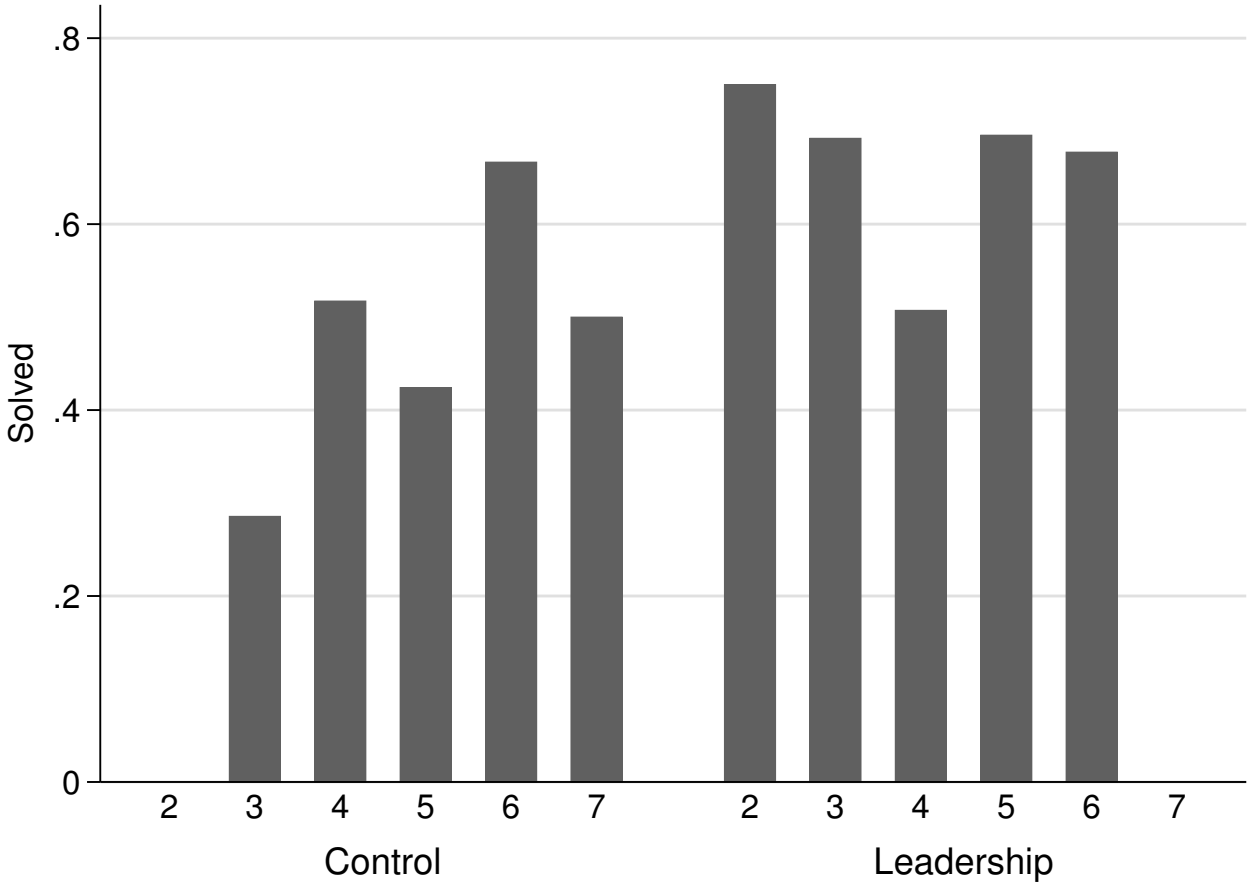
Back

Leadership: Results by Group Size

Some Descriptives of the “complying to Leadership” Subsample

- Share of Males …
  - …in teams: 49.2%
  - …as leaders: 54.3%
- Share of Females …
  - …in teams: 50.8%
  - …as leaders: 45.7%
- Average Age …
  - …of teams ca. 31.6 yrs
  - …of chosen leaders ca. 30.8 yrs

Success by group size



Observations Group Size:

n(2) = 9, n(3) = 27, n(4) = 98, n(5) = 102, n(6) = 43, n(7) = 2

Finishing time

	Tobit: Finishing time			
	(1)	(2)	(3)	(4)
Leadership	-4.545***	-4.568***	-5.257***	-3.244**
	(1.221)	(1.218)	(1.436)	(1.282)
Observations	281	281	281	281
Team Controls	No	Yes	Yes	Yes
Staff FE	No	No	Yes	Yes
Room, Day and Week FE	No	No	No	Yes

Notes: The table displays average marginal affects from Tobit regressions of finishing times on our treatment indicator (with Control as base category). Team controls (group size, share of males, experience, median age, language, private, natural leader, walkie talkie), staff, room, day and week fixed effects are step-wise included. Standard errors in parentheses are clustered on the date level, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

▸ GLM

Dynamics of Completion

	Hazard Ratios			
	First 45 min	45-50 min	50-55 min	55-60 min
	(1)	(2)	(3)	(4)
Leadership	— — —	1.543	2.821**	3.033***
	(— — —)	(1.082)	(1.411)	(0.864)
Observations	281	260	227	187
Team Controls	Yes	Yes	Yes	Yes
Staff FE	Yes	Yes	Yes	Yes
Room, Day and Week FE	Yes	Yes	Yes	Yes

Notes: The table displays hazard ratios from Cox proportional hazard regressions on our treatment indicator (with Control as base category). Team controls (group size, share of males, experience, median age, language, private, natural leader, walkie talkie), staff, room, day and week fixed effects are included. Standard errors in parentheses are clustered on the date level, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

RA Ratings (LCI and LNCI)

	Individual Search		Standing Together	
	(1)	(2)	(3)	(4)
Leadership	0.060		-0.162*	
	(0.119)		(0.081)	
Leader chosen immediately (LCI)		0.133		-0.124
		(0.127)		(0.099)
Leader not chosen immediately (LNCI)		-0.025		-0.207**
		(0.136)		(0.090)
Observations	279	279	279	279
Team Controls	Yes	Yes	Yes	Yes
Staff and RA FE	Yes	Yes	Yes	Yes
Room, Day and Week FE	Yes	Yes	Yes	Yes

Notes: The table displays coefficients from OLS regressions of individual search and standing together on our treatment indicator (with Control as base category). Team controls (group size, share of males, experience, median age, language, private, natural leader, walkie talkie), staff, RA, room, day and week fixed effects are included. Standard errors in parentheses are clustered on the date level, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

Happiness of the teams

We have upon the end a *customer satisfaction survey* where we in particular ask for the experience regarding *satisfaction* and *value-for-money*

	Satisfaction		Value for money	
	All	if solved	All	if solved
	(1)	(2)	(3)	(4)
Leadership	0.084	-0.125	0.377***	0.268
	(0.070)	(0.074)	(0.133)	(0.178)
Observations	279	157	278	156
Team Controls	Yes	Yes	Yes	Yes
Staff FE	Yes	Yes	Yes	Yes
Room, Day and Week FE	Yes	Yes	Yes	Yes

Notes: The table displays coefficients from OLS regressions of customer satisfaction and price effort evaluations on our treatment indicator (with Control as base category). Team controls (group size, share of males, experience, median age, language, private, natural leader, walkie talkie), staff, room, day and week fixed effects are included. Standard errors in parentheses are clustered on the date level, with * = $p < 0.10$, ** = $p < 0.05$ and *** = $p < 0.01$.

▸ Perception of Motivation/Coordination

Originality

Number of hints

- Teams can use up to 5 hints for solving the task
- Number of hints (inversely) related to finding original solutions
- Average number of hints taken does not differ statistically significantly across conditions

	Control Mean (SD)	Motivation Mean (SD)	Coordination Mean (SD)
Hints	3.42	3.09	3.35
	(1.40)	(1.56)	(1.42)
Hints (Solved)	2.90	2.54	2.87
	(1.45)	(1.54)	(1.41)
Observations	95	95	91

▸ Timing of Hints

Tournaments: Motivation

- Prominent feature of our task: **recombination of ideas**
- Probably leading model for understanding **innovation**
- Very often, innovation is motivated via tournaments (see e.g. Dan Gross' work)

Tournaments: Design & Treatments

Sample

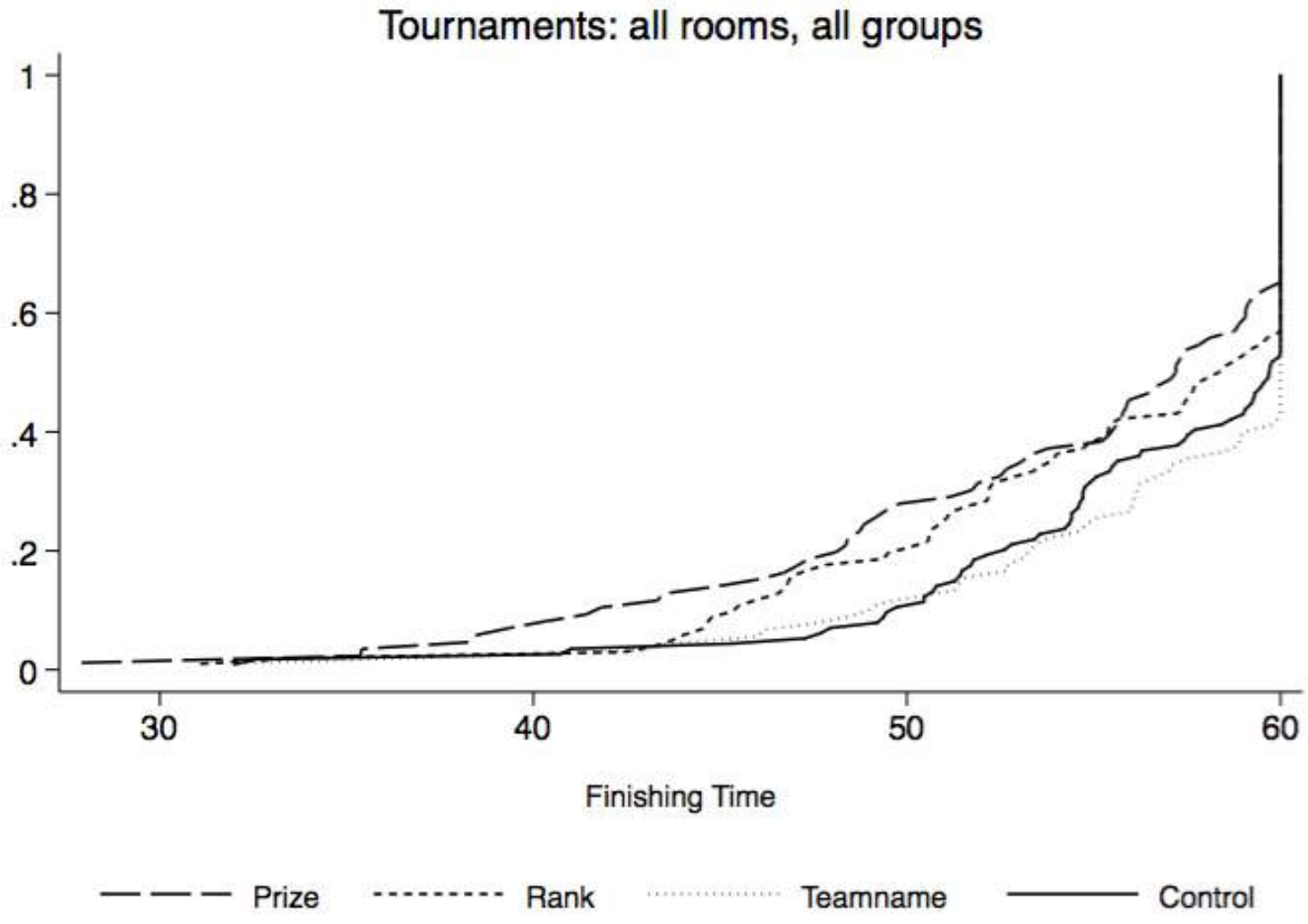
- 1831 Regular ETR customers (398 groups)
- Data collection from April 2018 to July 2018
- Randomization on weekly basis

Treatments

- *Control*: no intervention (114 groups)
- *Teamname*: Teamname (90 groups)
- *Rank*: Teamname + Ranking on Facebook-Fanpage (102 groups)
- *Prize*: Teamname + Ranking on Facebook-Fanpage + Prize (150 Euro) (92 groups)

Tournaments: Main Results

CDFs of finishing times



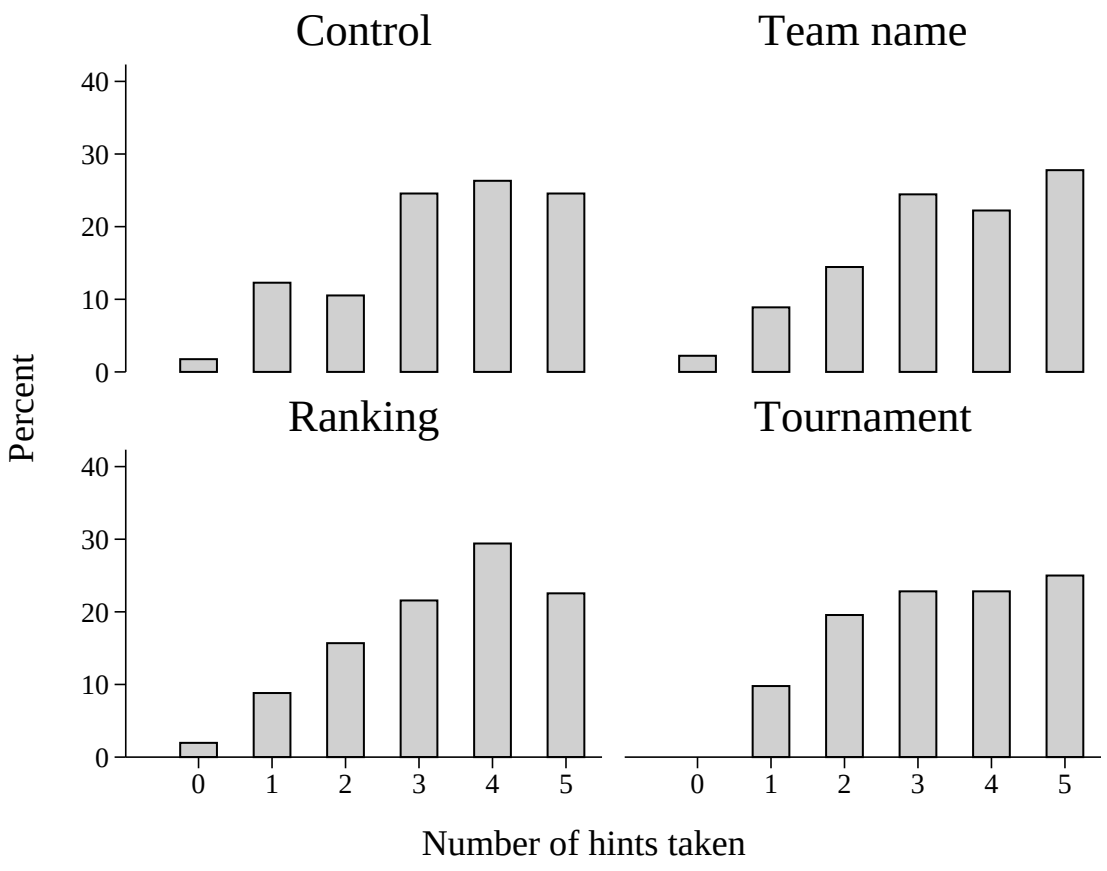
Tournaments: Main Results

Regression

	(1) Solved	(2) Solved	(3) Solved	(4) Solved
Team name	-0.262***	-0.278**	0.0595	0.190
	(0.0890)	(0.115)	(0.235)	(0.341)
Rank	0.0820	0.0604	0.255*	0.291
	(0.0847)	(0.0904)	(0.151)	(0.233)
Prize	0.210*	0.228**	0.355**	0.388**
	(0.108)	(0.0965)	(0.168)	(0.175)
N	398	398	398	398
Controls	No	Yes	Yes	Yes
Staff Fixed Effects	No	No	Yes	Yes
Room Fixed Effects	No	No	No	Yes

Coefficients from Probit regressions of whether a team solved the game on our treatment indicator (with *Control* as base category). Robust standard errors clustered on the week level reported in parentheses, with $^{*} = p < .10$, $^{**} = p < .05$ and $^{***} = p < .01$.

Tournaments: Originality



- Hint taking does not differ significantly across conditions.
- Tentative: Top-performing teams in tournament use **much** more hints!