1. Introduction and Summary

My research is centered at intersection of psychology and economics, employing the three core types of analysis in economics: theoretical modeling, empirical analysis (reduced form and structural), and experimental methods. I have four papers published or accepted in the "top-five" journals, three papers published or conditionally accepted in top-field or second-tier-general-interest journals, and a pipeline of three related papers.

My papers can be partitioned into three broad bins which illustrate the main themes in my work and demonstrate how my research has evolved over time. I first provide an overview of the three groups and then discuss each in the detail. Throughout, I favor casual language over formality in order to best explain my research succinctly to an unfamiliar reader.

(1) **Theory-driven identification of psychological biases in the field.** These three papers, all started in graduate school, identify psychological biases in economically-important situations through a theory-driven empirical analysis of behavior. The Sunk-Cost Fallacy in Penny Auctions uses a game-theoretic model to make predictions about bidding behavior in penny auctions and then empirically tests these predictions on a large dataset of actual bids, concluding that bidders suffer from the sunk cost fallacy. This paper was published in *the Review of Economic Studies*. Ballot Position, Choice Fatigue, and Voter Behavior, also published in *the Review of Economic Studies,* empirically exploits a natural experiment in which different voters see different contests at different points in the ballot, demonstrating that voters change their decisions when they are fatigued from previous decisions. For Using Competition to Elicit Cooperation in a Political Public Goods Game: A Field Experiment, published in *Economic Inquiry*, my co-author and I ran a field experiment that subtly varied the text on a politician's contribution-solicitation letter, showing that contributions can be increased by appealing to cooperative motivations, but increased even further when framed in a competitive light.

(2) **Structural identification of time preferences using laboratory experiments.** This cohesive set of three papers uses theoretically-guided structural analysis of laboratory experiments to answer fundamental questions about time discounting (the way that people value consumption received sooner relative to that received later). The projects all answer different questions following the same basic insight: the vast majority of the large empirical literature on time discounting examines tradeoffs between different time-dated *monetary payoffs* rather than time-dated *consumption*. The issue with the monetary-payoff methodology is that money is easily transferred over time and probably not converted in consumption at the time of receipt, so the way that people discount money is likely not representative of the way they discount consumption (which is what we often care about). To solve this issue, these three studies use an innovative method to examine decisions over time-dated non-transferable negative consumption in the form of the completion of unpleasant tasks. Working Over Time: Dynamic

Inconsistency in Real Effort Tasks, published as the lead article in *the Quarterly Journal of Economics*, demonstrates that people do in fact discount consumption differently than money. While monetary discounting is "time consistent," consumption discounting is "dynamically-inconsistent," meaning that people change their preferences as the time of consumption approaches. An Experiment on Time Preference and Misprediction in Unpleasant Tasks, recently accepted at *the Review of Economic Studies*, looks into people's perceptions of their own time discounting, finding that people are largely naive about their own time-inconsistency. Finally, Short-Term Discounting of Unpleasant Tasks is a recently completed paper that looks in great detail at exactly how people's preferences change over a week and estimates the short-term time-discounting curve, a fundamental objects in economics.

(3) **Creation and movement of beliefs: theory and empirics.** The third partition contains four papers that examine people's beliefs and how they change over time. The first two papers are distinctive, while the second two represent the start of a new and important research agenda. The Economics of Faith: Using an Apocalyptic Prophecy to Elicit Religious Beliefs in the Field, published in *the Journal of Public Economics*, uses theory-guided time-discounting experiments to measures the beliefs of a set of people who state that the apocalypse is imminent and tests if those beliefs change when the penalty for incorrect beliefs is increased, finding that these people hold seemingly non-elastic, extreme, and sincere beliefs in the end-of-the-world. To Reveal or Not to Reveal: Privacy Preferences and Economic Frictions , conditionally accepted at *Games and Economic Behavior*, theoretically examines the strategic interaction of two agents who want to determine if they are similar but are concerned about how revealing information changes the other agent's beliefs, identifying an optimal dynamic conversation protocol to reduce information revelation while still finding appropriate matches. The final two papers are highly related: both examine constraints on how rational people can change their beliefs over time given information. The key insight is that rationality implies that when people learn and move their beliefs, they must become — on average — concomitantly more certain about the world. Belief Movement, Uncertainty Reduction, and Rational Updating introduces this relationship, examines its implications, uses it to create statistical tests of rationality, studies the relationship between the test and four common psychological biases, and then employs these tests on datasets of beliefs, finding that individuals move their beliefs too much. Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence expands the idea to financial markets and risk-neutral beliefs, preliminarily finding that option prices are too volatile. I am bullish about the broad potential of this research vein.

Finally, I note that I am broadly successful at explaining core concepts of Economics in the classroom: I have repeatedly taught Core Strategy to MBAs, receiving a combined mean rating 6.16 out of 7 (compared to a combined mean rating of 4.1 by others for the same class in the same period).

2. Detailed Description of Research

I start with a summary table and then describe the individual papers in the three categories:

Paper Title	Co-Authors	Status
1) Theory-driven identification of psychological biases in the field		
a) The Sunk-Cost Fallacy in Penny Auctions	(none)	*Review of Economic Studies* (Jan 2016)
b) Ballot Position, Choice Fatigue, and Voter Behavior	Scott Nicholson	*Review of Economic Studies* (April 2016)
c) Using Competition to Elicit Cooperation in a Political Public Goods Game: A Field Experiment	Jesse Cunha	*Economic Inquiry* (Jan 2016)
2) Structural identification of time preferences using laboratory experiments		
a) Working Over Time: Dynamic Inconsistency in Real Effort Tasks	Muriel Niederle and Charles Sprenger	*Quarterly Journal of Economics* (Aug 2015)
b) An Experiment on Time Preference and Misprediction in Unpleasant Tasks	Matthew Rabin	*Review of Economic Studies* (Accepted)
c) Short-Term Discounting of Unpleasant Tasks	(none)	-
3) Creation and movement of beliefs: theory and empirics		
a) The Economics of Faith: Using an Apocalyptic Prophecy to Elicit Religious Beliefs in the Field	Jesse Cunha, Ernesto Dal Bo, and Justin Rao	*Journal of Public Economics* (Sept 2016)
b) To Reveal or Not to Reveal: Privacy Preferences and Economic Frictions	Aaron Bodoh-Creed	*Games and Economic Behavior* (Cond. accept)
c) Belief Movement, Uncertainty Reduction, and Rational Updating	Matthew Rabin	-
d) Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence	Eben Lazarus	Near Draft (intro only in the link)

2.1 Identifying Psychological Biases in the Field

At the start of my career, I empirically examined important field settings to identify psychological biases. This is a somewhat standard path in the field of Behavioral Economics: find an interesting environment in which people appear to be making a mistake and then document it. In line with my later work, the greatest differentiator of these three papers from the literature is probably the strong theoretical grounding, which I would argue allows for better pinpointing and measurement of each bias.

(a) "The Sunk-Cost Fallacy in Penny Auctions" is probably the best work in this group. It is also the most similar to my future work in using a somewhat-sophisticated theoretical analysis to create straightforward predictions about behavior under different behavioral assumptions, which are then tested using data.

More specifically, the paper analyzes behavior in the "penny auction," an auction format in which average empirical revenues exceed 150% of the good's value. This creates a puzzle — what causes people in this auction to overbid? In the auction, players pay money to become the "bid leader" for a good, and win the good if no other player places another bid in the following 30 seconds. Interestingly, the auction shares many characteristics with the "dollar auction," in which two players sequentially bid slowly escalating amounts to win a dollar bill but are both required to pay their last bid. The dollar auction is well-known as a classroom exercise that demonstrates the "irrational escalation of commitment" (also known as the sunk cost fallacy), in which players become less willing to exit a situation as their financial and mental commitments increase (even if these commitments do not increase the probability of success). This suggests that the sunk-cost effect also could be playing a role in penny auctions, as players make similarly escalating financial commitments (in the form of bid costs) as the auction continue. Of course, there are other explanations for the excessive bidding, from risk-seeking behavior to simple joy-from-winning. To differentiate these hypotheses, I theoretically solve for the equilibrium bidding behavior in the auction given these different possibilities and show that each implies different behavior over the course of the auction.

To test these models, I collect a large dataset of actual bids from one of the auctioneers' websites consisting of nearly 100 million bids on 166,000 auctions. Through reduced-form and structural estimation, I show that patterns that are consistent only with the sunk cost model: there is a significant deviation of the individual and auction-level hazard rates from the standard game-theoretic predictions, and this deviation increases with sunk costs. Finally, I posit a structural reason that competition between auctioneers does not reduce profits in the market, which (unfortunately, perhaps) still thrives.

(b): "Ballot Position, Choice Fatigue, and Voter Behavior" (with Scott Nicholson) also documents behavior that does not fit standard economic models in a meaningful environment using a large dataset. In this case, standard models predict that people make thoughtful decisions in all situations. However, many of us recognize that our ability to make good decisions gets poorer as we become tired or fatigued from making many previous decisions.

The paper shows that this type of "choice fatigue" can have important implications in a meaningful environment by creating a large novel dataset of voter behavior and empirically exploiting a natural

experiment in California. Specifically, due to voting laws that set the ballot ordering, voters in precincts with more local elections see the exact same state-level proposition at different points on the ballot. The paper shows that, as the ballot position of a race drops across precincts (so voters will have made more prior decisions when reaching the decision), voters are more likely to abstain or use decision heuristics (such as voting for the status-quo or voting for the first listed candidate). The effect does not appear to be driven by the endogeneity in the number of local contests in a precinct, as this variable is uncorrelated with voting behavior on commonly-positioned top-of-the-ballot contests. We finally note that the effects are not small: choice fatigue drives abstentions up by 8% and affected the outcome of 6% of the propositions in the dataset.

Finally, **(c) "Using Competition to Elicit Cooperation in a Political Public Goods Game: A Field Experiment"** takes a similar approach to study people's decisions about campaign contributions. The paper uses a theoretical framework to make predictions about the effect of different framing choices for contribution requests and then tests the theoretical predictions using a large field experiment with 10,000 potential contributors. In the experiment, potential donors receive a similar postcard that only varies in providing information about (1) the contributions of members of their own party, (2) contributions of members of the competing party, or (3) no information about the contribution of others (the control group). The results suggest that eliciting competitive rather than cooperative motivations can lead to higher contributions in intergroup public good settings. On the extensive margin, the contribution rates in the competitive, cooperative, and control treatments were 1.45 percent, 1.08 percent, and 0.78 percent, respectively. On the intensive margin, with the exception of one larger contribution, the distribution of contributions in the competitive treatment first order stochastically dominates that of the cooperative treatment. As a result of the combined effect, the cooperative and competitive treatments to yield 15 percent and 82 percent higher total monetary contributions than the control, respectively.

2.2 Identifying Time Preferences in the Laboratory

The next set of papers (started in my first few years at Haas) focus on time preferences, a fundamental topic in economics. I see these papers as important contributions, and the general topic as fertile ground for future projects.

Time preferences describe how a person chooses to allocate consumption and work over time (ie how much should I save for retirement? Should I complete this report today or put it off to tomorrow?). The majority of economic work describes time preferences using the work-horse exponential discounting model, which is useful for its simplicity and clean implications. In this model, the worth of future utility in

any period t+1 is worth a fraction δ of utility in the previous period t. In the past twenty years, other models of time discounting have been suggested in order to account for behavioral regularities, such as procrastination, which are not predicted by the standard model. The most common non-standard model is that of quasi-hyperbolic time preferences, which mirror exponential preferences but add a discounting factor β to all future consumption (but not to immediate consumption), leading to "present-bias" (essentially a desire for immediate gratification). This simple addition to the theory leads to a prediction of a variety of "time-inconsistent" behaviors in which the person's plans to maximize consumption and work can change over time, such as procrastination.

Given the importance of this topic, there are many studies on time discounting, both using controlled decisions in the lab and decisions in the field. The vast majority of this work examines choices between two time-dated monetary payments, largely because this provides a simple method to study time preferences. Many of these studies find large amounts of present-bias (a strong desire for immediate payments), which has been taken as support of the quasi-hyperbolic discounting. Unfortunately, the theory is fundamentally about consumption and not money — that is, the theory actually predicts that people (who are not cash constrained) will not be present-biased with respect to money because money is fungible and easily transferred across periods at a set rate. Consistent with the theory, a variety of recent papers have suggested that people are not present-biased with respect to monetary payments, and that the previous results were due to heuristics (mental shortcuts used to compare two simple options) or confounds (such as future payment reliability or future transaction costs — subjects are certain that the immediate payment will costlessly occur but are suspicious of future payments). This creates a gap in the literature — if all the time preference experiments with money are suspect, what is the evidence for hyperbolic discounting? There are a few experiments with consumption (rather than money), such as chocolates or real effort, but the literature is sparse.

This sets the stage for the three papers. The first two papers involve a six-week experiment with laboratory subjects concerning tradeoffs of actual work over time. In each week, subjects make decisions about current and future work that they must complete. The basic questions are 1) whether people say that they want to do more work in the future than they actually do when the future arrives, and 2) whether they are aware that their preferences might change.

a) "Working Over Time: Dynamic Inconsistency in Real Effort Tasks" (with Muriel Niederle and Charles Sprenger) establishes three main results. First, following recent work, people do not appear to be present-biased with respect to monetary payments over time. Second, people *do* appear to be present-bias with respect to work (negative consumption). This is important because it establishes that the theory does make correct predictions about behavior in the consumption domain. Furthermore, the paper uses a structural analysis to identify the additional discounting parameter β. Finally, the paper

attempts to rudimentarily determine if people that are present-biased *believe* that they are present-biased. This is accomplished by giving people the option to commit to work decisions in order to stop their "future self" from procrastinating in the future — if a person is aware of their own present-bias, they would choose to stop their future self from choosing to do less work in the future. There is a significant positive correlation between subject's choice to commit and the subject's estimate of β, suggesting that those that are more likely to procrastinate are (at least partially) aware of this proclivity (although there is very little desire to pay for this commitment).

b) "An Experiment on Time Preference and Misprediction in Unpleasant Tasks" (with Matthew Rabin) further explores these issues using a new methodology and focusing on the issue of people's beliefs about their own procrastination. The paper first establishes a similar result to the first paper with a more precise design and estimation strategy: subjects, on average, want to do less work when they are making decisions on the day of work rather than when they are making decisions on days before work. Second, the paper explores the exact level of sophistication of each subject about their own level of procrastination. Using a structural analysis of the data, we are able to assign each subject with two parameters — the actual level of the time discounting parameter β and the subject's perceived level of their time discounting parameter, denoted β-hat. There is a positive correlation between these measures. However, it turns out that people hold only partially sophisticated beliefs about themselves — for every amount of deviation from exponential discounting, people only recognize about 20% of this deviation (that is, I realize that I actually won't want to complete this referee report tomorrow, but I don't fully recognize the strength of the effect). Finally, the paper establishes evidence of "projection bias" in decisions, which has been difficult to identify in real-life data. Projection bias is the idea that people project their present state (such as tiredness) on their future state when making decisions about the future. To identify this effect, we exogenously vary the number of tasks people complete before making decisions (thus varying their current state of tiredness) and find that people that are more tired when making decisions choose to do less work in the future, presumably because the task seems harder at that point.

Finally, **c) " Short-Term Discounting of Unpleasant Tasks"** uses a similar methodology to the previous paper but asks a much more focused and simple question: what is the exact shape of the time-discounting curve in the short-term (a week)? This curve is a fundamental component of all models with intertemporal tradeoffs, and its estimation can be used to (1) discipline and test the myriad theoretical proposals of the shape of the discount function, (2) understand why time-discounting occurs or how it was formed, (3) refine theoretical models of behavior given non-exponential discounting, and (4) provide rough guidance on the empirical effect of using different time definitions for "now" and "later" to estimate the quasi-hyperbolic parameter β. However, there is almost no work about the shape of this short-term function in the literature. To fill this gap, I run a study in which students answer

questions about how many transcription tasks they want to do for different wages. In this study, they work at the end of the week and are asked these work-preference questions many times throughout the week via text message. By repeating this design for multiple weeks over 100 subjects and varying the times at which the students are asked, one can fully understand how decisions change as work approaches: that is, the entire time-discounting curve can be non-parametrically estimated.

The main contribution of the paper can be summarized in two figures showing the changes in task decisions over the week and the consequent estimated time-discounting curve. The former figure shows that, given an average choice of around 42 tasks at the start of the week, the task decisions fall gradually by around 2 tasks by a day before work, another 2 tasks by a few hours before work, and then another 2 tasks by the time work arrives. The estimated discount function follows a similar pattern to the raw task data: people discount the cost of work by 87% when work is a week away, with around a third of the discount occurring in the first few hours from work and another third in the first day. Interestingly, when the data is used to test the fit of a variety of commonly-suggested shapes for the discount curve, many do surprisingly poorly given the initial steepness of the curve.

2.3 Understanding Belief Formation and Changes Over Time

The final grouping of papers focus on the formation, change, and effect of people's beliefs. The first two papers (started in the first few years at Haas) are somewhat idiosyncratic. While they are both interesting, the second two papers (occupying much of my recent time) are more groundbreaking and provide a solid foundation for another long-term research agenda.

a) "The Economics of Faith: Using an Apocalyptic Prophecy to Elicit Religious Beliefs in the Field" (with Jesse Cunha, Ernesto Dal Bo, and Justin Rao) focuses on people's beliefs about the external world. Before the paper was in progress, my co-authors and I had been fascinated by self-serving beliefs: beliefs that are not supported by evidence, but which people seem to hold in order to make them happier. People have suggested that religious beliefs might fall in this category. The problem with religious beliefs is that it is very difficult to know if 1) the belief is incorrect and 2) people actually hold the beliefs they say that they hold. The main issue is that religious beliefs are statements about events which are often not falsifiable (ie "God exists").

To subvert this issue, we focus on a falsifiable religious statement. Specifically, we exploit the fact that there was a (somewhat large) group of people that believed that the Rapture and Apocalypse was going to occur May 21, 2011, due to the well-publicized Biblical interpretations of the owner and host of a popular network of religious radio stations named Harold Camping. This prediction allowed us to assess the beliefs of followers using monetarily-incentivized methods and determine if the follower's beliefs

tempered with a rise in the cost of holding the belief. If people are unwilling to hold these beliefs when the cost is relatively high, it suggests that there must be a tradeoff from holding the belief.

Specifically, we had followers make time-discounting decisions about receiving money on May 7th and receiving larger amounts of money of May 28th (after the Apocalypse was expected to happen). We varied the amounts provided on May 28th and the probability that we would pay people either amount. Somewhat amazingly, neither of these variables had an effect on people's decisions. Almost every subject choose $5 on May 7th rather than any amount (up to $500) on May 28th, regardless of the probability of payment. A control group (a set of Seventh Day Adventists who hold similar Apocalyptic beliefs but do not focus on May 21st as a special day) made time discounting decisions that largely match those seen in past studies with students. Therefore, it seems that the followers truly believed and appeared to have no elasticity in their beliefs.

To understand these results and guide the experimental design, we create a formal theory of religious beliefs that explicitly models the experimental intervention and offers a faith-rooted microfoundation for the demand side of religion.

b) "To Reveal or Not to Reveal: Privacy Preferences and Economic Frictions" (with Aaron Bodoh-Creed) is the only purely theoretical paper in my portfolio and is arguably the most distant from my normal work. The main connection between this and the other papers is the focus on beliefs and information as a primary driver — in this case, the desire to control the information revealed to others.

The basic idea of the paper is to explore how people can learn that they "match" with other people (that is, they share the same type) when they also prefer to keep information about themselves private. For example, a company needs to reveal information to a potential partner about its operations, but does not want to reveal this information to the potential partner in the case that they do not decide to become partners. Similarly, a revolutionary wants to find other people with similar goals, but does not want to reveal his goals to people sympathetic to the current regime.

We develop a theoretical model that explores the tension of this type of desire for privacy with a desire to find a matching partner, which requires revealing information. The main result shows that there is a specific form of "conversation" (dynamic revelation of information) that people prefer in order to dynamically screen out non-matching partners. In this preferred conversation, people reveal information slowly and reveal less sensitive information early on. Interestingly, when a certain type of person is very rare, the person might prefer to not reveal any information, leading to a "taboo" in which people of that type never identify themselves or find matches.

I am much more bullish about **c) "Belief Movement, Uncertainty Reduction, and Rational Updating"** (with Matthew Rabin), which studies the restrictions of rational belief changes. To understand the problem, suppose that we observe the evolution of a person's beliefs about some outcome (ie Will the stock market go up or down? Will Barack Obama become president? Will team A win the game?). For example, a person might believe that a team has a 20% chance of winning a game, see a goal and change their belief to 80%, see a save and change to 30%, and so on. Now, imagine that we see these belief movements before the event resolves. Is there some way we can say that these beliefs moving around too much? On one side, it seems impossible to ever conclude that beliefs are moving too much without a good model of the situation. After all, there is always some data-generating process that would rationalize any stream of beliefs. However, intuition suggests that a large amount of movement should be a relatively rare event for a rational Bayesian updater. For example, if a person's beliefs move over and over from 5% to 95% and back again, one might start to question the person's "Bayesianess."

In this paper, we formalize this intuition by developing simple measures of belief change and uncertainty and showing that a person cannot change beliefs (which suggests learning) without correspondingly becoming more certain in the outcome, at least on average. A simple corollary of says that every data-generating process with the same prior belief and ending with the person certain about the outcome (and thus having the same amount of uncertainty resolved) must lead a Bayesian to exhibit the same amount of total belief change on average.

Importantly (and surprisingly), these statements do not depend on the structure or meaning of the information that the person is observing over time. Therefore, even if we know nothing about the normative meaning of the person's information flow, we can statistically reject that a person is a Bayesian if their total belief movement is consistently above or below her uncertainty reduction.

Interestingly, the concepts of movement and uncertainty reduction seem very related to a set of commonly-studied psychological biases. To demonstrate this, we create a simple model that embeds four biases and show that (1) the biases universally lead to clear and intuitive patterned violations of the movement-uncertainty relationship, and (2) the individual concepts of movement and uncertainty reduction seem to provide a nice way to classify and understand the differences between the biases.

We then demonstrate the use of the statistical test by analyzing belief data from three datasets: a prediction market on sporting events, a set of individual beliefs over time, and a computer-based probabilistic prediction. One of the most interesting results is for prediction markets, in the early stages of events, there is not much change in uncertainty, suggesting that there is little true information revealed in these stages. However, there is a relatively significant belief movement in the periods

between these stages. Conversely, in the later stages, a large amount of uncertainty is resolved. However, there a relatively small amount of movement in the periods between these stages. That is, early pieces of information appear to be relatively uninformative, while later pieces of information are relatively informative, but it appears that people do not fully appreciate this time effect on the informativeness of information.

Finally, d) **"Restrictions on Asset-Price Movements Under Rational Expectations: Theory and Evidence"** [the link only includes the introduction] (with Eben Lazarus) extends the logic of the previous paper into the movement of prices of financial assets. The difficulty in this extension is that asset prices do not only reflect beliefs, but also people's changing value of money given different outcomes. For example, investors presumably value an additional dollar more during a recession (when they and everyone else has less money) than during a boom. Consequently, the amount they are willing to pay for a financial instrument that provides an additional dollar when a recession happens is based both on (1) their probabilistic belief in the recession and on (2) the *relative value* of money in a recession. Consequently, asset prices are said to represent *risk-neutral* beliefs rather than simple beliefs. This distortion unfortunately implies that the theory from the above paper is not applicable to asset prices.

However, the paper shows that it is still possible to make statements that bound the movement of risk-neutral beliefs. Specifically, given an assumption about the stability of the relative value of money across wealth states, one can identify the "worse-case" data-generating process of information flow that leads to the highest amount of risk-neutral-belief movement given these relative values. Furthermore, it is possible to show that there is some amount of movement that cannot be justified by *any* relative value of money. Consequently, movement in the prices above this amount is simply not possible in the model. We then take the theory to data by studying the empirical movement of risk-neutral beliefs implied by option prices. Our initial results suggest that the movement cannot be justified in the model or require unreasonably high relative values. Consequently, our initial results reject the predictions of the standard asset pricing model.