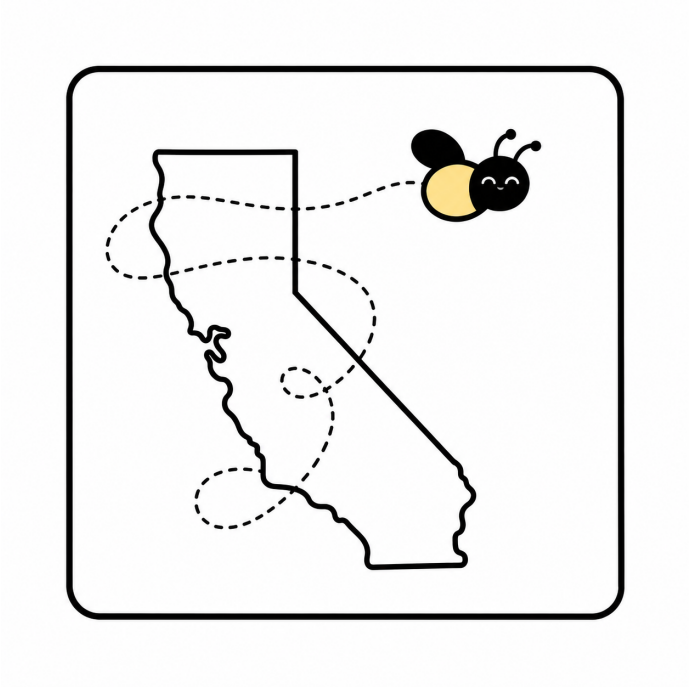
Behavioral and Experimental Economics
Stanford-Caltech-UC Student Conference
BEESCUitS 2026



Hosted at Caltech on November 20th and 21st.

Questions? Email Aaron Berman at aberman@caltech.edu.

Contents

1 Acknowledgments 2

2 Conference Schedule 3

2.1 Friday, November 20th 3

2.2 Saturday, November 21st 3

3 Session Format 4

4 Caltech and Pasadena 5

4.1 Venue and Location 5

4.2 Some Parking Options 5

4.3 A Thing or two to do 5

5 Presentation Abstracts 6

6 Attendees 13

1. Acknowledgments

We would like to thank The Linde Institute and The Center for Theoretical and Experimental Social Sciences for supporting BEESCUitS and making our mission of building a local behavioral and experimental economics graduate student community possible.

We are extremely grateful to Charles Sprenger, Kapaūhi Stibbard, Katie Fisher, Kirby Nielsen, Letty Diaz, and Marina Agranov for helping bring BEESCUitS 2026 together.

Aaron would also like to extend a special thank you to Camila Farres, Erik Scheuermann, and Francesco Crispano for generously helping to make the day run smoothly, and to Stefan Hansel for our group photo.

2. Conference Schedule

All sessions will be held in Broad 100. Lunches are organized outside of the Chen Neuroscience Building. Figure 1 provides a map marking these locations.

2.1. Friday, November 20th

Time	Activity
8:15 AM	Breakfast
8:55–9:00 AM	Welcome!
9:00–9:30 AM	Aaron Berman, "Accounting for Noise in First and Second Differences."
9:30–10:00 AM	Zander Memon, "Do Bipartisan Bills Lead to Higher Stimulus: Evidence From a Randomized Control Trial."
10:00–10:30 AM	Introductions
10:30–10:45 AM	*Break*
10:45–11:15 AM	Camila Farres, "Anticipating personalization."
11:15–11:45 AM	Zoë Arnaut, "Words versus numbers: Eliciting health and environmental risk beliefs on surveys."
11:45 AM–12:15 PM	Xinlu Zhao, "Who's Really Talking? Candidate-Side AI Assistance and Recruiter Judgments."
12:15–1:45 PM	*Lunch and a group photo!*
1:45–2:15 PM	Anastasiia Morozova, "The Lucky Leading the Blind: Advice and Imitation in Financial Markets."
2:15–2:45 PM	Austin Booksby, "Learning with competing models."
2:45–3:00 PM	*Break*
3:00–3:30 PM	Alisher Batmanov, "Fragile Learning From Others."
3:30–4:00 PM	Dexter Hsu, "Political Correctness and Political Polarization."
4:30 PM	*Social outing and dinner*

2.2. Saturday, November 21st

Time	Activity
8:15 AM	Breakfast
9:00–9:30 AM	Jindi Huang, "The Comparative Blind Spot: Misconceptions About Comparative Advantage."
9:30–10:00 AM	Yu-Hsiang (Sean) Wang, "Strategic Complexity in Decomposed Games."
10:00–10:30 AM	Bridget Galaty, "Incentive Structures and Strategic Uncertainty in Team Punishment."
10:30–10:45 AM	*Break*
10:45–11:15 AM	Vidhi Shah, "Social Networks as a Source of Risk: Evidence from Past Banking Crises."
11:15–11:45 AM	Milena Jessen, "Hierarchical Representations."
11:45 AM–12:15 PM	Clément Herman, "Small datasets as mental models."
12:15–12:20 PM	Closing and announcements
12:20 PM	*Lunch*

3. Session Format

All conference sessions will take place in Broad 100. A presentation clicker will be provided and we will use one central laptop to minimize transition time.

Each presenter is given 30 minutes, and you are welcome to structure that time however you wish. You may use the full 30 minutes for your presentation, present for 20 minutes followed by 10 minutes of questions and discussion, or take questions throughout. We encourage you to choose whichever format works best for your presentation. Before you present, Aaron will ask how you want to field questions.

To keep the conference on schedule, we will diligently enforce the 30-minute format. Please plan accordingly.

4. Caltech and Pasadena

4.1. Venue and Location

- All sessions are located in Broad 100, which is in the Broad Center, Building 96, circled in orange.
- Lunch on Friday and Saturday will be held outside of the Chen Neuroscience Research Building, Building 10, boxed in red.
- Our Friday social gathering and dinner is located in the Central Catalina Recreation Room, Building 111, circled in purple.



Figure 1: Map of Caltech.

4.2. Some Parking Options

- The closest day-pass paid parking is available in the Wilson North Parking Structure, Building 123, and in the Wilson South Parking Structure, Building 124.
- There is also some street parking on Wilson Avenue.

4.3. A Thing or two to do

We have a half-day on Saturday giving you some time to explore the city. I like the Norton Simon Museum and The Huntington Library, Art Museum, and Botanical Gardens. Also, we have several movie theaters on Colorado Boulevard, and the new Hunger Games movie releases on Friday!

5. Presentation Abstracts

We accept work at any stage. The abstracts range from project descriptions, experimental designs and pilot data to job-market papers. The abstracts are listed in the schedule's presentation order.

Aaron Berman, Caltech, "Accounting for Noise in First and Second Differences."

A central objective of empirical economics is to identify and understand the effects of treatments, interventions, and changes in economic environments on behavior. Beyond establishing whether an effect exists, researchers frequently seek to understand the conditions under which it persists, strengthens, or weakens. For example, does the "endowment effect" (Knetsch, 1989), characterized by a reluctance to exchange goods one owns, persist among individuals with substantial market experience? Interpreting such comparisons is complicated when individuals make decisions stochastically: observed differences in behavior may reflect differences in underlying preferences, variation in stochastic choice, or both.

This project investigates what researchers can learn about the existence and stability of underlying preference differences from stochastic choice data. We focus on second differences, which compare differences in population preference proportions between two decision problems across two contexts. We will characterize what can be inferred under alternative assumptions about stochastic choice and develop statistical tests accounting for choice noise and sampling uncertainty. We will then reanalyze evidence of the Allais paradoxes and exchange asymmetries, comparing conventional and noise-adjusted conclusions. Finally, we will investigate how richer choice data and new experimental designs can aid identification and inference.

Zander Memon, UCD, "Do Bipartisan Bills Lead to Higher Stimulus: Evidence From a Randomized Control Trial."

Rising political polarization in the United States has shaped the passage or failure of legislation across a wide range of goals. Polarization is expected in certain venues, but what about fiscal policies? Using bipartisanship as a proxy for political partisanship, this paper investigates how the degree of bipartisanship of a hypothetical fiscal policy shapes household's perceptions of and response to the fiscal stimulus. Using a randomized control trial, I provide all respondents with information about a hypothetical direct transfer – either a stimulus check or a tax rebate. I vary only the hypothetical associated partisanship, coding each bill as either bipartisan or contentious along party lines. To understand the implications of this information on the stimulus effect, I gather respondent's hypothetical, reported marginal propensities to consume (MPCs). I show that the degree of partisanship impacts beliefs about the fiscal policy and that these beliefs drive household level economic behavior. I provide evidence of mechanisms that underpin these belief and behavior changes, namely fiscal rules and concerns about reversal.

Camila Farres, Caltech, "Anticipating personalization."

Institutions often personalize using consumers' revealed preferences. This paper studies how consumers change their behavior when they anticipate personalization. I use an experiment that allows me to induce the consumer's information, disclose all consequences of sharing it for personalization, and exogenously manipulate whether personalization is incentive compatible. I find that personalization changes the way people share their information, but not in the way theory predicts; in particular, many share information accurately when they should not. Using a simple task, I can identify the misconceptions that drive suboptimal

behavior. Understanding how people respond to anticipated personalization is important for designing it effectively and evaluating consumer welfare.

Zoë Arnaut, UCSD, "Words versus numbers: Eliciting health and environmental risk beliefs on surveys." Joint work with Richard Carson and Dale Whittington.

Social scientists have long utilized individual expectations, or probabilistic beliefs, data to better understand how and why people make choices. But, what happens when we elicit these expectations using words, numbers, or both? We conduct three waves of a survey experiment with over 20,000 total participants across nine countries and two contexts and find that, for both health and environmental issues, on average, a "words only" treatment led to about a 9% increase in reported probabilities, as compared to the "words and numbers" treatment, and the "numbers only" treatment led to a decrease of 7% compared to "words and numbers."

Xinlu Zhao, UCSB, "Who's Really Talking? Candidate-Side AI Assistance and Recruiter Judgments."

As real-time AI assistance becomes increasingly accessible to job candidates, it may fundamentally change both how candidates perform in interviews and how employers evaluate their abilities. While AI tools may help candidates deliver more polished and persuasive responses, they may also alter the information recruiters obtain about candidates' underlying skills. This paper investigates how candidate-side AI assistance affects interview performance, recruiter evaluations, and the accuracy of employer beliefs.

We present preliminary evidence from a randomized pilot experiment in which job candidates were assigned to interview with or without access to a real-time AI copilot. Before the interview, all candidates completed a work-related task without AI assistance, providing an imperfect benchmark of job-relevant ability. Recruiters first reviewed candidates' résumés and predicted their task performance. After the interview, recruiters evaluated candidates' interview performance and updated their predictions based on combined résumé and interview information. This design allows us to distinguish changes in perceived interview performance from changes in the informational value of interviews.

Among 24 participants who completed the pilot, candidates assigned access to AI received substantially higher interview ratings. Controlling for GPA, the estimated difference was 1.44 points on a five-point scale (95% CI: 0.58–2.31). Recruiters also revised their predictions more favorably for AI-assisted candidates. However, differences in improvements in prediction accuracy were imprecisely estimated, leaving open whether AI assistance improves or reduces recruiters' ability to assess candidates' underlying skills. Selective completion, baseline imbalances, and incomplete assignment records limit the causal interpretation of these preliminary findings.

Building on these results, we propose a larger experiment involving 200 independently randomized candidates and 20 recruiters to examine how AI assistance affects candidate performance, employer belief updating, and screening accuracy. A supplementary design will investigate whether AI assistance generates spillover effects on the evaluation of competing candidates.

This research contributes to the literature on labor market screening, employer learning, and the economic consequences of AI by examining whether AI-assisted communication improves candidates' observed performance without necessarily improving the information employers obtain about their abilities.

Anastasiia Morozova, UCSB, "The Lucky Leading the Blind: Advice and Imitation in Financial Markets." Joint work with Alexey Upravitelev.

Recent surveys suggest that younger Americans are shifting toward riskier assets: cryptocurrencies, sports betting, prediction markets. We hypothesize that selective disclosure on financial social media contributes to this shift: gains and losses are amplified while ordinary returns go unshared, distorting the advice available to retail investors. We study both the supply and demand sides of retail financial advice in a laboratory experiment. In Individual sessions, conducted first, subjects complete a portfolio allocation task and choose whether to supply financial advice. In Social sessions, subjects observe the resulting leaderboard and are given the option to imitate. We study endogenous selection into both the supply of financial advice and the consumption of that advice through imitation, and how each interacts with task difficulty and performance — the resulting portfolio selection and risk. We hypothesize that limited understanding of market fundamentals, coupled with selective disclosure that goes unappreciated, results in excessive risk-taking on the part of consumers of retail financial advice.

Austin Brooksby, UCSB, "Learning with competing models."

Many consequential beliefs are formed when more than one model of the world is plausible and evidence arrives over time: jurors weigh rival accounts of a case, investors weigh rival advisors, voters weigh competing claims about a policy. Two classes of theories describe how people learn in such settings. Model averagers weight each model by how well it fits the evidence so far and revise gradually; the class ranges from Bayesian updating to rules that weight models by the wrong amount. Model selectors adopt one model, reason as if it were true, and abandon it wholesale when the evidence tips, so beliefs jump from one model's implication to the other's; the class's rules differ in when they switch. Experimental evidence on model uncertainty comes almost entirely from single updates, where wholesale switching cannot show itself. We study the dynamic problem in an online experiment. Participants predict the output of simulated machines, each governed by one of two disclosed candidate rules. Evidence arrives as a sequence of small inspection samples on which the rules nearly agree, while the predicted quantity is one on which they disagree sharply, so each prediction reveals the weight a participant places on each model and can be compared with the Bayesian weight given the same evidence. In a pilot with 32 participants (19 usable after exclusions), 7 are model selectors (6 usable) and 4 are model averagers (2 usable), and 27 of 32 receive the same type on simple and complex rules; the remaining participants respond little to the evidence. Selectors switch models on about a quarter of predictions, and about 40% of switches move to the model the evidence disfavors. Treatments that vary the strength of evidence, the sample size and the cost of applying the models separate the mechanisms behind selection and averaging.

Alisher Batmanov, UCSD, "Fragile Learning From Others."

Behavioral biases in decision-making are widespread and often persist despite experience and transparent feedback. This paper examines whether exposure to others' optimal decisions can correct initial misconceptions and facilitate learning in an environment where departures from the theoretical benchmark arise from neglecting an informative signal in a worker hiring task. Using a laboratory experiment, I show that such exposure substantially improves the quality of one's own decisions. The analysis of the underlying mechanisms shows that this improvement is not driven by mechanical imitation, as responses to exposure are asymmetric in the quality of observed choices, pointing instead to selective adoption of observed behavior.

Adopting others' decisions selectively, however, need not reflect an understanding of why those decisions are optimal, so I further evaluate the transfer of these learning gains to a modified environment once exposure ends. The main result of the paper is that learning from others is fragile: the improvements do not survive a change in the environment's primitives, and the acquired gains from exposure dissipate entirely. Comparing these effects with those of explicit guidance further underscores the limits of social exposure as an effective policy tool. These findings highlight the dual role of observational learning: while it can enhance the incidence of optimal behavior, the improvements it generates often fail to generalize beyond the observed context.

Dexter Hsu, UCD, "Political Correctness and Political Polarization."

People increasingly hold back their political views and avoid conversations across party lines. We study what this does to what people learn. In our model, everyone hears what others say in public, but public statements are noisy, so others appear farther apart than they are. Learning what someone actually thinks requires a private conversation, and political correctness makes that conversation costly when the other person appears to disagree. Because the value of a conversation also rises with disagreement, the cost falls hardest on the conversations people would learn the most from. Moderates seek out more extreme partners, but extremes do not reciprocate. In groups, the moderate in the middle either becomes the bridge between the two sides or is left out. Reaching agreement requires everyone to talk to everyone, and political correctness makes this less likely, so disagreement survives even though everyone starts with a common prior and reads information the same way. We ask whether this disagreement persists over time.

Jindi Huang, University of Zurich (visiting UCB), "The Comparative Blind Spot: Misconceptions About Comparative Advantage."

Comparative advantage is a foundational concept in economics, yet little is known about how individuals recognize and act upon it in economic decisions involving specialization and cooperation. This paper documents systematic misconceptions about comparative advantage using an experiment. In the experiment, subjects face production tasks in which mutually beneficial cooperation can be identified either from absolute advantage or only from comparative advantage. They first choose how to cooperate and then propose a transfer to divide the gains. The results reveal two patterns when cooperation depends on comparative advantage. First, when participants have an absolute advantage in all tasks, they are significantly less likely to cooperate. Second, they divide the gains between themselves and their partner sub-optimally, lowering their own realized gains from cooperation. To understand the cognitive foundations of these misconceptions, we elicit both beliefs about the gains from co- operation and preferences over the resulting outcomes. Additional vignette studies show that these misconceptions persist in real-world contexts such as household labor division, team collaboration, and firm production, though they appear less relevant for attitudes toward trade.

Yu-Hsiang (Sean) Wang, Caltech, "Strategic Complexity in Decomposed Games."

A common assumption in game theory is that each player's scalar-valued payoff is the sum of utilities from different aspects of an outcome. The predictions therefore depend only on the aggregated payoffs and are independent of how these payoffs are distributed across aspects. However, players may evaluate these aspects separately, and some aspects are easier to reason about strategically than others. I experimentally examine whether the strategic complexity of each aspect affects behavior by comparing games presented as

a single payoff matrix with the same games presented as two component matrices whose sum equals the actual payoff. I find that game play systematically shifts toward the action recommended by the strategically simpler dominant component, in which each player has a dominant action. To account for this pattern, I develop a behavioral weighting model that allows payoffs from different components to receive different weights. The model captures systematic variation in behavior across payoff-equivalent decompositions that cannot be explained by aggregated payoffs alone.

Bridget Galaty, UCSD, "Incentive Structures and Strategic Uncertainty in Team Punishment." Joint work with Georgy Kalashnov and Isabel Trevino.

Holmström (1982) posited that in the presence of multiple equilibria, an incentive scheme that penalizes any player choosing low effort through the threat of punishment can sustain coordination on the efficient outcome. However, we know from a large literature in experimental economics that people struggle to coordinate on the efficient equilibrium in coordination games with multiple Pareto-ranked equilibria (see e.g. Van Huyck, et at 1990, Cooper, et al 1990). More recent work suggests that the degree of strategic uncertainty associated with the efficient action, as measured by the basin of attraction, is an important predictor of cooperative actions (Fréchette et al 2021).

We design a laboratory experiment not just to test the predictions of Holmström empirically, but to understand how the degree of strategic uncertainty in this environment (via the basin of attraction) characterizes the validity of Holmström's predictions. We simplify the setup of Holmström to work with a sequence of 2x2 games that capture the tensions and trade-offs of the original model. Starting from a "seed" Prisoner's Dilemma that represents the no-punishment benchmark, we introduce punishment schemes of varying magnitudes and forms. These generate a sequence of related Prisoner's Dilemma and Stag Hunt games that correspond to different contractual environments derived from the same underlying structure. Both within and across treatments we vary the basin of attraction to separate the effects of strategic uncertainty from level of punishment.

Preliminary results show that the main predictor of efficient coordination is the basin of attraction, while the absolute level of punishment has little predictive power on its own. However, we observe some evidence of threshold-like responses to punishment at the aggregate level, consistent with Holmström's prediction. Ongoing analysis will further explore the interaction between strategic uncertainty and collective punishment in determining coordination outcomes.

Vidhi Shah, USC Marshall, "Social Networks as a Source of Risk: Evidence from Past Banking Crises."

Social networks can act as a source of run risk for banks by directing depositor attention. Using recent U.S. bank failures, I find that a geography's social connectedness to a failed bank's headquarters predicts search attention for the failure and for deposit insurance. Deposits follow attention: a one-standard-deviation increase in connectedness to Silicon Valley Bank's headquarters predicts a 1.6% larger decline in uninsured deposits. The effect survives controls for distance, local news, mark-to-market losses, funding structure, and the economic and demographic composition of a bank's depositor base. It also persists at banks with below-median fragility, suggesting that the outflows reflect depositor beliefs rather than weak fundamentals. Smaller failures generate no measurable search response and produce no outflows at connected banks. Migration-

based connectedness shows a similar pattern for the 2008 failures of IndyMac Bank and Washington Mutual Bank.

Milena Jessen, University of Bonn (visiting Caltech), "Hierarchical Representations."

Evidence from cognitive science suggests that people compress multi-dimensional information into hierarchies, yet we lack causal evidence on how these representations shape learning, beliefs, and choice. I study this question experimentally. Participants learn success probabilities of firm–project pairs over multiple rounds and repeatedly choose the pair most likely to succeed. I exogenously vary whether outcomes are grouped by firms or by projects, inducing the grouped dimension as the top level of participants' hierarchy while holding information constant. The representation governs how participants engage with the data and where they direct their attention: they process information conditional on the grouped dimension, forming summaries of the data - conditional averages - that they were never asked to form. I elicit behavior at both ends of the resulting hierarchy: beliefs about the conditional averages at the top and incentivized choices over pairs at the bottom. Participants hold accurate beliefs about the averages along their induced dimension, while compressing information along the other. Optimizing locally within these top-level groups, they thereby translate belief distortions into systematic distortions of pair-level choices. Procedural data confirm that attention tracks the representation. Endogenous experiments demonstrate that individuals are not passive recipients of imposed structure: they adapt their representations to both cognitive costs and environmental diagnosticity.

Clément Herman, Caltech, "Small datasets as mental models."

This paper studies how cognitively constrained agents form beliefs when they cannot retain all observable information. I model belief formation as a two-step process: individuals first compress data into small datasets (a handful of cases constructed from a larger dataset) and then predict using similarity-based reasoning over those stored cases. Despite their compactness, these small-dataset mental models can encode a rich set of statistical relationships, including central tendencies, pairwise correlations, and conditional correlations, at low cognitive cost.

To study which mental models endogenously emerge, I introduce a novel experimental paradigm in which participants observe datasets of 12 real-life items described by 3 attributes, construct 1 to 3 ""guessing aids"" of their own design, and later use only those aids to predict hidden attribute values for random observations from the original dataset. The experiment was conducted with 166 Princeton undergraduates using 80 randomly generated datasets varying systematically in correlation structure and noise. In a baseline treatment, the number of guessing aids is exogenously fixed; in an endogenous treatment, participants choose their number of aids at a cost per additional aid.

I find that participants distribute their guessing aids evenly across the data space, minimizing the total distance between any observation and its nearest stored case. In the prediction phase, they follow a nearest-neighbor rule, moderating toward less extreme guesses when the nearest stored case is an imperfect match. This behavior is broadly consistent with k-medians clustering (a canonical compression algorithm with strong theoretical guarantees) where observations are partitioned into groups each summarized by a representative prototype.

I document inefficiencies in prototype construction, including excess sensitivity to outliers and a tendency

to store observed exemplars rather than constructed prototypes, as well as failures to fully leverage stored information. Some inefficiencies are well-calibrated to prediction behavior; others are not. I further document heterogeneity in clustering strategies and limited ability to tailor mental model richness to the information environment.

These findings deliver systematic patterns for belief formation: participants preserve central tendencies well with a single case and capture pairwise correlations adequately with two or more, but poorly retain conditional correlations regardless of model size. Accuracy gains from richer models are concentrated in structured, low-noise datasets and for outlier observations; unstructured datasets yield no meaningful gains from additional cases.

6. Attendees

Thank you all for coming. This is just to keep a fun log of our growing BEESCUitS community.

Aaron Berman, Caltech; Alisher Batmanov, UCSD; Anastasiia Morozova, UCSB; Austin Booksby, UCSB; Brandon Kabernagel, UCI; Bridget Galaty, UCSD; Camila Farres, Caltech; Carl Romer, UCSD; Clément Herman, Caltech; Dexter Hsu, UCD; Erik Scheuermann, Caltech; Francesco Crispano, Caltech; Gayoung Koo, USC Marshall; Jack Adeney, ESI–Chapman University; Jamie Martini-Tibbs, UCSD; Jindi Huang, University of Zurich (visiting UCB); Kailin Zhu, UCSB; Kyoowon Yang, UCSD; Mario Morales, USC; Milena Jessen, University of Bonn (visiting Caltech); Odera Onyechi, USC; Olya Rospuskova, Caltech; Renee Wang, UCI; Saurav Lahon, Chapman University; Sota Minowa, Caltech; Vidhi Shah, USC Marshall; Wei-Chien Weng, UCD; Xinlu Zhao, UCSB; Yu-Hsiang (Sean) Wang, Caltech; Zander Memon, UCD; Zoë Arnaut, UCSD.

Faculty Lunch Attendees

Colin Camerer, Caltech.