Should I Stay or Should I Go: Microeconomic determinants of migration

Zack Barnett-Howell

October 26, 2017

Abstract

There appear to be inefficiencies in the distribution of people across space as individuals stay in locations that offer little observable welfare. To answer this question this paper develops a new microtheoretic model of decision-making under uncertainty. The model decomposes the migration decision into separate decisions over when to leave and where to go, stopping time and matching rules, respectively. These decisions are made using informative but noisy signals over a person's underlying type. This approach makes an intractable dynamic problem solvable using static rules, and the level of noise in the signal can produce ex-ante rational but ex-post sub-optimal decisions. This theory is tested in a lab experiment where participants make stopping time and matching decisions over locations in a game and receive monetary rewards according to their performance. This experiment is run at universities in the United States and Ethiopia, where the results demonstrate that participants solve the stopping time and matching rules according to the model, with no difference in average payoffs between study sites. Furthermore, an information treatment about a player's type shows the primary role that information constraints play in optimization behavior.

[1]Zack Barnett-Howell. Department of Agricultural & Applied Economics and Department of Political Science, University of Wisconsin – Madison. 301 Taylor Hall 427 Lorch St Madison, WI 53715. zbarnetthowe@wisc.edu. I appreciate the significant help provided on this project by Kait Sims, Matt Kimball, Getachew Alemayehu, and the faculty of Economics at Bahir Dar University. This research received clearance from the University of Wisconsin–Madison's Institutional Review Board via Protocol 2017-0448 and 2017-0448-CP001 ("Choice Under Uncertainty"). Financial support for this research was provided by a 2016 grant from the Behavioral Research Insights through Experiments (BRITE) Lab and Department of Agricultural and Applied Economics, University of Wisconsin – Madison.

1 Introduction

Income mobility over generations appears to be largely determined by geography, with strikingly different likelihoods of economic advancement across locations.[1] The distribution of people across space therefore appears inefficient when looking at individual migration patterns, because a large number of people remain in areas with low observable levels of welfare, despite few obvious barriers to movement. Location persistence in the face of significant unrealized welfare gains is the puzzle that motivates my work. I argue that location choice is impacted by information constraints, with immobility the result of poor information about a person's individual-specific benefit from residing in another location.

Beneath the impressive progress that has been made in studying migration, there is a challenge: migration data do not match theoretical predictions. Migration patterns routinely fail to fully reflect push and pull factors as rural areas are depopulated and urban areas grow unmanageably large; migration destinations seem to be over-determined by existing network structures; most problematically is the lack of migration from areas with clear push factors to areas with significant pull factors. The lack of migration from areas with little observable amenities or goods needs to be explained.

I construct a microtheoretic model of decision-making under uncertainty that decomposes migration into separate stopping time and matching decisions.[2] An individual chooses when and where to migrate as independent decisions that are a function of noisy signals about their type-specific payoff at their current and alternate locations. Over time people accumulate signals and update their beliefs about their own type, choosing to stop — i.e. move — and where to move based on these beliefs. I

[1] The work of Chetty et al. [28], Chetty, Hendren, and Katz [26], and Chetty et al. [27] enumerates the near-deterministic role of geography in income mobility. The authors show that while some areas the U.S. have income mobility comparable to the highest mobility countries in the world, such as Canada and Denmark, "others have lower levels of mobility than any developed country for which data are available." Chetty and Hendren [25] find that by moving to a better neighborhood, children's outcomes improve linearly as a function of exposure to the better environment. The causal mechanisms behind this stark disparity in opportunity are not yet well understood, but the role of geography is significant.

[2] A *stopping time problem* is the decision whether to advance a stochastic state and receive a continuation payoff or whether to freeze that state. Examples cover the entire lifecycle, such as the decision when to park a car [87, 86, 69], when to replace a bus engine [77], when to leave school [48], when to take a job [83, 55], when to have a child [91], when to quit a job [89], when to retire [18, 84], when to buy a cow [70], and when to retire a cow [59].

A *matching problem* is how to allocate one thing to another. Examples include matching tenants with landlords [7], matching houses to people [78, 79, 4], matching people to people (i.e. marriage) [15, 16, 17, 19], matching children to schools [3, 2, 6, 5, 1], matching colleges to students [39, 23, 24], matching doctors to hospitals [71, 72, 65, 58], matching kidneys to patients [74, 75, 73], and matching managers to companies [38]. For an overview of the extensive literature on matching see: Niderle et al. [64] and Chade, Eeckhout, and Smith [22].

predict how information constraints reduce mobility and high variance in the signal causes errors in sequential optimization.

The result is, ex-ante, people make rational choices about location choice, even though, ex-post, these decisions are not optimal. By focusing on the importance of noisy signals, this theory can explain the discrepancy between ideal migration trajectories and those that are actually observed. The result is a set of predictions that comports with observed human behavior. I construct these predictions without assumptions that other theories of migration rely on, such as the primacy of idiosyncratic preferences or the reliance on large but unobserved migration costs.

To test my theory I use a bandit algorithm as an approximation of the the process by which people solve decisions under uncertainty. These algorithms function under significant information constraints, and have update and decision rules that are analogous to those that a person could employ. Because bandit algorithms make mistakes as a part of their optimal functioning, contrary to dynamic programming solutions which are optimal at all periods of time, they can fit sub-optimal trajectories without programmatically introducing costs to migration as a way of rationalizing immobility and ex-post sub-optimalities.

To implement this test, I design a lab experiment involving sequential optimization. I create a computer game where participants make stopping time and matching decisions, receiving higher payouts for their participation based on their average score over several games. Despite tight information and time constraints, participants achieved a payoff equivalent to that of a bandit algorithm, exhibiting learning within and between games. In addition, some participants were subject to an information treatment which allowed them to perform better and arrive at solutions that are much closer to optimal. This demonstrates the degree to which additional information constraints are binding in reducing the realization of optimal outcomes.

This paper provides a new research tool for examining sequential optimization problems under uncertainty. I construct a parsimonious model of migration that shows how information constraints increase the distance between optimal and observed trajectories. I further show how bandit algorithms form a robust solution concept for this class of dynamic problems, and how this solution is qualitatively similar to how people behave in a laboratory setting. My model helps illustrate the fundamental

drivers of migration as endogenous to the individual signals a person receives and how they update their beliefs using their signals, a process that can extend to a large set of questions regarding location, employment, and education.

2 Background

Attempts to explain the fundamental drivers of migration go back to the late 1800s in Ravenstein [66, 67]. The “laws of migration" enumerated in these articles later emerge in in the eponymous Harris-Todaro model [88, 46], which can be understood as a generalization of the observed flows of people being pushed from poorer and pulled to richer areas.

What drives migration in these models is the fundamental variation of goods over space. As not all places are similar, movement from one area to another is a simple method to acquire access to a better job market, more educational opportunities, economically vibrant markets, and other location-specific goods. The individual decision can be thought of in a more atheoretic sense, as gravity models do,[3] while more sophisticated theories have expanded on the interplay of household and network decisions in deciding when and where people migrate [60, 57].

Tunali [90] focuses on the stay/go decision, but largely ignores the question of where the migrant is considering going. In contrast Dahl [32] allows many destinations, but treats migration as a one-shot decision. Gallin [40] builds a macro-level theory of migration, but does not consider the individual decisions that comprise these large movements.

This discrepancy between the potential for significant welfare improvement as a function of mobility and the observed stationary behavior is seen in empirical work. Clemens [30] argues that wage gaps are determined by location not by worker characteristics, and the potential gain of people effectively arbitraging these gaps would be tremendous [29]. [4] Gibson, McKenzie, and Rohorua [42] show that migration offers tremendous benefits. In the case of migrants from Tonga and Vanuatu, the authors show an increase in income, consumption, durable goods ownership, and subjective standard

[3] Simini et al. [80] defines a sophisticated gravity model, which estimates migration flows between locations using terrain features, mostly distance, to compute possible paths. This method can produce highly accurate results, but the inherent downside is that it does not describe the mechanism producing migration flows.

[4] Ashenfelter [8] finds a differential in wage rates of 10 to 1 for McDonald’s workers in differential countries, suggesting that location makes all the difference when holding skillset and position constant.

of living. Increases that they claim “dwarf those of other popular development interventions." Moving from a poor to rich area then is a singular action capable of dramatically increasing individual income, and therefore the decision not to do so can be puzzling.[5] Munshi and Rosenzweig [62] go on to argue for the fundamental “missallocation" of labor across countries, pointing to the vast productivity differential between rural and urban areas. If prior work was more concerned with too much mobility and migrants from rural areas appeared in the *urbs*, then later work has been concerned with the lack of mobility among the poorest. Munshi and Rosenzweig [62, p. 47] argue that the lack of take-up of migration as a technology is puzzling given the opportunity for “substantial opportunities associated with spatial wage differentials in India to move permanently to the city."[6]

Yet the persistence of wage gaps and the persistence of the decision *not* to migrate is not easily overcome. An experiment conducted by Beam, McKenzie, and Yang [14] failed to induce significant migration, despite offering information about migration, assistance with a job search, and assistance with the administrative difficulties of migrating (i.e. passport applications). The authors a zero impact of these treatments on the propensity to migrate overseas. This suggests that information over locations, the prospects one might have there, and other barriers are not the limiting factor to migration.

Bryan, Chowdhury, and Mobarak [20] implement an experiment to increase rates of migration in some of the poorest areas of Bangladesh. Their experiment proved successful in using a small conditional cash transfer to increase rates of migration for treated households, which then translated into higher average welfare outcomes conditional on migration. The authors find that seasonal rural-urban migration induced by the treatment increased consumption of an origin household by $20 per month. A gross return is conservatively estimated at 273%.

Yet their result leads to a more problematic question: if the welfare returns to migration were significant, why were households not migrating in the first place? The honesty with which the authors address this puzzle is refreshing. They write: “we cannot provide a fully satisfying explanation for why people in Rangpur had not saved up to migrate," and that future research was necessary to

[5] See Yang [92] for further evidence on the positive externalities to migration (remittances). For a long-run view of the impact of migration see Gibson et al. [43].

[6] Munshi and Rosenzweig ultimately decide that a combination of informal insurance and social limit mobility, finding culture as a limiting agent as in their previous work [63].

create a means of explaining the low rate of migration with its relatively high returns.[7] Explaining the mismatch between observed rates of migration and their relative payoff remains a fundamental question.

However, the framing of migration as a "technology" gives us some insight into its lack of adoption. Foster and Rosenzweig [36] find that experiential learning is a critical aspect of technology dispersion. [8] There are cases where observing the returns to the adoption of a technology by a neighbor is insufficient, where Duflo, Kremer, and Robinson [35] find little spillover effects from neighbors in fertilizer adoption. The reason for this may be found in Munshi [61] where knowledge of the relationship between observable factors and an underlying state space changed how farmers could learn socially from each other. Essentially, if a farmer understands how inputs and unobserved and unobservable farm characteristics interact then he or she can learn from others. But if factors are dissimilar across a broad number of dimensions then a farmer may be unable to learn by proxy through the experiences of others.

The work in Kennan and Walker [50, 49] provides a sophisticated approach to understanding the nature of the migration decision as a dynamic programming problem. Using individual-level data from the U.S. National Longitudinal Survey of Youth, Kennan and Walker estimate an average individual moving cost of 312,000 to 384,000 U.S. dollars. This cost parameter is necessary to explain low rates of mobility from poorer to richer areas, where a counterfactual move would offer significant welfare advantages.

When Kennan and Walker estimate the costs of realized moves, they recover a value that ranges from-148,000 to 138,000 dollars. The authors state that the "the typical move is not motivated by the prospect of a higher future utility flow in the destination location, but rather by unobserved factors yielding a higher current payoff in the destination location, compared with the current location."[9] We are then left in the position of attaching low moving costs to individuals who move and high moving costs to individuals who stay. This arrangement is practical, yet does not allow us to examine

[7] Bryan, Chowdhury, and Mobarak [20, p. 1675] further say: "our quantitative results show that we do not fully understand the migration choices of these households: there is some important aspect of their choices that we are not capturing. This final challenge leads us to briefly consider some departures from full information and rationality and other market imperfections (such as savings constraints). Ultimately, however, we lack the data to determine what ingredient would provide a fully satisfying characterization of the behavior we observe, and leave this to future research."

[8] For an overview of the learning literature see Foster and Rosenzweig [37].

[9] Gemici [41] incorporates the effect of marriage to explain some of these costs, but they still remain high overall.

the actual determinants of these costs. It suggests that any similar approach will arrive at a similar conclusion. We can populate a vector of location and personal characteristics at any resolution. We will find many significant predictors, and yet no obvious determinants of migration.

The dominance of idiosyncratic unobserved costs is a result of the dynamic programming solution concept that research from McCall and McCall [56] to Kennan and Walker [49] employs. Although dynamic programming solutions to discrete decisions under uncertainty have a long tradition in applied microeconomics, there is both a conceptual and function problem. The conceptual problem is that the method of solving dynamic programming problems through value function iteration is entirely alien to how people solve these types of problems. The second is that it relies on the model possessing complete knowledge of the probability distributions of all possible state transitions [85]. The solution is not robust to cases where the state transition probability is uncertain — a characteristic of most important problems. This is where my model provides a way forward. By formalizing the unobserved innovations driving migration, we can begin to understand what creates variation in migration outcomes.

3 A Model of Migration

This is a parsimonious model of how migration is determined at an individual level. This theory is deliberately reductionist in that it focuses on a single main driver rather than the myriad secondary and tertiary factors that influence migration. My goal is to predict migration behavior without using exogenous costs to limit mobility. For this, monotone comparative statics provides strong predictions for migration behavior without significant assumptions on the functional form of utility. The end result is a model that preserves rationality while permitting sub-optimal outcomes to emerge in sequential decision-making. The incorrect priors necessary to comport with rationality are endogenously generated within the model and maintained in equilibrium.

3.1 The Story

A person lives in a location. They receive noisy signals about who they are, and based on these signals and their beliefs over their own type they choose when and where to migrate. I decompose migration into two conceptually distinct processes:

- A stopping time rule: the binary decision of whether to migrate
- A matching rule: the multinomial decision of where to migrate

Taken together, the stopping rule and matching rule form the trajectory of locations that an individual chooses over their lifetime. These rules are determined by a set of signals that an individual receives over their underlying type, the value of which determines their optimal location.

I begin with a finite set of locations and a continuous distribution of individual types, the joint distribution of which form the basis of utility draws. I assume the existence of an optimal location for each type, and while there is no uncertainty over location values, the parameter value of the type is unknown to the individual and is therefore treated as a random variable. An individual chooses a trajectory of locations over their lifetime to maximize the expected value of their utility function. Lacking certain knowledge over their type, the individual uses signals to infer the most likely value for their own type. The notation is as follows: the individual sequentially solves the stopping time rule and matching rule in periods $t_1, \ldots, t_T$; they choose among an ordered set of locations $a_1, \ldots, a_N \in \mathscr{A}$ conditional on their beliefs over type $\omega \in \Omega \subseteq \mathbb{R}$ and signal signals received $x \in \mathscr{X} \subseteq \mathbb{R}$. The signal is affiliated with type, a property that will create the requisite monotonicity conditions in the stopping and matching rule.

Individual utility is based on the quality of the match between individual and location. For any type there exists an optimal location a^* such that expected utility is non-increasing as the individual moves away from this location. Any optimal migration history maximizes the time spent at the optimal location, and minimizes the time spent at other locations.

Uncertainty over which location is optimal is necessary to produce a model that comports with observed cases where ex-post migration trajectories fail to include locations with higher observable welfare. The locus of this uncertainty can be placed on either individual type or the value of a location.

I argue that uncertainty over location value cannot be sustained in equilibrium, while uncertainty over individual type is a natural function of experience.

To see this, imagine the case where location values were uncertain. In this world, uncertainty over the amenities and potential payoff of moving to a location prevents migration to *better* locations. This requires a mechanism that sustains uncertainty over time. At some starting point t_0 perhaps there exists insufficient information over the quality of a given location; over time, however, as the amount of information increases and initial migrants go to these places and send word back home, this uncertainty would evaporate. Furthermore, there is no hard limit on the amount of information an individual could acquire over an external location if they were sufficiently motivated to do so.

Experiences, and by extension time, act as a natural gating mechanism for acquiring information. If experience is necessary to generate information then there is no way to generate *additional* information. One can not generate extra experiences in a fixed amount of time. Moving uncertainty from location to individual doesn't change the result that the match-payoff to a given location is uncertain. But it does allow this certainty to be sustained in equilibrium.

3.2 Stopping Rule

The migration stopping rule is whether an individual should stay at or leave their current location. This is not a choice over where they should go, but rather whether they should go. This stopping rule is a probabilistic decision procedure that can be described by the functions ϕ_{stay} and ϕ_{go} with two points of support:

$$\phi_i(x) = \begin{cases} 0 \\ \\ 1 \end{cases} : \sum \phi_i = 1 \tag{1}$$

The value of leaving is increasing in type such that the stopping rule is monotone in the signal. Without loss of generality $\phi_1 = 1$ as x lies outside of an interval (x_i, x_{i+1}).[10] What this means is that there exists a threshold such that for any signal above that threshold the decision to *stop* is always chosen.

[10] For the proof see Lehmann [54].

3.3 Matching Rule

The matching rule is a choice over locations according to their expected payoff; it is only observed in cases where the stopping rule is activated. In cases where the continuation decision is made, then no matching result is unobserved — except in cases where the stopping rule is activated but the matching rule points to the same location.

The matching rule $a(x)$ is a function of the signal received and determines the utility draw the individual receives. The individual's utility function maps from the product of type and location into the reals $u : \mathscr{A} \times \Omega \to \mathbb{R}$. The joint density of type and signal is represented by the function $f : \Omega \times \mathscr{X} \to \mathbb{R}$. Since type is unknown the objective function is defined as:

$$U(a,x) = \int u(a,\omega) f(\omega;x)\, d\omega \tag{2}$$

Expected utility is monotonically increasing in type ω by two minimally sufficient conditions: [9, p. 200]:

$$u(\cdot,\cdot) \text{ Has the single crossing property in} \{a;\omega\} \tag{3}$$

$$f(\cdot,\cdot) \text{ is log-supermodular} \tag{4}$$

We can achieve these minimally sufficient conditions for cases where there exists positive assortive matching between individuals and locations. When the location-match utility is increasing as a function of type the single crossing property is stated as follows:

$$u\left(a',\omega\right) - u(a,\omega) \geq 0 \Rightarrow u\left(a'\omega'\right) - u\left(a,\omega'\right) \geq 0 \tag{5}$$

This states that any higher location choice that would be profitable for a lower type would also be profitable for a higher type.

For the joint distribution of type and signal to be log-supermodular, we require that $f(\cdot,\cdot)$ have the monotone likelihood ratio such that $\frac{f(\omega;x')}{f(\omega;x)}$ is nondecreasing in type for all $x' > x$.

The first assumption is merely a description of the optimal distribution of people across space. This

generalizes single crossing property to apply to optimal match locations such that high types optimally match with high places, and lower types optimally match with lower places. This is a constructive attempt to formalize the notion of idiosyncratic location preferences. Locations constitute a partially ordered set, but the utility function $U(\cdot, x)$ is maximized at the type-optimal location: $a^* \mid \omega = \Omega$.

I add an independently distributed disturbance term $\epsilon \overset{iid}{\sim} F(\mu = 0)$ scaled by $\sigma > 0$ that enters the utility function non-separably:

$$U(a, x) = U(a, x, \epsilon) = \int \int u(a, \omega, \epsilon) f(\omega; x) d\omega d\epsilon \tag{6}$$

This says that expected utility is unchanged: integrating over the support of ϵ removes it from the equation. However, a small sample history of observed utility will not allow the individual to identify their type.

The purpose of this noise is to better model reality; outcomes in life are rarely deterministic. This term represents any number of events that enter into an individual's utility yet are uncorrelated with who or where they are. These stochastic events can be significant, yet are not informative over the goodness of fit in the match between a person and a location.

3.4 Loss Function

An optimal migration trajectory is one that admits no alternative move that would produce a greater utility — analogous to Pareto optimality. Any time spent at a sub-optimal location would make a trajectory sub-optimal, even when visiting that location is necessary to determine its sub-optimality. If migration decisions are meaningful as a way of revealing information, then it is necessary to consider the information set of the actors making that decision. Are their decisions optimal *ex-ante* rather than appearing sub-optimal when viewed *ex-post*?

I suggest the use of a loss function for two reasons:

- It fits the context of a risk-adverse agent making significant choices over uncertain outcomes *ex ante*
- It helps transition to the notion of regret which will be used later in the paper as part of my

solution concept

In each period the individual receives a signal $x \in \mathscr{X}$, solves the stopping rule for $\phi_i \in \{0,1\}$ and their matching rule $a \in \mathscr{A}$. They receive a utility draw centered on the match value with variance determined by the disturbance term:

$$U_{a,\omega} \sim F\left(\mu_{\{a,\omega\}}, \sigma^2 \text{var}(\epsilon)\right) \tag{7}$$

By period t an individual is at location a and has observed a history of signals $\mathbf{x} = \{x_1, \ldots, x_t\}$. For the remaining periods $t+1, \ldots, T$ the individual attempts to minimize their loss function, defined by:

$$\mathscr{L}(\mathbf{x}) = \left[\int u(a,\omega) f(\omega, \mathbf{x})\, d\omega + \beta V\left(a, f_{\mathbf{x}_{t+1}}\right)\right] - \sup\left[u\left(a', \omega\right) f(\omega; \mathbf{x})\, d\omega + \beta V\left(a', f_{\mathbf{x}_{t+1}}\right)\right] \forall a' \neq a \tag{8}$$

Where $f_{\mathbf{x}}$ represents the distribution of the signal conditional on the signal history; $V(\cdot,\cdot)$ represents the future payoff of being at a given location; $\beta \in (0,1)$ is a discount factor; and $f_{\mathbf{x}'}$ represents the conditional distribution of the signal in the next period.

An optimal stopping rule and matching rule minimizes the loss function over time

$$\phi^*(\mathbf{x}) : \arg\min \sum_t^T \beta^t \mathscr{L}_t \tag{9}$$

$$a^*(\mathbf{x}) : \arg\min \sum_t^T \beta^t \mathscr{L}_t \tag{10}$$

An optimal rule describes the correct time to leave a location, as well as the best location to go to at that time. These choices are monotonically increasing in signal.

3.5 Payoff

This model describes migration as the combination of the decisions whether and where to go. I allow for a heterogeneous distribution of people across places based on their underlying type, and permit sub-optimal ex-post trajectories due to uncertainty over underlying type. What makes this uncertainty over type compelling in a way that exogenous costs or uncertainty over locations are not is that it is

sustainable in equilibrium.

Exogenous costs have to become large in a dynamic context to justify immobility. As noted above, it is not obvious how uncertainty over location attributes can be sustained in equilibrium; there is no hard limit on the amount of information an individual can acquire. But experience over time provides a natural limit to acquiring information about an individual type. A person cannot experience more in a fixed amount of time. Experiences are revealing, but can also incorporate bias into what they reveal. And if this bias serves as a limit to mobility, then there is no obvious reason why it could not be sustained in equilibrium. Using longitudinal data it is possible to estimate counterfactual trajectories, where an individual would realize higher wages and markers of welfare — but on the individual level these trajectories are unrealized *off-path equilibria*, and as a result never enter the individual's information set.

The following propositions can be derived from the theory:

Proposition 1. A sequence of sufficiently high signals will lead to a stopping (leave) decision.

$$\exists\{x_t, \ldots, x_{t+k}\} : \Pr\left(\phi_{\text{stop}} = 1\right) \to 1 \tag{11}$$

Proposition 2. The probability of stopping and matching with a higher location are affiliated as both functions are monotonically increasing in signal.

$$\forall a' > a \ \left[\Pr\left(a'\right) | \left(\phi_{stop} = 1\right)\right] > \left[\Pr\left(a'\right) | \left(\phi_{stop} = 0\right)\right] \tag{12}$$

The intuition of these propositions is as follows. You start at some location. You do not know your underlying type, but you know the distribution of types and locations. You know that you are more likely to receive a high signal if you are a high rather than a low type, and you know that higher types get more out of matching with higher locations.

You receive a signal x_1 that falls near the median of the distribution of signals. The signal is insufficiently large to either (a) set the stopping rule to "leave," or (b) if the stopping rule is set to leave, the matching rule points you to the same location $a_{t=1} = a_{t=0}$. Either way you stay at your current location.

In the following period you receive another signal $x_{t=2} > x_{t=1}$. This signal is higher than the previous one — sufficiently high that your stopping rule tells you to leave. Your matching rule has incorporated these signals, and points you to match with a higher location $a_2 > a_1$. You move to this new location and receive a new draw. These decisions are based on your beliefs over your own type as induced by the sequence of signals that you receive.

Inverting this story we describe an individual that makes decisions that are *ex-ante* rational but are *ex-post* suboptimal. If you receive low signals then you are less likely to migrate (via your stopping rule) and less likely to match with a higher location (via your matching rule).

Proposition 3. Observed utility may be misleading over finite periods of time, and increased levels of noise may further reduce the likelihood of average observed utility falling within some bound of its true match value. Any finite realization of signals and utility draws produces a biased estimation of type.

$$\underset{t\to\text{inf}}{p\text{-lim}}\, \bar{U}_{a,\omega} = \mu_{a,\omega} \tag{13}$$

Equation 13 is given by the weak law of large numbers. Define the probability of an average of utility draws being an arbitrary distance from the true location and type match value:

$$\Delta U \equiv \Pr\left(\mid \bar{U}_{a,\omega} - \mu_{a,\omega} \mid \geq \xi\right) \tag{14}$$

Then the probabilistic difference between location average utility is non-decreasing in the noise of the signal:

$$\frac{\partial \Delta U}{\partial \sigma^2} \geq 0 \tag{15}$$

Proposition 4. For $\epsilon \sim F(\mu), \sigma > 0$ in the case where $F(\cdot)$ is unimodal then it is also true that the probabilistic difference between location average utility is non-increasing in time:

$$\frac{\partial \Delta U}{\partial t} \leq 0 \tag{16}$$

Proposition 5. If there exists a sequence of utility draws that is minimally sufficient to identify an

individual's type then every subset of that sequence of draws is insufficient to identify an individual's type.

$$\exists \mathbf{U}_T = \{U_t\}_{t=1}^{T} \implies \omega \Rightarrow \forall \mathbf{U}_S \subset \mathbf{U}_T \not\Longleftrightarrow \omega \tag{17}$$

3.6 Theory Conclusion

The model is based on Lehmann [54], Athey [10, 9], and Athey and Levin [11] in their development of a method to order distributions and solve monotone decision problems under uncertainty. Their research is focused on decision processes with payoffs that exhibit increasing differences. The framework is useful for constructing a theory of migration in which returns to migration are increasing in type ordering, yet the decision rules are largely governed by signals. Counterfactual migration trajectories in the data point to unrealized gains, which are a combination of place and type-specific factors, and the fact that these gains go unrealized at the individual level suggests a set of signals affecting decision-making. The value of using monotone comparative statistics is that we avoid dependence on a specific utility representation. While many of these predictions would hold with commonly used utility functions, we instead create a more general structure while maintaining predictive power. This is useful because we are not using this model to recover point-estimates of signals or propensities, but rather as a tool to understand the factors that produce a *propensity* to migrate and a propensity for types of matching.

I divide the migration decision into two separate decisions. The choice to migrate is captured using a stopping rule. An optimal stopping rule minimizes expected loss over time, and using the ordering definitions from Lehmann [54] I can establish the monotonicity of the stopping rule as a function of signal. Separate from the stopping rule there is the matching rule. Once the individual has decided to migrate their choice of location will be monotonically increasing as a function of the signals they have received, a theoretic structure detailed in Athey [9] and Athey and Levin [11]. Again, this model predicts that higher matches are generated by higher value signals given the affiliated nature of signal and type.

The purpose of this model of migration is to generate outcomes similar to those we observe while minimizing extraneous assumptions. Returning to the high cost estimates in Kennan and Walker [50],

we can also understand those values as forgone welfare. The model here allows us to view actors as rational and with homogeneous preferences, yet based on their private information set — the signals they receive — they may rationally choose locations in such a way to produce these high residual costs.

4 Dynamic Allocation Index Problems

I define an optimal solution to the stopping time and matching decision rules as one that minimizes the loss function in Equation 10. However, this definition does not in of itself provide a method for arriving at a rule. Given the dynamic nature of the problem, where outcomes in one period affect subsequent decisions, there are three possible ways to derive a functional solution:

- Dynamic programming
- Gittins Index
- Bandit algorithms

Dynamic programming and the Gittins index are inappropriate methods in the context of migration, while bandit algorithms provide a tractable method of arriving at an approximate solution to an optimal stopping time and matching rule. The primary concern with dynamic programming solutions is that their computation requires information, in the context of my model, that the individual does not possess. Knowledge of the dynamics of the state-space is necessary to derive an optimal policy function, but assuming this knowledge is at odds with my description of the problem as characterized by uncertainty. A Gittins index allows for the derivation of an optimal policy despite uncertainty of the underlying state-space, yet its calculation is not feasible in cases with sufficient complexity to be realistic. This leaves bandit algorithms as the best method for describing an optimal solution to a dynamic process without prior knowledge of the state-space. Bandit algorithms are computationally simple, and can serve as a functional heuristic to human decision-making.

4.1 Dynamic Programming and Gittins

The traditional solution to sequential problems in the social sciences has been to employ dynamic programming methods, which approximate a solution to the sequential optimization problem using a Bellman Equation. The concern with this approach is that it requires a complete understanding of the underlying state-space. If the transition dynamics of the state-space are known, then an efficient solution can be computed through value function iteration. But if there is uncertainty over the transition probabilities from one state to another, then no solution can be reached; and if the transition probabilities are misspecified then the solution is wrong.

In the case of my migration model, the state-space is the probability distribution over possible types that the individual computes. This state evolves in a Markov fashion, in that the a function of all previous draws of signal and utility is sufficient for the entire vector. Yet the evolution of this state, which is to say the probability of future draws conditional on past draws, is the fundamental uncertainty faced by the individual making the migration decision.

Sutton and Barto [85] write that "if the environment's dynamics are completely known... in principle, its solution is a straightforward, if tedious, computation." Because dynamic programming requires computing expectations over all possible next states and rewards, a distribution model is necessary. Any model, however, may be sensitive to misspecification in transition probabilities due to sensitivity in the optimal solution to these probabilities. The result is a lack of robustness in the solution concept for what is already a highly reduced-form theoretical model of decision-making.

In taking a dynamic programming solution to the data further *a priori* assumptions over agents' preferences are necessary. Rust writes "if we are unwilling to make any parametric functional form assumptions about preferences or beliefs, then in general there are infinitely many different structures in more direct terms, there are many different ways to rationalize any observed pattern of behavior as being "optimal" for different configurations of preferences and beliefs."[76] Rust goes on to cite Ledyard's result that there exists some set of preferences and beliefs that make any undominated strategy optimal. The result is that solutions to dynamic programming problems are optimal by definition and not by any empirical test.[11] This highlights the limitations of optimality, as it exists

[11] See Ledyard [53] for the full proof.

in Kennan and Walker [50], where observed migration trajectories are optimal by definition, yet achieve measurably less welfare than counterfactual trajectories. A nuisance term must then be added to the result to preserve optimality. This term must grow arbitrarily large as the distance between observed and counterfactual optimal trajectories increases. These features arise because under dynamic programming solutions there are no mistakes, and learning does not take place.

The first of these is a direct result of the optimality of the underlying policy function. In a dynamic programming solution, an optimal policy is being implemented at all times by definition, so that other idiosyncratic terms must mechanically force the trajectory to be optimal. This optimality by assumption is unrealistic in a decision process like migration where mistakes are surely made. The second of these implies that the decision-maker does not learn from experience. Their decisions are the result of an optimal policy rule, and as such any stochastic utility or welfare realizations do not feed back in to the actor's decision-making.

The notion of perfect lookahead with no learning in dynamic programming is somewhat at odds with human decision-making process. The lack of learning or mistakes — conditional on perfect information over transition dynamics — cannot effectively describe a dynamic decision process like migration where dynamics are relatively uncertain, learning should occur, and mistakes may be made. In a dynamic programming world an optimal solution rule for my model:

$$\begin{cases} \text{If } a_{t=0} = a^* & \text{stop} \\ \text{Else} & a_{t=1} = a^* \end{cases} \tag{18}$$

This is to say that if an individual is at their optimal location they should remain there, and otherwise they should move to that location and stop. This decision rule is not consistent with observed mitigation trajectories. Individuals make more than one move in their lives, mistakes are made, and learning hopefully occurs.

Preserving the uncertainty around the transition dynamics of the state-space is characteristic of a dynamic index allocation problem. Gittins [44] offers a optimal solution to in the form of an index.

For an infinite-horizon discounted-reward Markov decision problem, an optimal policy can be defined:

$$F(x) = \max_{i:\in\{1,\ldots,n} \left\{ r_i(x) + \beta \sum_{y \in E_i} P_i(x_i, y) F(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n) \right\} \tag{19}$$

Where x is the state variable, and the policy selects among n choices. The Gittins index arrives at the optimal solution for the policy:

$$G(x) = \sup_{\tau>0} \frac{E\left[\sum_{t=0}^{\tau-1} \beta^t r(x(t)) \mid x(0) = x\right]}{E\left[\sum_{t=0}^{\tau-1} \beta^t \mid x(0) = x\right]} \tag{20}$$

Any policy that sequentially plays the largest index arrives at an optimal solution. There are some issues however: the state space must be discrete, discounting must be geometric, and computation time is $O\left(T^3\right)$. It is the computational cost that makes the Gittins index infeasible except in circumstances of simple distributions in a finite state-space. As such, it is elegant but largely remains a theoretical construct.

4.2 Optimality and Regret

Lai and Robbins [52] and Lai [51] which define *asymptotically* efficient allocation rules that could, over repeated play, achieve a solution equivalent to the Gittins index. What's notable about their solution is that while it does not produce an optimal trajectory, i.e. a sequence of choices such that there is no sequence with a higher payoff, it does minimize *regret*. Regret is defined as the distance between an optimal trajectory and any sequence of choices. Uncertainty over choice outcomes makes the realization of a zero regret strategy unlikely, but a bandit algorithm as a class of asymptotically efficient allocation rules, achieves this zero regret asymptotically.

The concept of *regret* originates in the decision making and operations literature. Regret as a concept has multiple definitions, but is best expressed as the difference between the value of the choices made and the value that an oracle would have realized having complete information over the dynamic problem. Bubeck and Cesa-Bianchi [21] define expected regret as the expected difference between the sum of rewards a given policy would generate $\sum r_t$, and the maximum realized rewards

over that same horizon for choices $i \in K$.

$$E\text{Regret} = E\left[\max_{i=1,\dots,K} \sum R_{i,t} - \sum r_t\right] \tag{21}$$

Pseudo-regret is a weaker but more useful notion of the difference *in expectation* between a sequence of optimal choices and those selected by a policy. For an arbitrary set of choices with reward distributions possessing a central moment such that $E[F_i] = \mu_i$, allow there to exist an option such that $\mu_i \geq \mu_k \forall k \neq i$. This is to say that one choice provides the highest return in expectation. Minimizing pseudo-regret then requires a policy that selects this option with the highest expected return.

$$\text{Pseudo Regret} = \max_{i=1,\dots,K}\left[\sum R_{i,t} - \sum r_t\right] \tag{22}$$

Any sequence of choices $\{r_1, r_2, \dots, r_T\}$ can be compared to the mean payoff from the optimal choice: $T \cdot E[r^*]$. As the optimality of the sequence of choice improves, this metric shrinks towards zero. This value can be calculated simultaneously $E[r^*] - r$, cumulatively $kE[r^*] - \sum_{t=1}^{k} r_t$, and the cumulative total at the end of any finite time sequence $TE[r^*] - \sum_{t=1}^{T} r_t$. In what follows, minimizing pseudo-regret will be the primary object of interest.

Lai and Robbins demonstrate that when following a regret minimizing strategy, the expected number of rounds at a suboptimal location within a given duration is:

$$E\left[n_a^T\right] \geq \left(\frac{1}{D(p_i||p_{i^*} + o(1))}\right)\log(T) \tag{23}$$

Where n_a^T is a random variable defined as the number of rounds spent at $a \neq a^*$ within a period of T, and the denominator is the Kullback-Leibler divergence between the density of the distribution of a and the density of the optimal location a^*. Sub-optimal choices are then necessarily selected with certainty, and any *consistent* optimal strategy is one that selects those choices the least — a bound defined by a logarithmic function of time, where regret grows at least logarithmically in time

according to:

$$R \propto \left[\sum_{a \neq a^*} \left(E[a^* - a] \cdot \left(\int f_a \log \frac{f_a}{f_{a^*}} \right)^{-1} \right) \right] \log T \tag{24}$$

Auer et al. [13, p. 73] proves regret bounds where the distance in reward distribution between locations a and a^* is small or otherwise poorly defined. With no statistical assumptions being made over the generation of rewards, a lower bound on pseudo-regret is

$$\Omega\left(\sqrt{KT}\right) \tag{25}$$

And an achievable regret bound is:[12]

$$\Theta\left(\sqrt{T}\right) \tag{26}$$

Therefore, under any optimal strategy regret grows logarithmically, with a defined lower bound of $\Omega(\log T)$. The definition assumes that mistakes are made in the sense that non-optimal choices will be chosen; this is necessary in fact to discover the non-optimality of those choices, a fact that cannot be known *a-priori*! The goal of an algorithm then is simply to minimize the time necessary to determine the suboptimality of any non-optimal choice.

4.3 Optimistic Algorithms

One of the largest families of bandit algorithms is the upper confidence bound (UCB) algorithms that make decisions optimistically when faced with uncertainty. This optimism yields payoffs which achieve logarithmic regret uniformly over time — not just asymptotically. This strategy implements the decision rule for choices $i \in K$:

$$i = \arg\max_{i \in K} \frac{\sum_i^K r_i}{n_i} + P_i \tag{27}$$

[12]The lower bound is $\frac{1}{20}\sqrt{KT}$. Further reading includes 3.3 in Bubeck and Cesa-Bianchi [21]. Dekel et al. [34, p. 14] shows for bandit algorithms with switching costs a lower bound of $\tilde{\Theta}\left(T^{\frac{2}{3}}\right)$ or $\frac{k^{1/3}T^{\frac{2}{3}}}{50\log_2 T}$. However, this paper will stick with the more general result of Equation 26 to evaluating human performance.

Where P_i is a “padding function" designed to optimistically approximate the uncertainty about the potential upper reward bound that arm i could offer.

Developed by Auer, Cesa-Bianchi, and Fischer [12] the UCB1 strategy sets $P_i = \sqrt{\frac{2\log t}{n_i}}$, which produces a uniform regret bound:

$$8\left[\sum_{i:\mu_i<\mu^*}\left(\frac{\log t}{\Delta_i}\right)\right]+\left(1+\frac{\pi^2}{3}\right)\left(\sum_{i=1}^{K}\Delta_i\right) \tag{28}$$

There are a significant number of extensions to the UCB1 strategy, mostly involving tuning it to various contexts. Regardless, the principle remains the same, where an optimistic vector of choices given observed payoffs can produce a regret-minimizing decision rule.

4.4 lil’UCB

Jamieson et al. [47] propose a UCB-type algorithm based on the law of the iterated logarithm such that the number of samples required to identify a best arm is within a constant factor.

The decision rule defines the arm-specific optimism P_{it} as:

$$P \equiv (1+\beta)(1+\sqrt{\epsilon})\sqrt{\frac{2\sigma^2(1+\epsilon)\log\left(\frac{\log((1+\epsilon)n_i}{\delta}\right)}{n_i}}; \delta>0, \epsilon>0, \lambda>0, \beta>0 \tag{29}$$

The double logarithmic factor is necessary to ensure that the probability the best arm is not selected falls below a constant confidence level δ. The lil’UCB algorithm achieves bounded regret then with a high confidence level, making it the best performing member of the UCB family across a wide variety of uses.

4.5 Bandits with Constraints

An important element to the migration decision is the constraint set under which individuals choose their locations. It easy to imagine an algorithm readily switching between wildly different locations, but individuals are less likely to behave similarly in their location choices. I can introduce this behavior

by including a switching cost that is incurred when the bandit chooses a new location.

Srivastava, Reverdy, and Leonard [81] and Reverdy, Srivastava, and Leonard [68] develop the Bounded Upper Credible Limit (B-UCL) to minimize cumulative regret in a bandit problem with switching costs. By extending a Bayesian UCB algorithm (UCL) to make block allocations — sequences in which the same choice is made repeatedly — increasing the length of blocks allows for the minimization of moving costs.

B-UCL works by dividing the total amount of time into frames and further subdividing those frames into blocks. The number of choice instances are allocated into frames $\{f_k\}_{k\in N} : f_k \subseteq \left\{2^{k-1}, \ldots, 2^k - 1\right\}$. Therefore the length of each frame f_k is 2^{k-1}.

The frame is then further subdivided into blocks where the same choice is made throughout. Each frame f_k has $b_k = \frac{2^{k-1}}{k}$ blocks. Over the entire history k denotes the frame and r denotes the block with the tuple $(k, r) \mid k \in \{1, \ldots, \ell\}; r \in \{1, \ldots, b_k\}$. At time $t_{k,r}$ B-UCL selects the arm with the highest upper credible limit on the distribution of its rewards, and chooses it k times in a row.

B-UCL by itself achieves an expected cumulative regret

$$R[B-UCL] \leq \log(T) + o(\log(T)) \tag{30}$$

The takeaway is that B-UCL retains the same optimism that powers UCB, but has a clear method for minimizing switching costs by increasing the duration it stays at any one location.

4.6 Human Analogue

Bandit algorithms provide a close analogue to human decision-making. The algorithms make mistakes, as a sub-optimal choice must be chosen a minimum number of times before it's established as sub-optimal. The algorithm also learns about the payoff distribution of the available choices through experience. This is different from dynamic programming solutions in which the optimal policy rule is computed ahead of time; as a result no learning takes place as the state-transition matrix is already known, and no mistakes are ever made.

Dynamic allocation problems such as an optimal stopping time problem or a dynamic matching problem are interesting because under uncertainty an optimal trajectory cannot be computed. The

Gittins index provides a theoretical solution to these types of problems, but for any problem sufficiently complex as to be interesting, it cannot be computed. Likewise, in the case where the state-transition matrix is unknown, and therefore the underlying uncertainty over choices can be computed, then a dynamic programming solution is optimal. Yet this only holds for cases in which perfect lookahead is possible, which is not a useful method of characterizing many real world problems.

Therefore, evaluating whether an individual solution to a given dynamic allocation problem requires an appropriate benchmark. Comparing almost any realized trajectory against a perfectly optimal solution is unsatisfying as people lack the perfect lookahead and information over complex probability distributions and state changes to possibly compute an optimal trajectory. Srivastava, Reverdy, and Leonard [81] argue then that the effectiveness of any sequential decision problem should be measured by comparing those decisions to what a multi-armed bandit algorithm would have reached.[13]

5 Lab Application

My theoretical model makes a number of predictions. The first is that individuals make optimal decisions conditional on their information set. The second is that the way in which individuals update their information set and construct a decision rule is analogous to a bandit algorithm. The third is that individuals are information-constrained such that the addition of extra information produces a significant improvement to their optimization behavior.

In order to test these three predictions I construct a multi-armed bandit simulation game for participants to play. The game is designed for players to make a series of discrete location decisions. At the beginning of each game a random location is assigned as the optimal location, and the player attempts to find this location in the smallest number of turns. Each turn they receive feedback as the result of a randomly drawn reward pulled from the underlying distribution at their current location.

[13] It is important to emphasize that bandits primarily serve as an *approximation* to human decision-making and a reasonable benchmark when evaluating any sequence of decisions. The work by Daw et al. [33] is an early attempt to match human behavior to bandit algorithms. Using a 4-armed bandit game with fourteen volunteers they found that softmax and epsilon-greedy algorithms most closely matched the sequence of play by humans. Work by Steyvers, Lee, and Wagenmakers [82] similarly finds softmax the best analogue to human decision making, although notes the disparity between bandit rewards and human decisions. Cohen, McClure, and Angela [31] further confirms the close approximation of human decision making with simple bandit strategies.

Table 1: Number of Games

	Control	Treatment
Level 1	2	2
Level 2	13	5
Level 3	0	8

As the player moves closer to their optimal location the mean of this distribution increases.

The player's task is to maximize the number of points they received over a fixed number of rounds. The player is allowed to come up with their own strategies to achieve this goal, which may include a relatively sedentary accumulation of points or a more active search process. Over a number of games the player will improve on their strategy, typically moving from one involving significant randomization, to a more measured approach.

5.1 Design

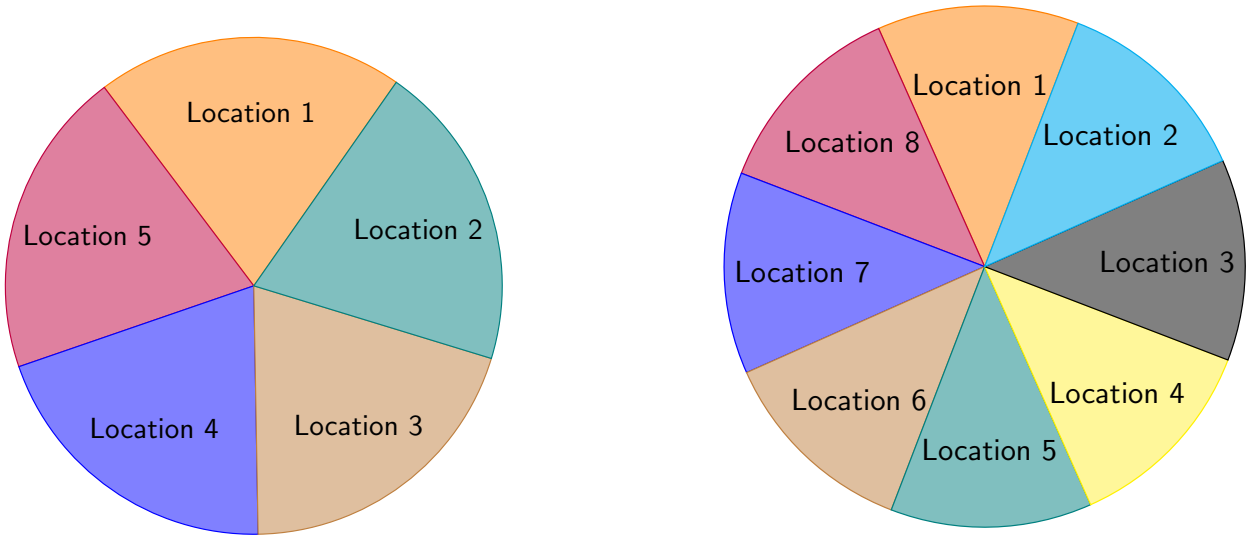
The setup of the simulation is similar to previous multi-armed bandit studies. There are two modes, a starting mode and a more complex mode, each offering 5 and 8 location choices and 2 and 3 sub-choices, respectively. The player selects a location and a sub-choice[14] and then receives a reward drawn from a distribution defined by the intersection of those two choices. The further the player is from their optimal location and sub-choice, defined independently, the lower the average of their draw. Each round the player can change their sub-choice, change their location, or neither. As the player makes choices they begin to get a better sense of the underlying distributions.

The participant plays two games at the starting 5 location/2 sub-choice (Level 1), and then plays the remaining games with 8 locations/3 sub-choices (Levels 2 and 3).[15] This is done to give participants a simpler problem to solve initially as a means of reducing overall frustration.

The tradeoff the player must make is between exploration and exploitation. Spending too long at any one location may result in the player failing to discover locations with higher average payoffs. Spending too much time exploring new locations may leave promising locations under-utilized. There is also the possibility to spend more time making tradeoffs on the intensive margin, which is the

[14] Described in the game as a "technology" or "technique."
[15] Level 3 offers the player access to extra information.

Figure 1: Location Placement



Note: Arrangement of locations for each game type. The optimal location is randomized each game, with rewards at the optimal location drawn from a Poisson distribution with the highest mean, and than mean linearly decreasing in either direction. In the first two games the maximum distance from the optimal location is 2, while in the subsequent 13 games that participants would play the maximum location would be 4. Circular distance is used to keep average starting distance from the optimal location consistent across games.

sub-choice common across all location choices, or the extensive margin, which is when a player changes locations.

Distance between locations is defined circularly as shown in Figure 1 so that for games with five and eight locations, the distance between Location 5 and 8 to Location 1 is one, respectively. The reward distribution of each location is distributed Poisson($\lambda = 10$), and the location parameter is reduced by the distance between their optimal location and the current location. For example, players start with a choice over five locations and two sub-locations. At the beginning of the game the optimal location is randomized to location 1. If the player chooses location 3 they draw a reward from the distribution Poisson($\lambda = 10 - 2$), as 2 is the distance between their current location and their optimal location. If they move to location 2 in the subsequent round they draw a reward from the distribution Poisson($\lambda = 10 - 1$).

Sub-choices are denoted $\{A, B, C\}$ with C only an option for games with 8 locations. These are common across all locations, which means that if the expected value of A is greater than the expected

value of B then it is greater at all locations. Formally:

$$E[A \mid \ell_i] > E[B \mid \ell_i] \Rightarrow E[A \mid \ell_j] > E[B \mid \ell_j] \forall i, j \in K \tag{31}$$

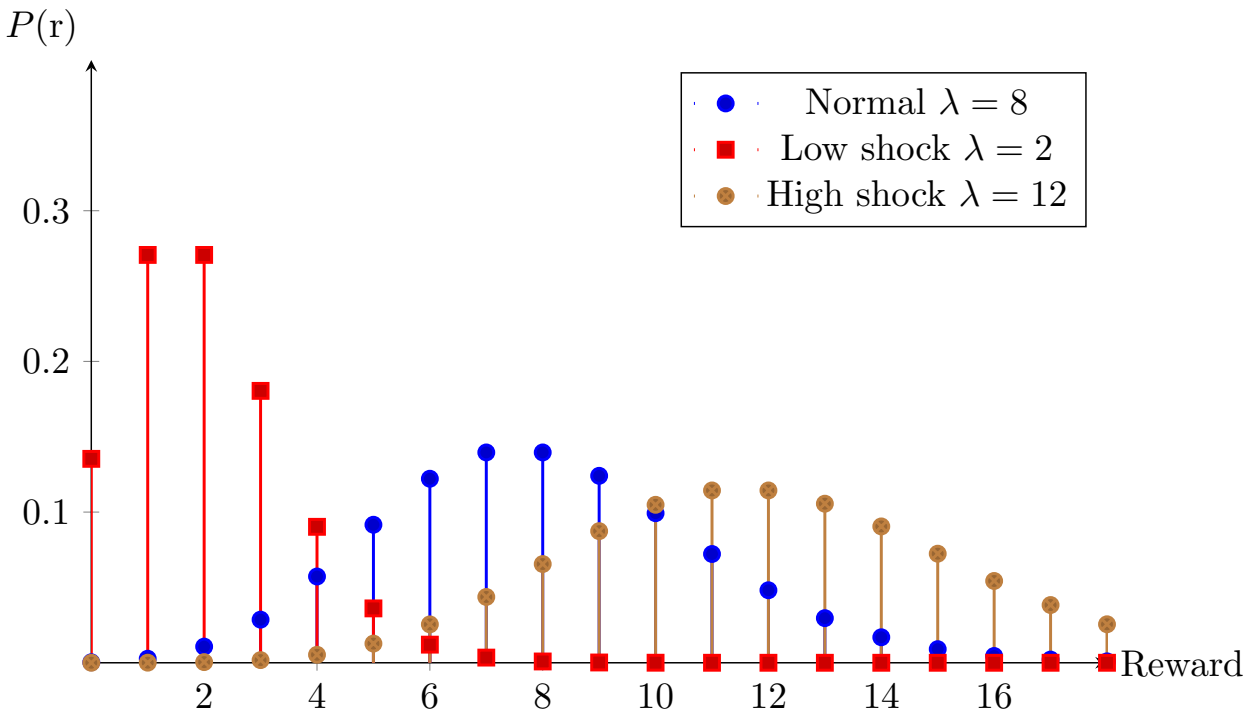
Sub-choices affect the reward distribution location additively, so that choosing the correct sub-choice leaves the location parameter purely as a function of distance, and the sub-optimal sub-choice reduces the location parameter by one. So if the participant is at the optimal location and the optimal sub-choice they would draw from the distribution Poisson($\lambda = 10 - 0$), and if they were at the sub-optimal sub-choice they would draw from Poisson($\lambda = 10 - 1$).

5.1.1 Frictions

The game incorporated two frictions into location choice to more closely represent the costs associated with migration. The first is a fixed movement cost between locations of 10 points. Every time the player selects a new location they pay this upfront cost. The second friction is an opportunity cost, where subsequent rounds spent at the same location increase the value of that draw. Each round the participant observes a reward, but then receives a scaled reward as a function of the duration of time they have spent at that location. This scaling starts out at 50%, increasing 10% a turn to a maximum of 100%. For example, the participant will draw a reward of 8 on the first round — and they will observe this reward — but then their score will only increment by $(8 \cdot 50\%) = 4$ for that turn. If they stay in that location for the next turn, they will then observe another draw from that same distribution, with their received reward incremented by 60% of that value.

5.1.2 Shocks

To match the inclusion of an uncorrelated noise term I added noise into the reward distribution in the form of an uncorrelated low shock and high shock. The player draws from a reward distribution parametrized by their distance from their optimal choice and sub-choice 75% of the time. However, 20% of the time they receive a high shock and draw from the distribution Poisson($\lambda = 12$) and 5% of the time they receive a low shock and draw from the distribution Poisson($\lambda = 2$) regardless of their current location.



Note: Demonstrates the three potential distributions that the player is drawing from each round, with *Normal* arbitarily defined as $\lambda = 8$. In the case where the player was at a location/technology combination with $\lambda = 8$ they would receive a draw from that distribution 75% of the time, 20% of the time from the *high* distribution, and 5% of the time from the *low* distribution, the latter two probabilities invariant across locations.

These shocks were added to represent the innovation term ϵ introduced in Equation 6 to serve as a noise component separate from the reward draws induced by the location and sub-location specific choices.

5.1.3 Encouragement

The primary means of inducing more optimal behavior is through a sliding scale of financial rewards. Participants were given a reward as a function of their average score over fifteen or more games. The lowest payout was $9 USD, increasing to a maximum of $15 USD depending on their average score.[16]

In order to further induce optimal play a graphical interface was constructed to give the player a sense of context for their decisions.[17] Familiar instruments such as buttons, noise feedback, and some data visualization in the terms of bar graphs and density plots were added to further aid participants

[16] In Ethiopia the payouts were scheduled between 50 to 100 Ethiopian birr. This adjustment was done to account for difference in local purchasing power parity.

[17] For work on the psychology of optimization behavior see Green et al. [45] where the use of a explanation of the stochasticity of a reward stream allowed participants to optimize better than in cases where they lacked such an explanation.

Table 2: Payout Menu

Average Grade	Wisconsin (USD)	Ethiopia (ETB)
A+	15	100
A	13	90
B	12	80
C	11	70
D	10	60
F	9	50

Note: 100 Birr ≈ 4.27 USD

in remembering the set of payoffs from their choices.

At the conclusion of the game the player received a letter grade representing the optimality of their performance. The grade presents an easy method of communicating their performance. Furthermore, the grade was adjusted to be fairly harsh: the multi-armed bandit algorithm would receive Bs and Cs most of the time. The intention was to further motivate participants to improve upon their results.

5.2 Treatment

The study incorporates a value of information treatment. Every fifteen rounds the player was allowed to observe a single draw from five locations for all sub-choices for free. Participants did not receive the rewards from these draws, but benefited doubly:

1. By not paying the movement cost to draw from these locations

2. By not having to pay the opportunity cost to draw from these locations

The purpose of the treatment is to get at the actual value of information in terms of its impact on human performance in an information constrained environment. If the cause of sub-optimal play is behavioral then there is no reason to expect extra information to have a significant impact on performance. In that case sub-optimal trajectories are the product of functional errors in optimization. However, if the extra information improves optimization behavior then it demonstrates the importance of information constraints in limiting performance as well as the exact degree to which extra information improves optimization.

Table 3: Participants across Treatment Arms

	Control	Treatment
Wisconsin	129	110
Ethiopia	73	0
Total	202	110

5.3 Protocol

Data was collected over a two week period in Madison, Wisconsin with labs being run on June 24th, 27th, and 28th and then July 8th and 12th 2017. Over these five days I had 228 Wisconsin students participate in the experiment. Later in the summer at Peda Campus in Bahir Dar University I was able to add 79 Ethiopian undergraduate students to the study on August 3rd and 4th 2017.

The lab protocol involved arranging set times for groups of fifteen to twenty-five students to arrive at the computer lab. Research members would sign in participants and give them instruction sheets written in the local language.[18] When all students had entered I would conduct a brief overview of the experiment, emphasizing that their payment would be a function of their overall score. Participants were asked to play a total of fifteen games each consisting of approximately 100 rounds.[19] The first two games are a simple setup of five locations and two sub-choices, the remaining thirteen games had participants making choices over eight locations and three sub-choices. Players were also given a checklist to keep track of the games that they had completed.

6 Data

Participants were undergraduate and graduate students at the University of Wisconsin – Madison and the University of Bahir Dar. The students from Wisconsin were drawn from the SONA system, a participant database maintained by the BRITE Lab. The students in Bahir Dar were largely from the Economics faculty, but as well as other departments in Arts and Sciences. A total of 302 students

[18] English for Wisconsin, Amharic for Ethiopia. See Appendix A for instruction sheets.

[19] The number of rounds was 100 ± 3 to prevent the possibility that a participant might want to solve the dynamic index problem through backward induction. Not knowing when the game would end created a setup more similar to a game with an infinite number of rounds — since the player would not know which round would be their last. Reward parameters were further scaled by a small random factor every game to prevent participants from realizing the true underlying maximum reward distribution of Poisson(10).

participated in the study as detailed in Table 3, with the majority taking part in the control group.

I construct a bandit algorithm using the B-UCL algorithm and lil'UCB algorithm. The B-UCL algorithm controls the location choice and the lil'UCB algorithm controls the the technology choice. As these two decisions are independent both algorithms can function independently, and pass their results to each other.

7 Results

I test the following hypotheses:

H1 Do players follow a bandit or dynamic programming solution to the optimal stopping time and matching problem?

H2 Do players learn over time and update their beliefs to arrive at a regret minimizing solution?

H3 Does a higher level of variance in reward draws reduce overall optimization performance?

In Equation 32 I estimate instantaneous regret using a linear model, where regret in round t is a function of a constant α across participants, a cubic function of round, the game number (1-15), whether the player is in the treatment group $j(i) = 1$ interacted with the type of game the participant is playing (introductory, no information treatment, and information treatment), whether they are playing in Ethiopia $\eta = 1$, and an error term.

$$r_{it} = \alpha + t + t^2 + t^3 + \beta \text{game} + \gamma \left(\tau_{j(i)} \times \text{gametype} \right) + \eta_i + \epsilon_{it} \tag{32}$$

To further increase the robustness of the model I implement it with participant-level fixed effects to account for unobserved individual characteristics θ_i. This however removes time-invariant features, such as being Ethiopian.

$$r_{it} = t + t^2 + t^3 + \beta \text{game} + \gamma \left(\tau_{j(i)} \times \text{gametype} \right) + \eta_i + \theta_i + \epsilon_{it} \tag{33}$$

To account for autocorrelation in the error term I cluster standard errors at the participant level across all rounds and games. Since the game is played independently by all participants and is a single

player game, there is no concern of spillovers between participants or groups. I am interested in four dependent variables: cumulative regret in game, cumulative regret at the end of a game, probability of moving, and number of moves per game.

Figure 2 and 3 show the average instantaneous regret incurred by participants by round and treatment group. It is apparent that instantaneous regret declines within a game, and that it declines at an increasing rate. In 3by round 15, where participants get their first access to extra information we see a sharp decline in instantaneous regret. Eventually the treatment and control groups converge, but not until significantly later in the rounds. It is also important to note that neither group ever achieves an average instantaneous regret near zero, which is what a dynamic programming solution would necessarily achieve.

Figure 4 shows the percentage of players who are at the best location and sub-choice. This value is increasing over time, but also diverges sharply by inclusion in the treatment group. This further demonstrates the learning that takes place within a game, as well as the degree to which information constraints reduce the capacity for players to fully optimize.

The distribution of cumulative regret at the end of a game achieved by participants is better on average than that of the multi-armed bandit algorithm. Figure 5 shows the tight distribution of cumulative regret achieved by the combination of the B-UCL and lil'UCB algorithms in selecting locations, one in which is beaten in expectation by participant performance. This suggests that over short time frames, such as the 100 rounds provided in each game, participants are capable of constructing an optimal sequence of stopping and matching decisions — in comparison to those achievable by an asymptotically efficient allocation rule. The variance of player regret is much higher than what can be achieved algorithmically, and a plot of the empirical cumulative distribution functions in Figure 8 demonstrates second order stochastic dominance by human participants of the bandit ecdf.

8 Conclusion

Migration is the product of decision-making under uncertainty. By decomposing the migration decision into independent stopping time and matching rules I show how uncertainty over type can produce post-facto inefficiencies in location choice while maintaining sequential optimality. The use of bandit

algorithms as a solution concept allows ex-post optimization errors in a way that dynamic programming does not. By relaxing the assumption of perfect knowledge over the evolution of the dynamic system, I show how individuals learn and optimize in an environment characterized by uncertainty. I further demonstrate this behavior in a lab experiment in Wisconsin and Ethiopia. My results demonstrate the surprising effectiveness of this decision-making model in describing how people make sequential decisions under uncertainty, and how migration fits this model as an *increasingly optimal* sequence of decisions.

The model that I propose should be considered minimally sufficient. That is to say that by making use of the work that Lehmann, Athey, and Levin have accomplished on monotone comparative statics under uncertainty, a combination of ordered types and places and a set of signals is sufficient to describe how individuals choose when and where to migrate. This model predicts the importance of accurate signals in allowing individuals to choose their optimal location, and how monotonicty in the stopping and matching rule can endogenously produce results that comport the limited mobility often observed among people who would stand to gain the most by moving. However, the effect of moving costs, start-up costs, and networks play an undeniably important role in structuring migration, and are necessary to apply my model to any context.

Extending this model would involve incorporating a more nuanced understanding of signals and the quality of signals as discussed in Athey and Levin [11] which sets out the conditions under which information is valuable in monotone decision problems. The lab test that I run evaluates the overall impact of additional information on optimization behavior, the *literal* value of information. However, by allowing for a more nuanced signal, a study should investigate the dimensions along which information induces different choices.

This paper further adds to the literature in behavioral economics on using laboratory experiments to test specific microtheoretic predictions. The value of running my study in multiple locations, Madison and Bahir Dar, is to demonstrate the cross-cultural and context validity of the behavior that the model predicts. Optimization behavior varies surprisingly little between undergraduate students in the midwest of the United States and those in northern Ethiopia when tasked with the same optimization under uncertainty test.

Table 4: Distance between participant and bandit reward distributions

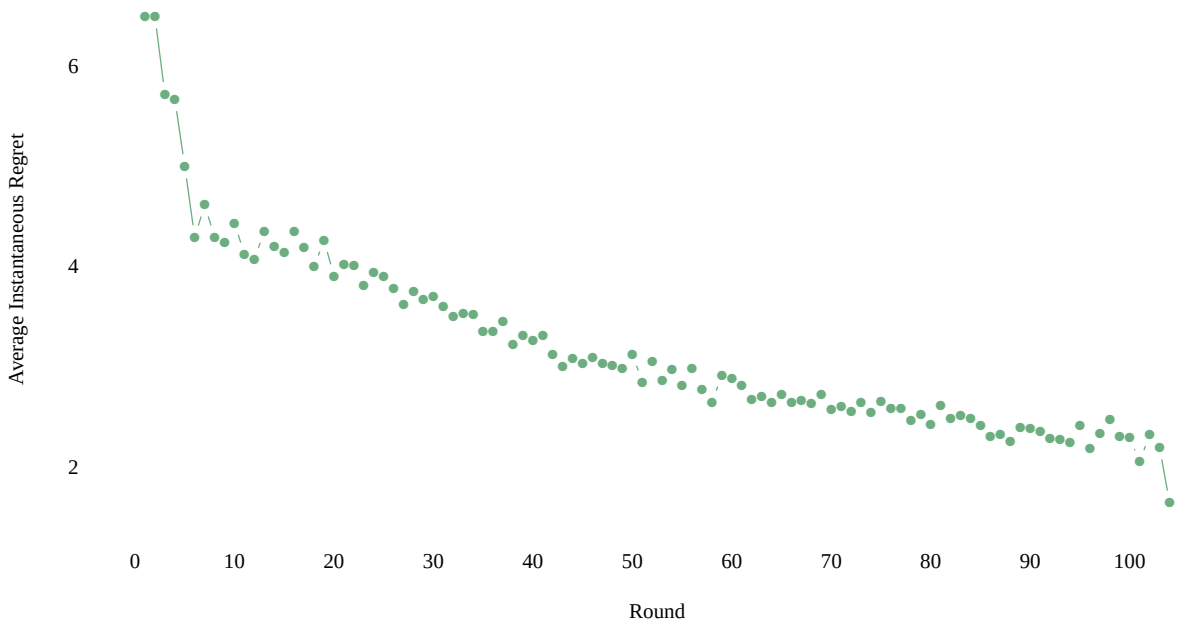
Difference between Bandit and Human Distributions	
Difference in means	-74.06^{***}
CDF overlap	45.43
PDF overlap $\in [0, 1]$	0.50

An evaluation of this work using observational data would require defining a suitably informative signal over type, as well as acquiring information over a sufficient migration history for a large number of individuals. The data demands are high, but this model can offer a direct method of estimating the likelihood of migration without requiring delving into overly-specific demographic data with little predictive power.

Bandit algorithms serve as a suitable analogy for human decision-making under uncertainty, and their use as a benchmark method for optimal decision making may help move economics research away from a focus on idiosyncratic, and therefore largely inexplicable, costs to the specific process by which individuals learn and make decisions. In an environment where decisions are repeatedly made the bandit analogy is especially useful as it provides for a simple learning heuristic with both an update and decision rule.

Ultimately, this paper seeks to understand migration as the product of two simpler dynamic decisions. This decomposition allows me to provide simple rules for optimal behavior, and then test those rules in a lab. The results of my lab test show the validity of bandits as a human analogue, and the crucial role that information constraints play in sequential optimization decisions.

Figure 2: Average Instantaneous Regret by Round



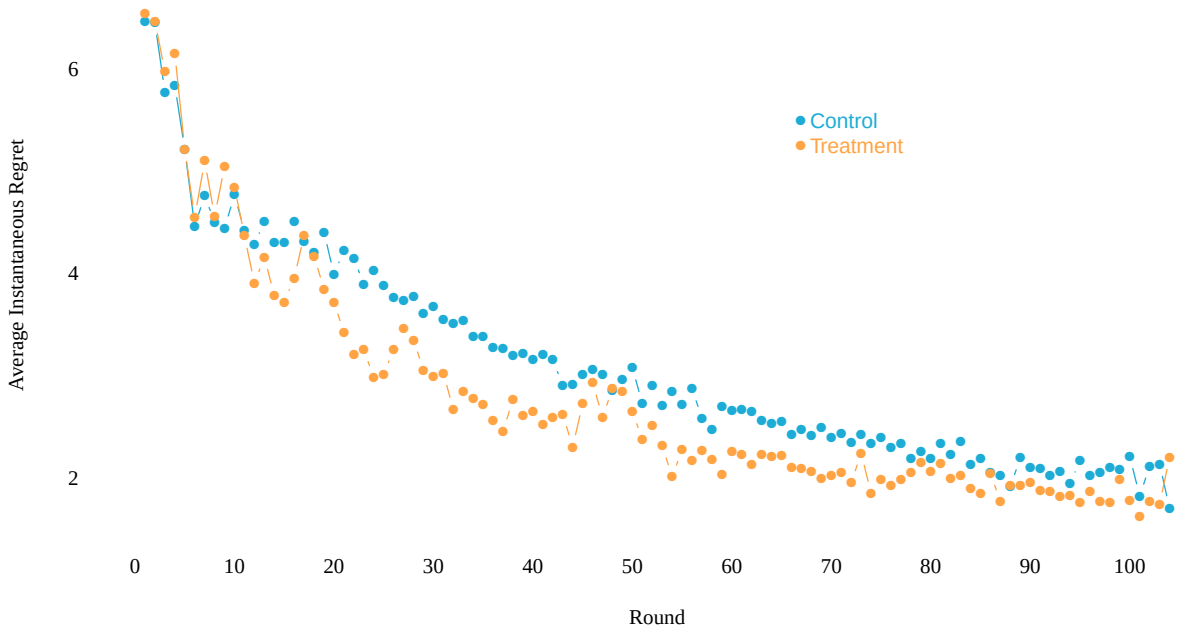
Points represent the average instantaneous regret by participants by round. Instantaneous regret is defined as the difference between the optimal choice mean and the reward received in that round: $\mu^* - r_t$, where the received reward is further modified by moving costs and opportunity costs. The decreasing slope in instantaneous regret shows the increasing optimality of player choices as they converge on the correct location. However, instantaneous regret never reaches an average of zero, suggesting that the average participant never fully optimizes within the time horizon.

Table 5: Cumulative Regret

	Average	Std. Dev.	q25	q50	q75
Participants (control)	333	(177)	208	317	431
Bandit Algorithm (B-UCL + lil'UCB)	415	(68)	368	414	463
T-Test H_A:	$\hat{\mu}_{\text{participants}} \neq \hat{\mu}_{\text{bandit}}$				
$t =$	-73.36				
df $=$	4743				
p-value $<$	$2.2e-16$				

Note: Average cumulative regret between participants and bandit algorithms at the end of the game, where cumulative regret is defined as the sum of the difference between the expected value of the optimal decision and the sum of the rewards received: $T\mu^* - \sum_{t=1}^{T} r_t$. The T-test is run on the difference in means in cumulative regret between the participants without access to the information treatment and the bandit algorithm. The results show that participants scored statistically significantly better on average than the algorithm did over the 100 round timeframe.

Figure 3: Average Instantaneous Regret by Round and Treatment

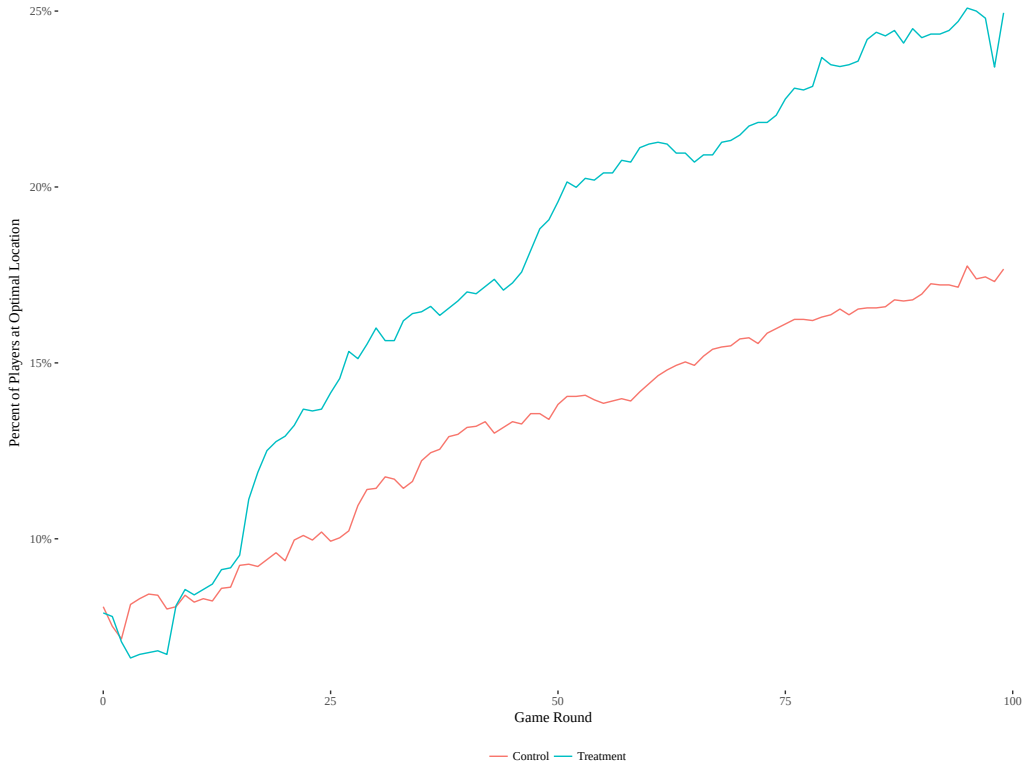


Points represent the average instantaneous regret by treatment and control participants by round. Instantaneous regret is defined as the difference between the optimal choice mean and the reward received in that round: $\mu^* - r_t$, with lower values corresponding to more optimal devisions.The divergence between treatment and control groups by round 15 shows the impact of additional information on optimizing behavior. The convergence in instantaneous regret between the control and treatment by round 80 suggests that participants in the control group eventually gain sufficient information to match the treatment group.

Table 6: Cumulative Regret by Treatment and Region

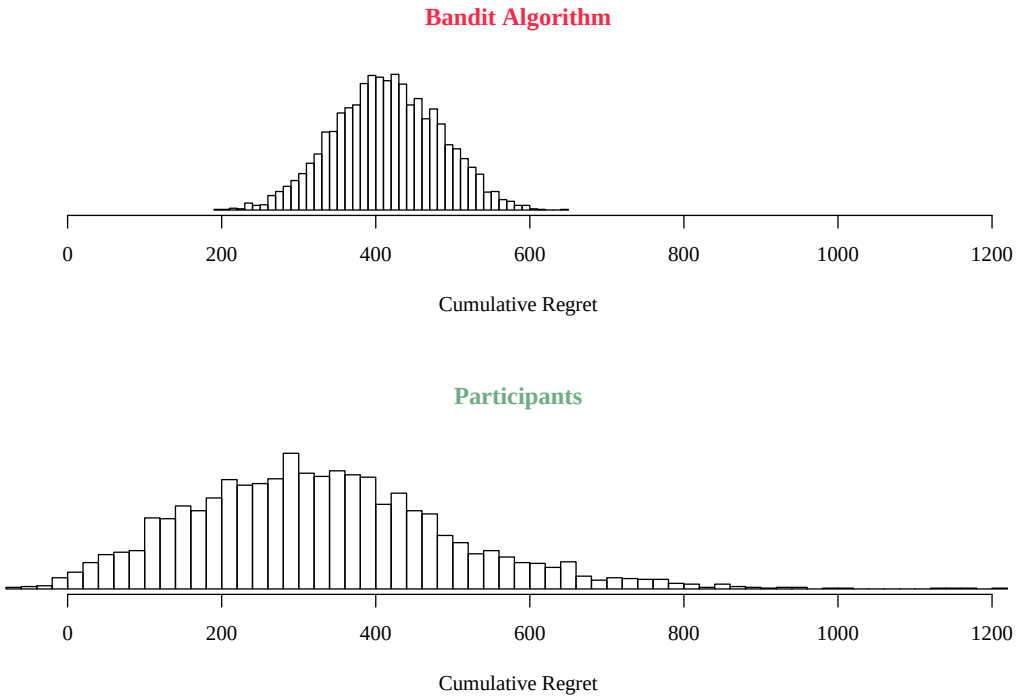
	Average	Std. Dev.	q25	q50	q75
Control (Wisconsin)	330	(166)	218	318	424
Treatment (Wisconsin)	294	(157)	184	274	385
T-Test H_A:	$\hat{\mu}_{\text{control}} \neq \hat{\mu}_{\text{treatment}}$				
$t =$	-72.51				
df $=$	417860				
p-value $<$	$2.2e-16$				
Wisconsin (control)	330	(166)	218	318	424
Ethiopia (control)	341	(207)	186	308	460
T-Test H_A:	$\hat{\mu}_{\text{wisconsin}} \neq \hat{\mu}_{\text{ethiopia}}$				
$t =$	-13.03				
df $=$	107360				
p-value $<$	$2.2e-16$				

Figure 4: Percentage of Participants at the Optimal Location and Sub-Choice by Round



The percentage of participants at the optimal location and sub-choice across all rounds during the experiment. This definition of is fairly restrictrive, only counting cases where the player has selected the optimal location *and* sub-choice. Two things are evident from this graph: first is the increasing probability with which participants are able to make the optimal choice as they accumulate information. Second is the divergence in optimization between the treatment and control gorup. This separation in optimization performance between the treatment and control groups demonstrates the degree to which information constraints bind, and how relaxing them produces an immediate increase in optimization.

Figure 5: Histogram of Cumulative Regret

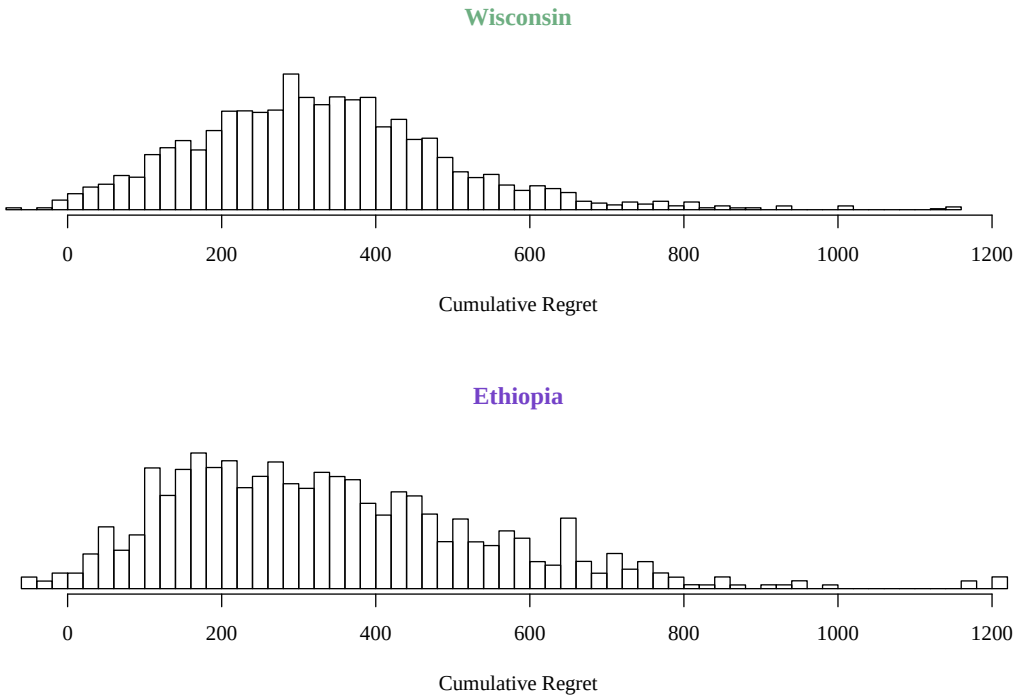


The distribution of cumulative regret between participants and bandits, where cumulative regret is defined as the difference between receiving the optimal average payoff every round and the actual reward received: $T\mu^* - \sum_{t=1}^{T} r_t$. Lower cumulative regret represents more optimal play, with a cumulative regret of zero representing optimal play. The cumulative regret achieved by the bandit algorithm, a combination of B-UCL and lil'UCB, produces a tight regret distribution. In contrast the cumulative regret achieved by participants has a significantly higher variance, but with an average that outperforms the bandit algorithm in this time horizon.

Table 7: Instantaneous Regret by Treatment and Region

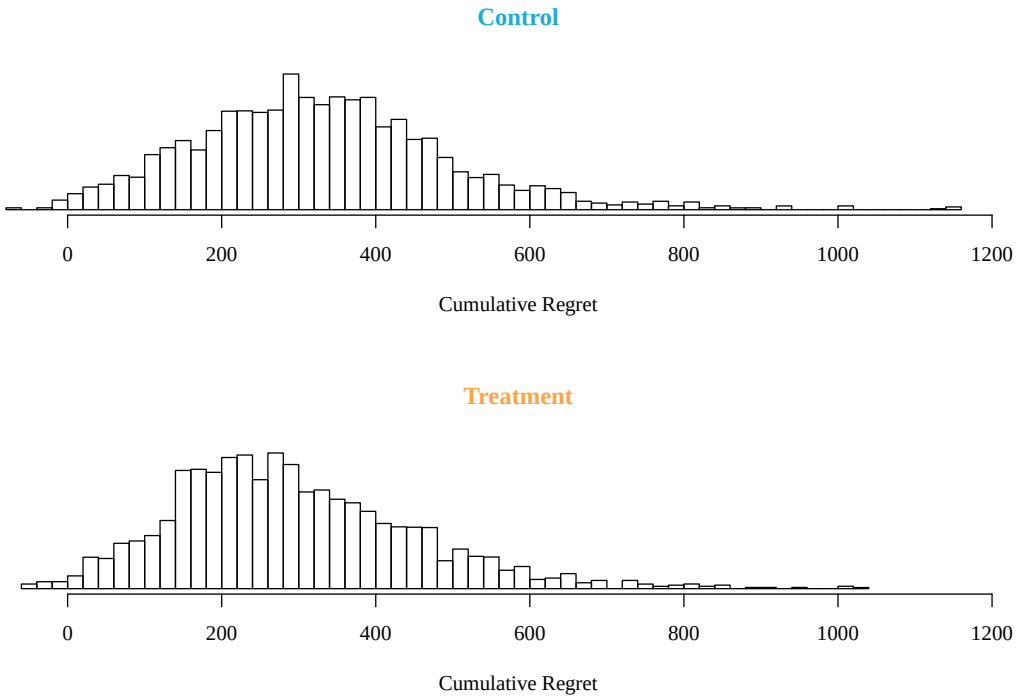
	Average	Std. Dev.	q25	q50	q75
All Control Participants	3.28	(5.03)	0.00	3.07	6.11
Control	3.27	(5.14)	0.00	2.93	6.00
Treatment	2.92	(5.03)	0.00	2.63	5.67
T-Test H_A:	$\hat{\mu}_{\text{control}} \neq \hat{\mu}_{\text{treatment}}$				
$t =$	-21.807				
df =	413570				
p-value <	$2.2e-16$				
Wisconsin	3.27	(5.14)	0.00	2.93	6.00
Ethiopia	3.34	(4.70)	0.00	3.51	6.38
T-Test H_A:	$\hat{\mu}_{\text{wisconsin}} \neq \hat{\mu}_{\text{ethiopia}}$				
$t =$	-3.539				
df =	136470				
p-value <	0.0004018				

Figure 6: Histogram of Cumulative Regret by Region



The distribution of cumulative regret between Wisconsin and Ethiopian participants, where cumulative regret is defined as the difference between receiving the optimal average payoff every round and the actual reward received: $T\mu^* - \sum_{t=1}^{T} r_t$. Lower cumulative regret represents more optimal play, with a cumulative regret of zero representing optimal play. Wisconsin participants on average achieved a statistically significant average lower regret than Ethiopian participants — but only by a very small margin (6.09).

Figure 7: Histogram of Cumulative Regret by Treatment



The distribution of cumulative regret between control and treatment groups in Wisconsin, where cumulative regret is defined as the difference between receiving the optimal average payoff every round and the actual reward received: $T\mu^* - \sum_{t=1}^{T} r_t$. Lower cumulative regret represents more optimal play, with a cumulative regret of zero representing optimal play. Participants in the treatment group with access to more information achieved a statistically significant lower average regret than participants in the control group by 39.69 points.

Figure 8: Empirical CDF of the Distribution of Cumulative Regret

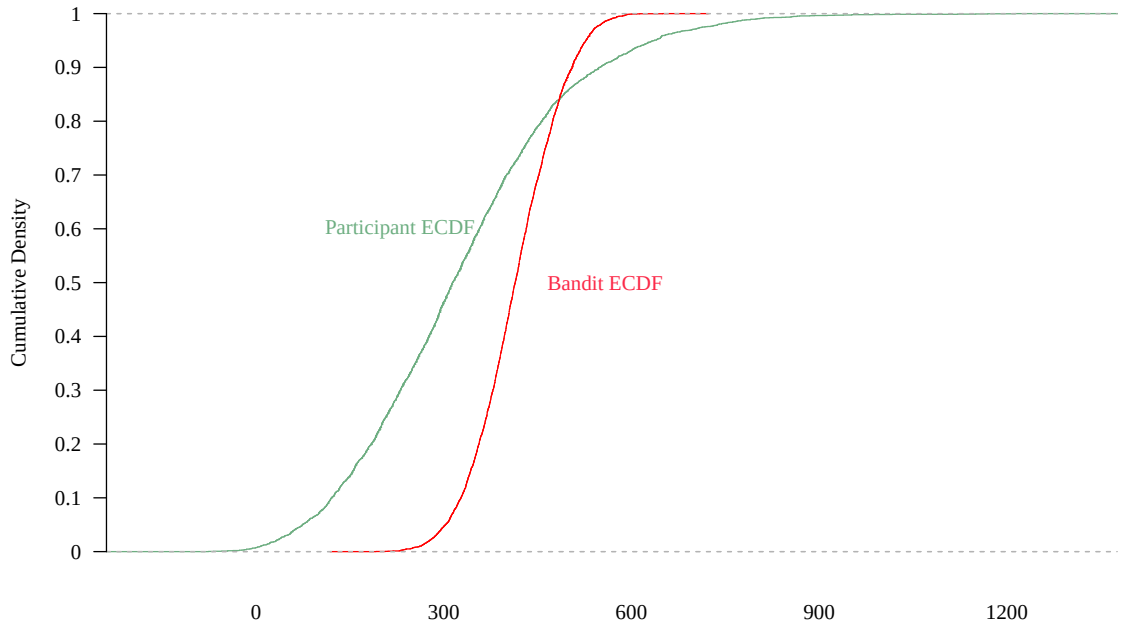


Figure 9: Distribution of Cumulative Regret

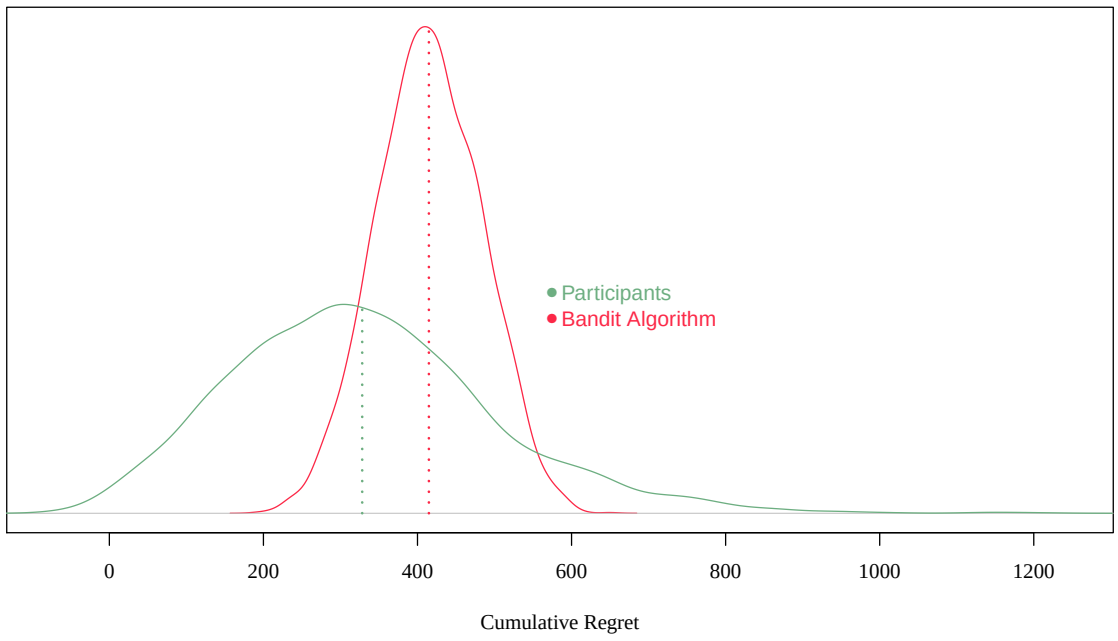


Figure 10: Distribution of Cumulative Regret by Treatment

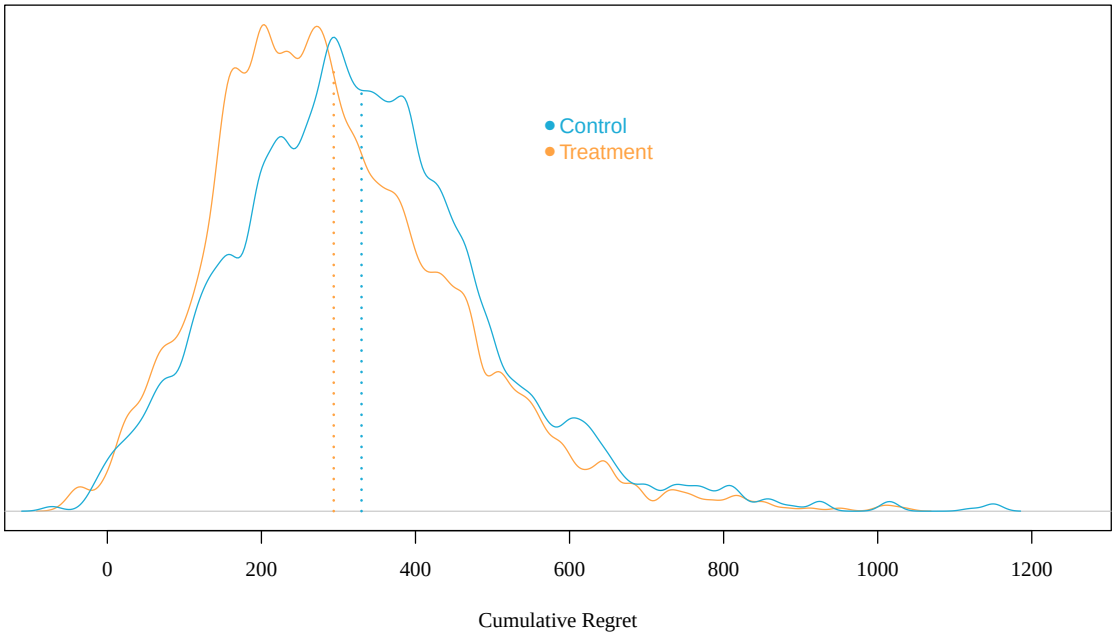


Figure 11: Distribution of Cumulative Regret by Region

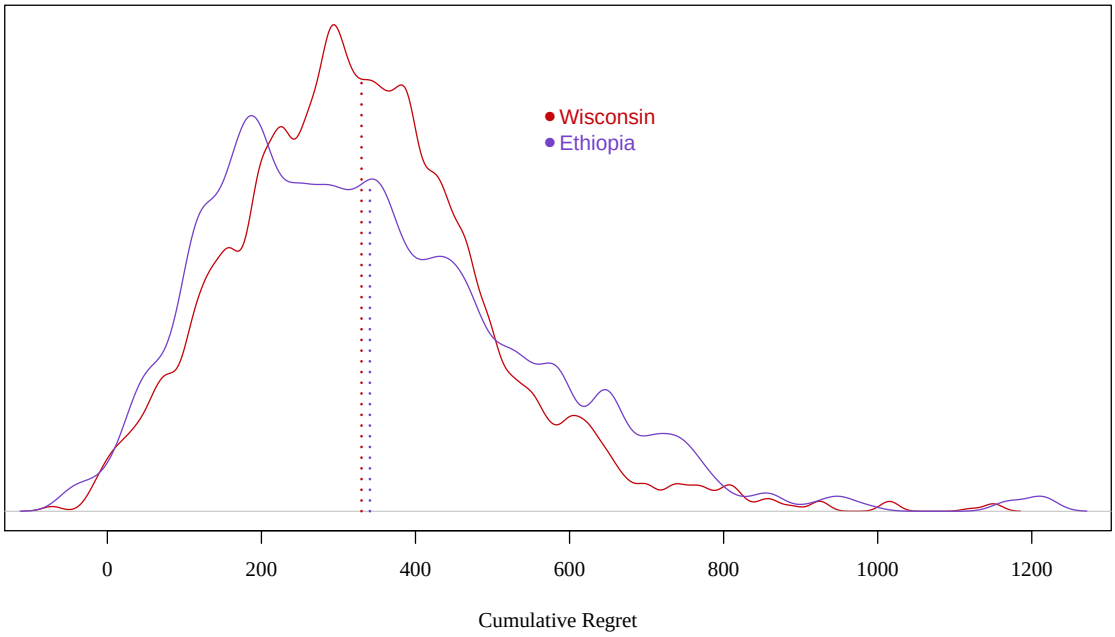


Table 8: Instantaneous Regret

	Model 1	Model 2	Model 3
Intercept	7.64***	7.35***	
	(0.11)	(0.10)	
Time effects			
Ten rounds	−2.35***	−2.35***	−2.35***
	(0.06)	(0.06)	(0.06)
Ten rounds squared	0.39***	0.39***	0.39***
	(0.01)	(0.01)	(0.01)
Ten rounds cubed	−0.02***	−0.02***	−0.02***
	(0.00)	(0.00)	(0.00)
Game number	−0.06***		
	(0.01)		
Ethiopian (0/1)		0.01	
		(0.10)	
Treatment and placebo			
Treatment × Intro		0.11	−1.14
		(0.14)	(0.77)
Treatment × Middle		0.07	−1.19
		(0.11)	(0.76)
Treatment × Treated		−0.76***	−2.08**
		(0.08)	(0.76)
Fixed Effects	No	No	Yes
Adj. R^2	0.06	0.07	0.06
Num. obs.	398389	398389	398389

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered at the individual level.

Note: Linear regression of instantaneous regret as a function of time and treatment. Instantaneous regret is defined as the difference between the expected value of the optimal location less the reward in any round: $\mu^* - r_t$ so smaller values indicate more optimal play. The model is estimated with and without covariates and participant fixed effects. Results show the non-linear effect of additional rounds on optimization behavior, as well as the impact of additional information on average instantaneous regret.

Table 9: Cumulative Regret in Game

	Model 1	Model 2	Model 3
Intercept	46.24***	31.84***	
	(4.05)	(3.03)	
Time effects			
Ten rounds	49.77***	49.75***	49.70***
	(1.27)	(1.27)	(1.26)
Ten rounds squared	−3.06***	−3.05***	−3.04***
	(0.24)	(0.24)	(0.24)
Ten rounds cubed	0.12***	0.12***	0.11***
	(0.01)	(0.01)	(0.01)
Game number	−3.30***		
	(0.50)		
Ethiopian (0/1)		−19.78*	
		(8.02)	
Treatment and placebo			
Treatment × Intro		10.83	−61.64
		(7.47)	(35.49)
Treatment × Middle		−1.24	−69.41
		(6.95)	(36.42)
Treatment × Treated		−42.22***	−111.57**
		(5.94)	(36.51)
Fixed Effects	No	No	Yes
Adj. R^2	0.40	0.40	0.44
Num. obs.	398389	398389	398389

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered at the individual level.

Note: Linear regression of the cumulative regret the player has achieved at every round in the game, where cumulative regret is defined as the difference between the expected value of the optimal choice and the current choice for all periods *up to the current period*. Formally: $t\mu^* - \sum_{t=1}^{t} r_t$. The model is estimated with and without covariates and participant-fixed effects. Standard errors are clustered at the participant level. Results show that cumulative regret is increasing over time, but it does so at a decreasing rate. Being in the treatment group caused a statistically significant lower amount of cumulative regret at any round.

Table 10: Cumulative Regret at the End of Game

	Model 1	Model 2	Model 3
Intercept	392.40***	389.42***	
	(10.79)	(12.71)	
Time effects			
Game number	−14.75***	−12.18***	−11.39***
	(2.86)	(3.12)	(3.38)
Game number squared	0.55**	0.52**	0.49*
	(0.17)	(0.18)	(0.19)
Ethiopian (0/1)		−1.35	
		(10.54)	
Treatment and placebo			
Treatment × Intro		−17.76	−60.79
		(15.14)	(81.00)
Treatment × Middle		2.39	−43.74
		(10.98)	(78.88)
Treatment × Treated		−64.00***	−112.19
		(8.75)	(79.92)
Adj. R^2	0.03	0.05	
Num. obs.	3975	3975	3975
Adj. R^2 (full model)			0.13
Adj. R^2 (proj model)			−0.04

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered at the individual level.

Table 11: Effect of Reward Variance on Cumulative Regret at the End of Game

	Model 1	Model 2	Model 3
Intercept	262.54***	258.58***	
	(10.75)	(12.89)	
Reward variance	7.40***	7.39***	7.42***
	(0.41)	(0.40)	(0.41)
Time effects			
Game number	−12.30***	−9.85***	−9.11**
	(2.63)	(2.89)	(3.12)
Game number squared	0.43**	0.41*	0.38*
	(0.16)	(0.17)	(0.18)
Treatment and placebo			
Treatment × Intro		−17.76	−51.53
		(14.39)	(59.73)
Treatment × Middle		5.48	−31.18
		(10.37)	(57.50)
Treatment × Treated		−62.49***	−101.28
		(8.71)	(57.94)
Ethiopian (0/1)		2.26	
		(9.76)	
Fixed Effects	No	No	Yes
Adj. R^2	0.14	0.17	0.10
Num. obs.	3975	3975	3975

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. Standard errors clustered at the individual level.

Note: Linear model of cumulative regret at the end of a game as a function of reward variance. According to the theoretrical predictions, regret should be increasing in variance of draws. This result is particularly surprising given that for the Poisson distribution mean and variance are given by the same parameter, so that higher rewards should be correlated with higher variances. However, these models show that regret is significantly increasing in variance of the draws, suggesting that players have difficulty identifying the optimal location/sub-choice.

8 References

[1] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. “Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match”. In: *The American Economic Review* 99.5 (2009), pp. 1954–1978.

[2] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. “The new york city high school match”. In: *American Economic Review* (2005), pp. 364–367.

[3] Atila Abdulkadiroglu and Tayfun Sönmez. “School choice: A mechanism design approach”. In: *The American Economic Review* 93.3 (2003), pp. 729–747.

[4] Atila Abdulkadiroğlu and Tayfun Sönmez. “House allocation with existing tenants”. In: *Journal of Economic Theory* 88.2 (1999), pp. 233–260.

[5] Atila Abdulkadiroglu et al. *Changing the Boston school choice mechanism*. Tech. rep. National Bureau of Economic Research, 2006.

[6] Atila Abdulkadiroğlu et al. “The Boston public school match”. In: *American Economic Review* (2005), pp. 368–371.

[7] Daniel A Ackerberg and Maristella Botticini. “Endogenous matching and the empirical determinants of contract form”. In: *Journal of Political Economy* 110.3 (2002), pp. 564–591.

[8] Orley Ashenfelter. “Comparing real wage rates”. In: *The American Economic Review* 102.2 (2012), pp. 617–642.

[9] Susan Athey. “Monotone comparative statics under uncertainty”. In: *Quarterly Journal of Economics* (2002), pp. 187–223.

[10] Susan Athey. “Single crossing properties and the existence of pure strategy equilibria in games of incomplete information”. In: *Econometrica* 69.4 (2001), pp. 861–889.

[11] Susan Athey and Jonathan Levin. “The value of information in monotone decision problems”. In: *Research in Economics* (2017).

[12] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. “Finite-time analysis of the multiarmed bandit problem”. In: *Machine Learning* 47.2-3 (2002), pp. 235–256.

[13] Peter Auer et al. “The nonstochastic multiarmed bandit problem”. In: *SIAM Journal on Computing* 32.1 (2002), pp. 48–77.

[14] Emily A Beam, David McKenzie, and Dean Yang. “Unilateral facilitation does not raise international labor migration from the Philippines”. In: *Economic Development and Cultural Change* 64.2 (2016), pp. 323–368.

[15] Gary S Becker. “A theory of marriage: Part I”. In: *Journal of Political Economy* 81.4 (1973), pp. 813–846.

[16] Gary S Becker. “A theory of marriage: Part II”. In: *Journal of Political Economy* 82.2, Part 2 (1974), S11–S26.

[17] Theodore C Bergstrom and Mark Bagnoli. “Courtship as a waiting game”. In: *Journal of Political Economy* 101.1 (1993), pp. 185–202.

[18] James Berkovec and Steven Stern. “Job exit behavior of older men”. In: *Econometrica* (1991), pp. 189–210.

[19] Venkataraman Bhaskar and Ed Hopkins. “Marriage as a Rat Race: Noisy Premarital Investments with Assortative Matching”. In: *Journal of Political Economy* 124.4 (2016), pp. 992–1045.

[20] Gharad Bryan, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. “Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh”. In: *Econometrica* 82.5 (2014), pp. 1671–1748.

[21] Sebastien Bubeck and Nicolo Cesa-Bianchi. “Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems”. In: *Machine Learning* 5.1 (2012), pp. 1–122.

[22] Hector Chade, Jan Eeckhout, and Lones Smith. “Sorting through search and matching models in economics”. In: *Journal of Economic Literature* 55.2 (2017), pp. 493–544.

[23] Hector Chade, Gregory Lewis, and Lones Smith. “Student portfolios and the college admissions problem”. In: *Review of Economic Studies* 81.3 (2014), pp. 971–1002.

[24] Yeon-Koo Che and Youngwoo Koh. “Decentralized college admissions”. In: *Journal of Political Economy* 124.5 (2016), pp. 1295–1338.

[25] Raj Chetty and Nathaniel Hendren. *The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects*. Tech. rep. National Bureau of Economic Research, 2017.

[26] Raj Chetty, Nathaniel Hendren, and Lawrence F Katz. “The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment”. In: *The American Economic Review* 106.4 (2016), pp. 855–902.

[27] Raj Chetty et al. “The fading American dream: Trends in absolute income mobility since 1940”. In: *Science* 356.6336 (2017), pp. 398–406.

[28] Raj Chetty et al. “Where is the land of opportunity? The geography of intergenerational mobility in the United States”. In: *The Quarterly Journal of Economics* 129.4 (2014), pp. 1553–1623.

[29] Michael A Clemens. “Economics and emigration: Trillion-dollar bills on the sidewalk?” In: *The Journal of Economic Perspectives* 25.3 (2011), pp. 83–106.

[30] Michael A Clemens. “Why do programmers earn more in Houston than Hyderabad? Evidence from randomized processing of US visas”. In: *The American Economic Review* 103.3 (2013), pp. 198–202.

[31] Jonathan D Cohen, Samuel M McClure, and J Yu Angela. “Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration”. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 362.1481 (2007), pp. 933–942.

[32] Gordon B Dahl. “Mobility and the return to education: Testing a Roy model with multiple markets”. In: *Econometrica* 70.6 (2002), pp. 2367–2420.

[33] Nathaniel D Daw et al. “Cortical substrates for exploratory decisions in humans”. In: *Nature* 441.7095 (2006), pp. 876–879.

[34] Ofer Dekel et al. “Bandits with switching costs: T 2/3 regret”. In: *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM. 2014, pp. 459–467.

[35] Esther Duflo, Michael Kremer, and Jonathan Robinson. “How high are rates of return to fertilizer? Evidence from field experiments in Kenya”. In: *American Economic Review* 98.2 (2008), pp. 482–488.

[36] Andrew D Foster and Mark R Rosenzweig. “Learning by doing and learning from others: Human capital and technical change in agriculture”. In: *Journal of political Economy* 103.6 (1995), pp. 1176–1209.

[37] Andrew D Foster and Mark R Rosenzweig. “Microeconomics of technology adoption”. In: *Annual Review of Economics* 2.1 (2010), pp. 395–424.

[38] Xavier Gabaix and Augustin Landier. “Why has CEO pay increased so much?” In: *The Quarterly Journal of Economics* 123.1 (2008), pp. 49–100.

[39] David Gale and Lloyd S Shapley. “College admissions and the stability of marriage”. In: *The American Mathematical Monthly* 69.1 (1962), pp. 9–15.

[40] Joshua Hojvat Gallin. “Net migration and state labor market dynamics”. In: *Journal of Labor Economics* 22.1 (2004), pp. 1–21.

[41] Ahu Gemici. “Family migration and labor market outcomes”. 2008.

[42] John Gibson, David McKenzie, and Halahingano Rohorua. “Development impacts of seasonal and temporary migration: A review of evidence from the Pacific and Southeast Asia”. In: *Asia & the Pacific Policy Studies* 1.1 (2014), pp. 18–32.

[43] John Gibson et al. “The Long-term impacts of international migration: Evidence from a lottery”. In: *The World Bank Economic Review* (2017).

[44] John C Gittins. “Bandit processes and dynamic allocation indices”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1979), pp. 148–177.

[45] CS Green et al. “Alterations in choice behavior by manipulations of world model”. In: *Proceedings of the National Academy of Sciences* 107.37 (2010), pp. 16401–16406.

[46] John R Harris and Michael P Todaro. “Migration, unemployment and development: a two-sector analysis”. In: *American Economic Review* (1970), pp. 126–142.

[47] Kevin Jamieson et al. “lil’UCB: An optimal exploration algorithm for multi-armed bandits”. In: *Conference on Learning Theory*. 2014, pp. 423–439.

[48] Michael P Keane and Kenneth I Wolpin. “The career decisions of young men”. In: *Journal of Political Economy* 105.3 (1997), pp. 473–522.

[49] John Kennan and James R. Walker. “Modeling individual migration decisions”. In: *International Handbook on the Economics of Migration*. Ed. by Amelie Constant and Klaus Zimmermann. Edward Elgar, 2013.

[50] John Kennan and James R. Walker. “The effect of expected income on individual migration decisions”. In: *Econometrica* 79.1 (2011), pp. 211–251.

[51] Tze Leung Lai. “Adaptive treatment allocation and the multi-armed bandit problem”. In: *Annals of Statistics* (1987), pp. 1091–1114.

[52] Tze Leung Lai and Herbert Robbins. “Asymptotically efficient adaptive allocation rules”. In: *Advances in applied mathematics* 6.1 (1985), pp. 4–22.

[53] John O Ledyard. “The scope of the hypothesis of Bayesian equilibrium”. In: *Journal of Economic Theory* 39.1 (1986), pp. 59–82.

[54] E.L. Lehmann. “Comparing Location Experiments”. In: *The Annals of Statistics* 16.2 (1988), pp. 521–533.

[55] Steven A Lippman and John McCall. “The economics of job search: A survey”. In: *Economic Inquiry* 14.2 (1976), pp. 155–189.

[56] Brian P McCall and John J McCall. “A sequential study of migration and job search”. In: *Journal of Labor Economics* (1987), pp. 452–476.

[57] David McKenzie and Hillel Rapoport. “Network effects and the dynamics of migration and inequality: theory and evidence from Mexico”. In: *Journal of Development Economics* 84.1 (2007), pp. 1–24.

[58] C Nicholas McKinney, Muriel Niederle, and Alvin E Roth. “The Collapse of a Medical Labor Clearinghouse (and Why Such Failures Are Rare)”. In: *American Economic Review* 95.3 (2005), pp. 878–889.

[59] Mario J Miranda and Gary D Schnitkey. “An empirical model of asset replacement in dairy production”. In: *Journal of Applied Econometrics* 10.S1 (1995).

[60] Kaivan Munshi. “Networks in the modern economy: Mexican migrants in the US labor market”. In: *The Quarterly Journal of Economics* 118.2 (2003), pp. 549–599.

[61] Kaivan Munshi. “Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution”. In: *Journal of Development Economics* 73.1 (2004), pp. 185–213.

[62] Kaivan Munshi and Mark Rosenzweig. “Networks and misallocation: Insurance, migration, and the rural-urban wage gap”. In: *American Economic Review* 106.1 (2016), pp. 46–98.

[63] Kaivan Munshi and Mark Rosenzweig. *Why is mobility in India so low? Social insurance, inequality, and growth*. Tech. rep. National Bureau of Economic Research, 2009.

[64] Muriel Niderle et al. “Matching and market design”. In: *The New Palgrave Dictionary of Economics, 2nd Edition* (2008).

[65] Muriel Niederle and Alvin E Roth. “Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match”. In: *Journal of political Economy* 111.6 (2003), pp. 1342–1352.

[66] Ernest George Ravenstein. “The Laws of Migration”. In: *Journal of the Statistical Society of London* 48.2 (1885), pp. 167–235.

[67] Ernest George Ravenstein. “The Laws of Migration”. In: *Journal of the Royal Statistical Society* 52.2 (1889), pp. 241–305.

[68] Paul B Reverdy, Vaibhav Srivastava, and Naomi Ehrich Leonard. “Modeling human decision making in generalized Gaussian multiarmed bandits”. In: *Proceedings of the IEEE* 102.4 (2014), pp. 544–571.

[69] Frank Riedel. “Optimal stopping with multiple priors”. In: *Econometrica* 77.3 (2009), pp. 857–908.

[70] Mark R Rosenzweig and Kenneth I Wolpin. “Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India”. In: *Journal of Political Economy* 101.2 (1993), pp. 223–244.

[71] Alvin E Roth. “The evolution of the labor market for medical interns and residents: a case study in game theory”. In: *Journal of Political Economy* 92.6 (1984), pp. 991–1016.

[72] Alvin E Roth and Elliott Peranson. “The redesign of the matching market for American physicians: Some engineering aspects of economic design”. In: *The American Economic Review* 89.4 (1999), p. 748.

[73] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. “A kidney exchange clearinghouse in New England”. In: *American Economic Review* (2005), pp. 376–380.

[74] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. “Kidney exchange”. In: *The Quarterly Journal of Economics* 119.2 (2004), pp. 457–488.

[75] Alvin E Roth, Tayfun Sönmez, and M Utku Ünver. “Pairwise kidney exchange”. In: *Journal of Economic theory* 125.2 (2005), pp. 151–188.

[76] John Rust. “Dynamic Programming”. In: *The New Palgrave Dictionary of Economics*. Ed. by Steven Durlauf and Lawrence E. Blume. Palgrave MacMillan, 2008.

[77] John Rust. “Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher”. In: *Econometrica* (1987), pp. 999–1033.

[78] Lloyd S Shapley and Martin Shubik. “The assignment game I: The core”. In: *International Journal of game theory* 1.1 (1971), pp. 111–130.

[79] Lloyd Shapley and Herbert Scarf. “On cores and indivisibility”. In: *Journal of mathematical economics* 1.1 (1974), pp. 23–37.

[80] Filippo Simini et al. “A universal model for mobility and migration patterns”. In: *Nature* 484.7392 (2012), pp. 96–100.

[81] Vaibhav Srivastava, Paul Reverdy, and Naomi E Leonard. “On optimal foraging and multi-armed bandits”. In: *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*. IEEE. 2013, pp. 494–499.

[82] Mark Steyvers, Michael D Lee, and Eric-Jan Wagenmakers. “A Bayesian analysis of human decision-making on bandit problems”. In: *Journal of Mathematical Psychology* 53.3 (2009), pp. 168–179.

[83] George J Stigler. “Information in the labor market”. In: *Journal of political economy* 70.5, Part 2 (1962), pp. 94–105.

[84] James H Stock and David A Wise. “Pensions, the Option Value of Work, and Retirement”. In: *Econometrica* (1990), pp. 1151–1180.

[85] Richard S Sutton and Andrew G Barto. “Reinforcement Learning: an introduction”. 2017.

[86] Mitsushi Tamaki. “Optimal stopping in the parking problem with U-turn”. In: *Journal of Applied Probability* 25.2 (1988), pp. 363–374.

[87] Mitsushi Tamari. “An optimal parking problem”. In: *Journal of Applied Probability* 19.4 (1982), pp. 803–814.

[88] Michael P Todaro. “A model of labor migration and urban unemployment in less developed countries”. In: *American Economic Review* 59.1 (1969), pp. 138–148.

[89] Robert H Topel and Michael P Ward. “Job mobility and the careers of young men”. In: *The Quarterly Journal of Economics* 107.2 (1992), pp. 439–479.

[90] Insan Tunali. “Rationality of migration”. In: *International Economic Review* 41.4 (2000), pp. 893–920.

[91] Kenneth I Wolpin. “An estimable dynamic stochastic model of fertility and child mortality”. In: *Journal of Political Economy* 92.5 (1984), pp. 852–874.

[92] Dean Yang. “International migration, remittances and household investment: Evidence from Philippine migrants’ exchange rate shocks”. In: *The Economic Journal* 118.528 (2008), pp. 591–630.

A Instructions

Instructions:

- Please enter the first part of your *wisc.edu* email address into the userid box.
    - Double-check this, as it is necessary to receive payment for your participation
    - For example "zbarnetthowe" @wisc.edu
- Then select: **GROUP X**
- **First**: Play 2 games at Level 1, after which Level 2 will unlock
- **Then:** Play 13 games at Level 2
- Afterwards you will complete a brief survey

Figure 12: Game Instructions for Ethiopia

<u>መመሪያ</u>

- በመጀመሪያ የሞባይል ቁጥርዎን ወይም የኢሜይል አድራሻዎን የተጠቃሚ መለያ ቁጥር በሚለው ውስጥ ይጻፉ
- በመቀጠል የሚወክሉትን ቡድን ይምረጡ
- ሁለት ጨዋታዎችን በመጀመሪያው የጨዋታ ዙረጃ ከተጫወቱ በኋላ የሚቀጥለው የጨዋታ ዙረጃ ይከፈትልዎታል
- በሁለተኛው ዙረጃ 5 ጨዋታዎችን ይጫዎቱ
- በሶሥተኛው ዙረጃ 8 ጨዋታዎችን ይጫዎቱ
    - ሦስተኛውን ዙረጃ ሲያልፉ ስለፕላኔቶቹ የተሻለ መረጃ ለመፈለግ የሚያስችል ዕድል ያገኛሉ።
- ከዚያ በኋላ የተዘጋጀውን መጠይቅ ይሙሉ

<u>ይህ ሙከራ መረጃንና መረጃ መመምረጥን የሚመለከት ነው።</u>

ዋናው ዓላማ በጨዋታው የተሻለ ውጤት ማግኘት ነው።

ብዙ ነጥብ/ውጤት ባገኙ ቁጥር የሚያገኙት የገንዘብ መጠን ይጨምራል።

- ሁለት ጨዋታዎችን በመጀመሪያው የጨዋታ ዙረጃ ከተጫወቱ በኋላ የሚቀጥለው የጨዋታ ዙረጃ ይከፈትልዎታል። በመጀመሪያው የጨዋታ ዙረጃ ያገኙት አማካይ ዉጤት አይያዝም።
- በሁለተኛው ዙረጃ 13 ጨዋታዎችን ይጫዎታሉ
- ከፈለጉ ከ15 በላይ ጨዋታዎችን መጫወት ይችላሉ።

የሚያገኙት ክፍያ መጠን የሚወሰነው በዙረጃ ሁለት ላይ በሚያገኙት አማካይ ውጤት ይሆናል፤ ያም ማለት ብዙ ነጥብ ባገኙ ቁጥር ብዙ ገንዘብ ያገኛሉ።

- እጅግ በጣም ጥሩ/**Avg. A+** = 100 ብር
- በጣም ጥሩ/**Avg A** = 90 ብር
- ጥሩ/**Avg B** = 80 ብር
- በቂ/**Avg C** = 70 ብር
- ዙካማ/**Avg D** = 60 ብር
- መጥፎ/**Avg F** = 50 ብር

<u>ውድ ተሳታፊ፤ ይህንን ጨዋታ መሳተፍ/መጫዎት ይፈልጋሉ?</u>

ዋና ዓላማዎ አዲሱን ዓለም መቆጣጠር ነው። የተሻሉ ፕላኔቶችን ለመምረጥ በአማካይ 100 ዕድል ይኖርዎታል። ከእነዚህ ፕላኔቶች ውስጥ ውጤታማ የመሆን ዕድልዎ በአንዱ ብቻ ይሆናል። የሰፋሪዎችዎ ቁጥር በበዛ መጠን ወደተሻለው ፕላኔት እየተጠጉ መሆኑን ይወቁ።

የእርስዎ ዓላማ ወደ አዲሱ ፕላኔት የሚሄዱትን ሰፋሪዎች ቁጥር መጨመር ነው። ምንም እንኳ የተወሰነ ዕድል ቢኖርዎትም የሚፈልጉትን ውሳኔ ለመወሰን ግን በቂ ነው።

በጨዋታ ዙረጃ 3 ላይ ስለፕላኔቶች መረጃ ለማግኘት የሚያስችለውን አማራጭ ዙጋግመው መጠቀም ይችላሉ። አማራጩ ስለሌሎች ፕላኔቶች ሁኔታ መረጃ ይሰጥዎታል። አማራጭ ጠቋሚውን መጠቀም ከጀመርንበት ጀምሮ ምን ያህል ዕድል እንደቀረን በሂደት ጠቋሚ ግራፍ ያሳየናል፤ ምክንያቱም እነዚህን አቅጣጫ ጠቋሚዎች አንዴፈለግን ስለማናገኛቸው ነው።

በመጀመሪ በካርታው ላይ በሚገኙት ፕላኔቶች ወይም ቁጥሮች (1፣ 2፣ 3፣ 4፣ ወዘተ) ላይ ክሊክ የሚለውን ቁልፍ በመጫን ፕላኔቶችን ይምረጡ። በመቀጠል የመቆጣጠሪያ ቴክኖሎጅዎን ለመምረጥ ፊዴሎች ("A"፣ "B"፣ ወይም "C") ላይ ክሊክ የሚለውን ቁልፍ ይጫኑ። በመቀጠልም ፕላኔቱ ወይም ቴክኖሎጅው ላይ ክሊክ የሚለውን ቁልፍ በመጫን ይምረጡ። ምን ያህል ውጤታማ ሰፋሪዎችን እንዴላኩ የሚያዩት ባለአረንጓዴ ቀለም ቁጥር ሲወጡ ይሆናል።

እነዚህ ፕላኔቶች ያላቸው አየር ንብረት በጣም አስቸጋሪ በመሆኑ የላኳቸው ሰፋሪዎች ቁጥር የሚወሰነው በእያንዳንዱ ዕድል/ሙከራ ወቅት ፕላኔቶች ላይ ባለው የአየር ንብረት ይሆናል። እያንዳንዱ ጨዋታ ከአዲሱ ፕላኔት/ዓለም እና ከአዲሱ አካባቢ ተጽዕኖ ነጻ ነው።