A Theoretical Model for Grit in Pursuing Ambitious Ends

Avrim Blum[1], Emily Diana[2], Kavya Ravichandran[1], Alexander Williams Tolbert[3]

[1]Toyota Technological Institute at Chicago
[2]Carnegie Mellon University, Tepper School of Business
[3]Emory University, Data and Decision Sciences
avrim@ttic.edu, ediana@andrew.cmu.edu, kavya@ttic.edu, alexander.tolbert@emory.edu

Abstract

Ambition and risk-taking have been heralded as important ways for marginalized communities to get out of cycles of poverty. As a result, educational messaging often encourages individuals to strengthen their personal resolve and develop characteristics such as discipline and grit to succeed in ambitious ends. However, recent work in philosophy and sociology highlights that this messaging often does more harm than good for students in these situations. We study similar questions using a different epistemic approach and in simple theoretical models – we provide a quantitative model of decision-making between stable and risky choices in the *improving multi-armed bandits* framework. We use this model to first study how individuals' "strategies" are affected by their level of grittiness and how this affects their accrued rewards. Then, we study the impact of various interventions, such as increasing grit or providing a financial safety net. Our investigation of rational decision making studies the competitive ratio between the accrued reward and the optimal reward.

Extended version — https://arxiv.org/abs/2503.02952

1 Introduction

Scholars across various fields have long been interested in understanding how humans make decisions between immediate and long-term rewards in the face of uncertainty. Costs associated with a given action can also greatly influence how agents make these decisions, and so we wish to understand how to support exploratory and ambitious behavior, particularly in groups who have historically faced lack of access to such opportunities. One important factor that has received much attention in recent years is *grit*, with researchers across various fields studying the role resilience and optimism play not only in an individual's success but also in lifting disadvantaged communities out of bad circumstances.

While it is difficult to give a single, unifying definition of grit, philosophers Jennifer Morton and Sarah Paul provide a thesis that serves as our guiding qualitative description. They highlight that grit is rational (distinguishing it from delusional optimism) and that it is an outcome of *beliefs* an agent holds about their circumstances. In more detail:

> "Grit is not simply the ability to withstand the pain of effort and setbacks, or to resist the siren song of easier rewards; it is a trait or capacity that consists partly in a kind of epistemic resilience.
>
> This is a descriptive rather than a normative claim, and it has not gone unnoticed by psychologists who study perseverance. Angela Duckworth emphasizes the relevance of hope in underwriting the capacity for grit, where hope is defined as the expectation that one's efforts will pay off. And Martin Seligman touts the importance of optimism, which involves a distinctive style of explaining to oneself why good and bad events happen." (Morton and Paul 2019)

Individuals with grit are able to persevere despite the lack of immediate reward: for instance, they are more willing to practice a math or sport skill until it is mastered, or they are willing to tolerate a significant degree of financial austerity while starting a company until revenue is accrued.

For a long time, prevailing guidance in educational settings focused heavily on personal characteristics such as grit, discipline, and resilience as the way to succeed in ambitious ends (Gibbon 2020), often even at the protest of the scholars whose work was used to justify this perspective (Kamenetz 2015). Subsequent longitudinal work paints a more ambivalent picture, studying outcomes for students from disadvantaged backgrounds who follow those lessons in attempting to succeed at ambitious long-term ends such as college; these works find that simply pushing grit can often have negative effects (Morton and Paul 2019; Wooten 2022). These findings highlight an empirical tension, that grit is useful in pursuing long-term ends that require investment but can make those who cannot cushion early losses susceptible to the precarity associated with such ends. The puzzle posed by these findings posit that grit is not *uniformly* positive but perhaps instead *conditionally* useful. What, then, are the situations in which grit is productive? Harmful?

In this work, we attempt to bridge this analytic gap by isolating where additional grit helps and where it hurts via a simple theoretical quantitative model. We also investigate how outcomes are influenced by having financial supplements. Our goal is to study decision-making dynamics in a controlled, quantitatively-defined setting. By characterizing the landscape of grit and its interplay with external support, we unify the diverse empirical observations in this area and provide a model in which further quantitative analysis can take place. Our work (1) provides further understanding of the relationship between grit and financial support in succeeding when pursuing ambitious outcomes and while doing so, (2) introduces a simple two-armed bandit theoretical model that shows promise as a formalization of a decision-making problem that juxtaposes stable reward against ambition. Our analysis isolates three crucial parameters – level of grit, amount of external support, willingness to tolerate discomfort – and develops an understanding of outcomes as a function of these, reproducing with proof several patterns observed by qualitative scholars.

1.1 Our Approach To Studying Grit

Morton and Paul emphasize that grit leads agents to *rationally* stick with an option that others are not willing to stick with. We consider two formal models of rationality in decision-making. First, we study agents who maximize their competitive ratio. Measuring the ratio between achieved outcome and best possible outcome in hindsight, the competitive ratio being maximized implies that the agent has minimal multiplicative "regret" about the past. Thus, this model of rationality is a backward-looking model. In the extended version of the paper, we also study agents who take a Bayesian perspective to uncertainty quantification. In this view, an agent has an explicit prior probability distribution over the possible outcomes and updates their posterior each time they receive new information. In constructing a prior, the agent explicitly quantifies what they think the future holds, making this a forward-looking model of rationality. We compare and contrast the aspects of grit we are able to study in each of these formal models of rationality and see how the conclusions we draw from each on similar formal models relate.

In order to understand the impact of grit, we study two things – first, we consider the effect of grit on the "policy" or "strategy" that the agent follows, i.e., the actions they take; second, we study what the impact of that "policy" is on the reward they achieve. This modular breakdown will become especially useful when we consider the impact of a trust fund on the actions of an agent, allowing us to disentangle the effect of grit as a trait and other interventions that lead agents to behave similarly to those who are gritty.

1.2 Our Contributions

In this work, we propose studying grit in the *improving multi-armed bandits* framework. We develop an instance that allows us to satisfactorily investigate the impact of grit as a characteristic the agent has, and we show that the instance we propose is the simplest instance in which the strategy is non-trivial. In order to understand strategies resulting from grit, which as discussed above is rational, in this model, we must formalize what we mean by "rational," and we do so by appealing to two notions of rationality well-studied in computer science. In particular, we first consider the competitive ratio, a standard notion in the analysis of online algorithms, which is optimized when a strategy minimizes multiplicative regret in hindsight; in the extended paper, we also study a Bayesian notion of rationality in which an agent has a prior that helps quantify their uncertainty about the future, which they update to a posterior based on evidence from their environment. Between these two notions, we study how gritty behavior is reflected in both forward- and backward- looking formalizations of rationality.

With the model and notions of rationality in hand, we study how grit affects an agent's strategy and how that in turn affects their outcomes. When grit is related to the optimistic outlook of an agent, we can quantitatively derive cases in which grit helps and hurts the agent, showing that an excess of grit harms the agent by causing them to net less reward than their less-gritty counterparts. When grit is reflected in how willing an agent is to tolerate discomfort, we conclude that though agents who require comfort can minimize their multiplicative regret in hindsight pretty easily, agents who can tolerate more discomfort can explore for longer. This section culminates with studying how financial support changes agents' behavior, essentially showing that financial support allows agents to expand their exploration horizon while allowing them to still receive comparable reward to their less gritty counterparts.

The rest of the paper is structured as follows. First, we survey related work from several fields, including philosophy, sociology, and computer science. Then, we formally introduce the multi-armed bandits instance we study. In Section 3, we study the competitive ratio-based notion of rationality, following which in Section 4 we introduce models for financial support and study the effects of the financial safety net on behavior and reward.

1.3 Related Work

First, we discuss work from the social sciences that observes grit in people and analyzes its role and impact in society. One of the most popular studies of grit is psychologist Angela Duckworth's book (Duckworth 2016). Relatedly, psychologist Martin Seligman has a large body of work encouraging positive attitudes as a path to success and good outcomes (Seligman 2002, 2006). More recently, sociologist Tom Wooten studied the impact of the "no excuses" educational approach, which draws on the above ideas, in his dissertation (Wooten 2022), particularly exploring the mechanisms by which these educational systems perpetuate poverty. Abstracting findings from several such observational studies, Morton and Paul develop a philosophical theory of grit (Morton and Paul 2019). We build heavily on the abstractions derived in this work.

Next, we survey computer science literature that we draw on in order to quantitatively study the question of decision-making with grit. Our formalizations of rationality are inspired by objectives that are commonly optimized in the computer science literature, including the competitive ratio studied in online learning (Borodin and El-Yaniv 2005) and the Bayesian approach to uncertainty quantification (Bernardo and Smith 2000). The framework we use to represent the decision problem is an instance of the multi-armed bandits (MAB) problem, a well-studied framework for making decisions when the payoff is unknown (Slivkins 2022). In particular, we suppose the bandit arms have structured reward, and the structure of interest is improving, first formalized by (Heidari, Kearns, and Roth 2016) and studied by (Patil et al. 2023; Blum and Ravichandran 2025).

Key Features of Grit According to Morton and Paul Since we base our development of a theory of grit on the account of (Morton and Paul 2019), it will be helpful to summarize the key points from their work. In their work, Morton and Paul describe gritty behavior as displaying a form of "epistemic resilience." An important part of this is how an agent redefines their goal in the presence of encouraging or discouraging evidence. They also reason that grit is *rational*, and therefore agents displaying grit cannot *ignore* evidence but rather should be sensitive to failure. Citing works from psychologists Angela Duckworth and Martin Seligman, Morton and Paul further highlight the importance of hope and optimism. They culminate by providing a description of an "Evidential Threshold" that captures the decision-making of a gritty agent. The Evidential Threshold as conceptualized by them asks how compelling evidence must be to change the actions of an agent; for a gritty agent, the Evidential Threshold is higher than that of an "impartial observer." Since this threshold hinges on how compelling the agent finds the evidence, Morton and Paul argue that Permissivism applies, and different agents can witness the same evidence but come to different conclusions about what implications that evidence should have on their actions. Throughout our paper, we will connect back to these various facets described by Morton and Paul, including by providing a quantitative analog of the Evidential Threshold.

2 Formal Setting: Improving MAB

We propose studying grit in the improving multi-armed bandits framework (Heidari, Kearns, and Roth 2016). In this setting, there are k actions (modeled as "arms"), and at each time step, the agent chooses one of the arms to "pull," i.e., play. Each arm has associated with it a reward function that increases as a function of the amount of time for which it has been played. This allows us to model situations where engaging with an option increases its payout, for instance when learning a skill or developing a new technology. For most of the paper, we consider a continuous-time version of this framework. Formally:

Definition 2.1. *An instance of the improving multi-armed bandits problem comprises k bandit arms, each of which is associated with a reward function that increases as a function of the (possibly fractional) amount of time for which it has been played.*

In our models for grit, it is natural to think of options having different payoffs, some of which are static over time and some of which take a while to start paying off but then pay off well once they do. Thus, we propose a two-armed bandit instance in which to study gritty behavior. All of our models will include a stable arm, $f_1(t) = 1 \,\forall\, t$ that represents an option that starts paying off immediately and consistently rewards the agent the same amount (i.e., little scope for growth). The other arm is one in which there is no reward at first (or in certain cases where specified, there is actually a *cost* to striving), after which the arm starts providing non-negative reward. We refer to this arm as the "striving" arm, and it provides a reward of 0 units for the first θ time steps that it is played (θ being unknown to the agent[1]), following which it increases linearly at a slope of α (agents will have beliefs about α). Formally, the two bandit arms are:

$$f_1(t_1) = 1 \,\forall\, t \qquad f_2(t_2) = \begin{cases} 0 & t_2 < \theta \\ \alpha(t_2 - \theta) & t_2 \geq \theta \end{cases}, \quad (1)$$

where t represents the amount of time for which that arm has been played.

In most of the paper, we use this model, though in some sections we set $\alpha = 1$. Later on, in Section 3.3, we also define a notion of "comfort" which places a restriction on how often f_2 can be played, requiring that f_1 be played frequently enough to build up a buffer.

This choice of model is natural: the improving multi-armed bandits problem is well-studied, and there is a clear understanding of what we could hope to achieve in the general case (Patil et al. 2023; Blum and Ravichandran 2025). The structure in the reward function allows us to capture the fact that *investing* time into an option may change its payoff. Finally, the instance described is abstract and flexible, allowing us to model a wide range of real-world settings. On the other hand, a limitation is that this instance only allows for studying an agent's decision between two options. The stable option is arguably over-simplified, since stable options can also lead to growth in the real world. Overall, however, we believe this is a good starting point for formally modelling the decision problem of interest. Further, within the improving two-armed bandit setting, this is the simplest model in which we see non-trivial behavior: if the second arm's payoff were flat instead of linear, the trivial strategy of playing f_1 all along would suffice for optimizing the competitive ratio.

3 Rationality in Terms of Competitive Ratio

3.1 Rationality in This Model

The first formal model for rationality we study is one in which an agent minimizes regret in hindsight by optimizing the competitive ratio. Competitive ratio measures the relationship between the achieved reward and the optimal reward. This is a commonly-studied notion in theoretical computer science, and particularly in online algorithms (Borodin and El-Yaniv 2005). Achieving a competitive ratio of 1 ensures that the agent did as well as they could hope. Formally, we define the competitive ratio as follows:

Definition 3.1. *An algorithm achieves* competitive ratio g *if the ratio of its reward ALG to the optimal achievable reward OPT is at least g, i.e., if $ALG/OPT \geq g$.*

In certain situations where we are interested in the accumulated reward up to a certain time point t of the algorithm, we will refer to it as ALG_t.

We will consider agents who have beliefs about α, the potential for payoff, and who optimize their competitive ratio over an unknown θ, the time after which the payoff begins.

[1] for instance, during a PhD or while starting a business

Suppose an agent has a deterministic algorithm to choose how to play arms f_1 and f_2 as defined above. Each time the agent plays f_1 and then switches back to arm f_2 , they could instead have played f_2 followed by f_1 while gaining the same reward but possibly witnessing the increase sooner. Thus, there is no benefit to interweaving steps of the arms, and any deterministic strategy for playing this instance can be boiled down to the point at which it switches from f_2 to f_1 . We formalize this in the following lemma.

Lemma 3.2. *Any strategy that interweaves plays of f_1 and f_2 can be converted into a strategy that plays only f_2 followed by only f_1 that achieves at least as much reward.*

Proof. Suppose the interweaving strategy plays a total of t_1 steps on f_1 and s steps on f_2 . Consider two cases: in the first case, $s < \theta$. Then, the total reward of the policy is t_1 , and playing s steps of 0-reward-accruing f_2 followed by t_1 steps of f_1 achieves this reward, the same as any interweaving version. In the second case, $s \geq \theta$. Now, by the previous argument, the policy that plays θ steps of f_2 followed by t_1 steps of f_1 still receives t_1 reward. However, after θ steps on f_2 , the agent witnesses the increase and therefore has no incentive to switch to the stable arm. The reward of playing the remaining $s-\theta+t_1$ steps on the striving arm is $\frac{1}{2}(s-\theta+t_1)^2$, which is greater than t_1. Thus, we have shown that for each interweaved strategy, there is a non-interweaved one that accrues at least as much reward. □

As a result of this lemma, it suffices to study strategies that play f_2 for a while and then permanently switch to f_1. Thus, we will extensively study this switch point, and it will be an interesting quantity to study as a proxy for gritty and non-gritty strategies.

Morton and Paul discuss an "Evidential Threshold," writing "In a given context, how much evidence is required – that is, how compelling must the evidence be –before the thinker comes to a conclusion about what to believe or revises her current beliefs?" (Morton and Paul 2019). In our proposed framework of analysis, the switch point reflects this evidential threshold. The agent's strategy can be summarized as "if I don't see evidence that striving is going to pay off until time s, I will give up." The threshold s is different for different agents, and so the policy the agent follows exactly corresponds to their evidential threshold as described by Morton and Paul.

For a fixed switch point s , there are two "worst" cases – the first is when the arms are such that playing the stable arm the whole time would provide the optimal reward, and so the longer the agent spends exploring before switching, the worse the competitive ratio gets. On the other hand, if the striving arm pays off right after the agent switches, then staying on the arm just a little longer would have paid off, so this competitive ratio is increasing with s . In order for our strategy to minimize overall regret, we pick s such that the ratio is the same regardless of which extreme case we are in, i.e., we solve for s when the two extreme cases are equal.

3.2 Modelling Grit: Optimism

Now, let us discuss the relationship between the grittiness of an agent and their approach to the multi-armed bandit problem above. A gritty agent is optimistic about the potential for reward of the risky action, or they would not persevere in taking it. Accordingly, for this setting, we consider an agent with higher guess for the slope of the increasing portion of arm f_2 to be more gritty. Based on this, we can investigate consequences (in terms of both strategy and reward) of demonstrating grit and offer mechanistic insight as to why these consequences exist. The guess for the slope affects how long the agent is willing to play the striving arm. Formally:

Definition 3.3. *In the "grit-as-optimism" setting, an agent is $\tilde{\alpha}$-gritty if they guess that the slope of the increasing portion of the striving arm is $\tilde{\alpha}$.*

Lemma 3.4. *Suppose an agent guesses a value for α that we call $\tilde{\alpha}$, i.e., is $\tilde{\alpha}-$gritty. Assume their goal is to maximize the competitive ratio. Then, they play f_2 for $T-\sqrt{\frac{2T}{\tilde{\alpha}}}$ steps, following which they switch to f_1 permanently.*

Results Now, let us consider A, an $\tilde{\alpha}_A$-gritty agent. A is not particularly gritty, so $\tilde{\alpha}_A$ is small. On the other hand, B is somehow privy to perfect information, so $\tilde{\alpha}_B = \alpha$. Finally, consider C, an $\tilde{\alpha}_C$-gritty agent. C is very gritty, and so $\tilde{\alpha}_A < \tilde{\alpha}_B = \alpha < \tilde{\alpha}_C$. (We are simply instantiating the agents in this way to study "high" and "low" grit as they compare to perfect information.) In order to understand the impact the grit-induced strategy has on the reward the agent accrues, we are interested in understanding (1) what level of grit witnesses the striving arm paying off; (2) what level of grit results in good stable reward.

Observation 1: Duration of Attempt Applying Lemma 3.4, we have that agent A switches at time $s_A = T - \sqrt{\frac{2T}{\tilde{\alpha}_A}}$, agent B at $s_B = T - \sqrt{\frac{2T}{\alpha}}$, and agent C at $s_C = T - \sqrt{\frac{2T}{\tilde{\alpha}_C}}$. Note that $s_C > s_B > s_A$. Our first conclusion, therefore, is that the duration for which an agent explores is longer for a grittier person.

Observation 2: When Does Increased Grit Benefit the Agent? Let us now study under what conditions each agent comes out on top. The reward achieved by any agent depends on the relationship between θ , the threshold beyond which the striving arm starts increasing, and s , the agent's switch point as stated in the below proposition.

Proposition 3.5. *If $\theta \leq s$, then an agent switching at s receives $\frac{\alpha}{2}(T-\theta)^2$ reward, but if $\theta > s$, then the agent receives $T-s$ reward.*

Proof. If $\theta \leq s$, then the arm starts paying off while the agent is still playing it. This means that the agent accrues reward starting at time θ up until time T as the function increases linearly. Hence, the reward is $\frac{\alpha}{2}(T-\theta)^2$. On the other hand, if $\theta > s$, then the agent gains no reward from the striving arm. They gain reward from the stable arm from time $t = s$ to time $t = T$, which is $T-s$ units of reward. □

Let us consider how different levels of grit affect reward:

1. **Case 1: $\theta < s_A$.** Everyone's reward is the same in this case, since all agents receive $\frac{\alpha}{2}(T-\theta)^2$ reward.
2. **Case 2: $s_A < \theta < s_B$.** In this case, A has given up and switched to the stable arm. As a result, they receive $\sqrt{\frac{2T}{\tilde{\alpha}_A}}$ reward. However, agents B and C stay on the striving arm long enough to witness θ , and so they receive $\frac{\alpha}{2}(T-\theta)^2$ reward. We see that this is a situation where a lack of grit fares worse than being rather gritty.
3. **Case 3: $s_B < \theta < s_C$.** In this case, A and B have both given up and switched to the stable arm, but C valiantly perseveres. Here, A receives $\sqrt{\frac{2T}{\tilde{\alpha}_A}}$ reward, B receives $\sqrt{\frac{2T}{\tilde{\alpha}_B}}$ reward, and C receives $\frac{\alpha}{2}(T-\theta)^2$. C outshines even B, who had perfect information about the rate of reward increase. In this case, curiously, the least gritty person actually fares better than someone with perfect knowledge of the payoff. We can understand this as follows: there are two reasons why the received reward would be small – one is location of θ and one is size of α. In this case, someone that is pessimistic about the value of the reward magnitude due to pessimism about α ends up reaping the side benefit when the reward is indeed small, but it is because θ is large i.e., they are right about the reward being small but for the wrong reasons[2].
4. **Case 4: $s_C < \theta$.** In this case, no one receives the reward from the striving arm. However, since C has stuck around for so long, they actually also receive less reward overall from the stable arm. This indicates that there is a failure mode when an agent is *too* optimistic. Further, the resulting strategy of sticking it out for a long time can fail when θ is quite large. This reflects what (Wooten 2022) calls the "effort paradox," where students encouraged to be gritty often end up burnt out.

Observe that when the first agent switches, all agents have the same evidence about the striving arm. Likewise, when each agent switches, the remaining agents all have the same evidence, though different agents act differently, albeit all rationally, in response to this information. While it may at first seem disconcerting that people with access to the same evidence have *different* yet completely rational responses to it, this is is consistent with the philosophical thesis of Permissivism, which argues that "some bodies of evidence permit more than one rational doxastic attitude toward a particular proposition" (Jackson and LaFore 2023). Thus, it is reasonable for different agents to have different beliefs about the underlying state of the world upon receiving a set of evidence (i.e., different agents have different beliefs about how long it is worth staying on an arm after receiving reward 0 for the first s_A time). In fact, Morton argues that Permissivism exactly allows for grit to be conceived of as a factor in shifting people's behavior from the norm even when presented shared evidence. Indeed, in our model, we explicitly model this aspect of how beliefs + evidence $\rightarrow$ actions by having each agent hold different beliefs about the payoff slope α.

This perspective also provides us a natural way in which to consider the optimal level of grit for this setting. In particular, if we can remove the uncertainty in the guess for α , then the only remaining uncertainty has to do with when the function will start improving. This suggests that agents who have guesses for α closer to the true α will fare better.

3.3 Modelling Grit: Discomfort Tolerance

In this setting, we study another relevant aspect of grit, namely discomfort tolerance. To do so, let us first introduce a cost to playing the striving arm.

Cost to strive Let us consider a setting in which there is a cost to strive. The reward profile / bandit arms f_1 and f_2 look almost as before, with $\alpha = 1$: $f_1(t) = 1 \, \forall \, t$ and $f_2(t) = \begin{cases} -1 & t < \theta \\ t-\theta & t \geq \theta \end{cases}$.

The negative reward models the cost of striving, for instance financial debt or effort expenditure that can be offset by rewards from the stable arm to ensure the agent is not "in the negative," as the agent is not allowed have negative reward at any point in time. This corresponds to not being able to take out a loan. Thus, an agent who does not have any "savings" coming in must play f_1 before they ever play f_2 . We suppose that smallest piece of time an agent can split between the arms is 1 unit. Later, in Section 4, we consider what happens when the cost of striving is subsidized by a "trust fund."

Comfort Now, further, let us suppose an agent always wants to have average reward at least γ , that is, at time t , the agent wants their accrued reward to be at least $\gamma \cdot t$. They still aim to optimize the competitive ratio as before, except now they do so subject to the constraint that $\forall \, t$, the reward accrued up to that point is at least $\gamma \cdot t$.

Definition 3.6. *We say an agent desires γ-comfort if at any time $t > 0$, they require net reward at least $\gamma \cdot t$.*

The agent playing this game aims to solve:

$$\max \qquad \frac{ALG}{OPT} \quad \text{ s.t. } \quad \frac{ALG_t}{t} \geq \gamma \, .$$

Since the agent can play an arm for fractional amounts of time, an agent who requires γ comfort will play α_γ time on the stable arm and then $1-\alpha_\gamma$ time on the striving arm, alternating between the two as soon as possible, for $\alpha_\gamma := (\gamma+1)/2$, the time duration that guarantees the desired average reward:

$$ALG_t = \begin{cases} t & t \leq \alpha_\gamma \\ 2\alpha_\gamma - t & \alpha_\gamma < t \leq 1 \end{cases}$$

$$\Rightarrow \qquad \frac{ALG_t}{t} = \begin{cases} 1 & t \leq \alpha_\gamma \\ \frac{2\alpha_\gamma}{t} - 1 & \alpha_\gamma < t \leq 1 \end{cases}$$

$$\text{for } \alpha_\gamma < t \leq 1 \, , \gamma = 2\alpha_\gamma - 1 \leq \frac{2\alpha_\gamma}{t} - 1 \leq 1 \, .$$

[2]This is an interesting consideration for further modeling – in this particular setting, the model does not disentangle between these reasons for the reward to be small. More broadly, a competitive ratio-based model that studies reward in general rather than particular kinds of reward may not be able to disentangle this at all.

This is the most rational thing for them to do, since any additional steps on f_1 simply serve to offset future costs of f_2 but might prevent the agent from seeing the increase phase as quickly. We formalize this below, and the full proof proceeds via similar casework to the proof of Lemma 3.2).

Definition 3.7. *We call a strategy* minimally accumulating *for an agent who wants to be γ-comfortable if the average net reward at time t , ALG_t/t is exactly 1 until reaching time $\alpha_\gamma := (\gamma+1)/2$ when playing the stable arm and strictly decreasing until reaching* value γ *when playing the striving arm. In other words, the agent plays the stable arm for α_γ time followed by the striving arm for $1-\alpha_\gamma$ time and repeats.*

Lemma 3.8. *For each strategy that "stockpiles" reward along the way, there exists a minimally accumulating strategy that nets total reward at least as much as the stockpiling strategy.*

We present the result for the competitive ratio and reward for a γ-comfortable agent.

Lemma 3.9. *Suppose an agent who wants to be γ-comfortable plays f_1 for α_γ time followed by f_2 for $1-\alpha_\gamma$ time before reverting back to f_1 and continuing the process. Then, the agent wanting to maximize their competitive ratio subject to the constraint of the average reward always being at least γ will switch after absolute time $T - \frac{\gamma}{2} - \frac{1}{2}\sqrt{\gamma^2 + 4T(2-\gamma)}$, achieving competitive ratio $\gamma + \frac{\gamma(1-\gamma)}{2T} + \frac{(1-\gamma)\sqrt{\gamma^2-4T(2-\gamma)}}{2T}$. This corresponds to $\frac{1-\gamma}{2} \cdot \left(T - \frac{\gamma}{2} - \frac{1}{2}\sqrt{\gamma^2 + 4T(2-\gamma)}\right)$ time on the striving arm.*

Remark 3.10. *Consider a concrete numerical example: if $T = 150$, $\gamma = 0.5$, then, the agent will switch after about time 135. However, of that time, only about 34 would have been spent exploring, with the remaining 101 time spent on the stable arm. The competitive ratio achieved is 0.55.*

Remark 3.11. *For insight, let us next consider the behavior for extreme values of γ : if $\gamma = 0$, the first two terms go away, and the competitive ratio is $\frac{\sqrt{2T}}{T} = \frac{2}{\sqrt{T}}$. This exactly aligns with our computation before. On the other hand, if $\gamma \to 1$, the competitive ratio actually nears 1! This is because the best possible thing to do for an agent who requires complete comfort is to always play the stable arm. Indeed, to better understand this outcome, let us also investigate the total amount of time spent exploring as a function of γ . Taking the derivative of the expression for exploration time with respect to γ, we can see that it is negative for all $\gamma \in [0,1]$. This tells us that as an agent requires more "comfort," they spend less time exploring on the striving arm, and so while their competitive ratio improves, their chance of witnessing the growth in the striving arm and benefiting from it is low.*

4 Financial Support

A natural question following this discussion is how to incentivize a gritty agent to explore for longer. We have already seen that simply encouraging more grit could do more harm than good. We could encourage the agent to give up more comfort, but often a baseline level of comfort cannot be foregone – for instance, one may have to pay rent, buy food, etc. For this, we study the setting where there is a cost to striving and the agent must have discomfort tolerance $\gamma > 0$. In this section, we consider an implementation of a "trust fund," i.e., financial support we could provide an agent, and then show what conclusions we can draw about the strategy followed by an agent and their eventual reward. Here we consider the simplest possible implementation, which already has interesting outcomes, and more nuanced implementations have similar qualitative outcomes.

4.1 No Safety Net

Now, since an agent without a safety net must alternate between f_1 and f_2 , the competitive ratio maximizing strategy is computed as follows: first, if f_2 never increases, the agent incurs the competitive ratio on the left below, and if f_2 increases right after the agent switches, they receive the competitive ratio on the right.

$$\frac{(1-1)\cdot\frac{s}{2} + T - s}{T} = \frac{(1-1)\cdot\frac{s}{2} + T - s}{(1-1)\cdot\frac{s}{2} + \frac{1}{2}(T-s)^2}\,. \quad (2)$$

$$\Rightarrow s = T - \sqrt{2T}\,. \quad (3)$$

The duration of time after which the switch happens is $T - \sqrt{2T}$, but the proportion of that time actually spend exploring the striving arm is half that, namely $\frac{T-\sqrt{2T}}{2}$. Thus, if $\theta > \frac{T-\sqrt{2T}}{2}$, this agent will not be able to reap the benefits of striving.

4.2 Free Reimbursement

In this model, an agent with a support network gets reimbursed for free each time they taking a striving step. This is analogous to having a benefactor who financially supports the agent as much as needed to prevent their net reward from being negative. In this case, the effective arms for an agent with such unconditional support are now:

$$\hat{f}_1(t) = 1\,\forall\, t \qquad \hat{f}_2(t) = \begin{cases} 0 & t < \theta \\ t-\theta & t \geq \theta \end{cases}.$$

Thus, the analysis is as before (Lemma 3.4 with $\alpha, \tilde{\alpha} = 1$), and the agent will spend $T - \sqrt{2T}$ time on the striving arm before switching to the stable arm. Remarkably, the duration of time after which the agent without a safety net and the agent with unconditional support "give up" is the same! However, due to the safety net, the agent with it can explore the striving arm for twice as long. In particular, if $\theta \in [\frac{T-\sqrt{2T}}{2}, T - \sqrt{2T}]$, then the latter agent gets $\frac{1}{2}(T-\theta)^2 \geq \frac{1}{2}\cdot 2T = T$ reward, while one without the safety net only gets $\sqrt{2T}$.

We can also extend this to a model where the benefactor only promises support for a fixed amount of time. In that case also, a similar qualitative result holds – agents with and without support "give up" on striving at the same absolute time but the agent with the financial support gets to explore for a multiplicative factor longer. Note also that by mapping this onto the discomfort tolerance perspective, we can see that in a world where the agent must maintain a positive discomfort tolerance but the agent lacks financial support, they must spend less time exploring, whereas if they have financial support, they can explore longer.

4.3 Which Wins? Trust Fund or Grit?

In this section, we combine the pieces from the previous sections to understand the interplay between grit and financial support. We primarily present conclusions in this section. Here, the rate of increase is unknown, and there is also a cost to striving. Let us make two comparisons. First, let us investigate what happens when someone without a trust fund becomes more gritty. Then, we will study the effect of a trust fund on two people, one of whom is grittier than the other.

No trust fund, increased grit. For this, let us compare the first two rows of the Table below. Immediately, we see that increased grit, as before, leads to increased exploration time. In the last column, we present the reward that the agent receives if they don't witness the start of the payoff before switching, which we call "stable reward" for short. There, we can also see that the stable reward is *lower* when the agent is grittier.

Introduce trust fund, same grit. Now, let us compare the first and third rows of the table below. In this case, we see that at the same grit level, the presence of a safety net allows for a much longer exploration horizon. Both in the presence of and in the absence of the safety net, the stable reward is the same. This shows that the presence of a safety net allows for essentially "free" exploration.

grit level	safety net	explore time	stable reward
$\tilde{\alpha}_1$	none	$\frac{T}{2} - \sqrt{\frac{T}{2\tilde{\alpha}_1}}$	$\sqrt{\frac{2T}{\tilde{\alpha}_1}}$
$\tilde{\alpha}_2 > \tilde{\alpha}_1$	none	$\frac{T}{2} - \sqrt{\frac{T}{2\tilde{\alpha}_2}}$	$\sqrt{\frac{2T}{\tilde{\alpha}_2}}$
$\tilde{\alpha}_1$	yes	$T - \sqrt{\frac{2T}{\tilde{\alpha}_1}}$	$\sqrt{\frac{2T}{\tilde{\alpha}_1}}$

Table 1: Exploration time and stable reward (reward resulting from switching to f_1 following unsuccessful exploration on f_2) for various grit levels and safety nets.

Discussion We can view these results from two perspectives. One perspective is descriptive: we observe that providing a safety net increases exploration time essentially "for free." This happens since the agent does not have to split their time to ensure they are not in the negative. The other perspective is prescriptive: in settings where an external entity aims to encourage exploratory behavior, encouraging increased grittiness could lead to worse outcomes if the risk doesn't pay off. While we don't study this in this work, this could lead to future agents being discouraged from taking the grittier course of action. On the other hand, providing support to an agent who is already gritty encourages exploration without the prospect of bad outcomes in case the taken risk doesn't pay off. Thus, telling a gritty agent to be grittier is worse than giving them financial support to extend their exploration horizon. This corroborates and provides a simple mechanistic explanation for what (Morton and Paul 2019; Wooten 2022) observe.

5 Discussion of Modelling Choices

Modelling Perspective In this work, we provide a theoretical model for the following aspects of grit: we define a two-armed improving bandit instance that represents the decision-making problem agents are faced with. We further define two different objectives a decision-making agent might prioritize. Finally, we model different aspects of grit with careful study of variations of parameters in the bandit instance. While there are many folk notions of "grit", the aspects we choose to model are based on the abstraction and formalization provided by (Morton and Paul 2019). We do not seek to evaluate whether or not their account is valid but rather to quantitatively formalize their notions and study the implications of their analysis. In doing so, we find additional confirmation for empirical work regarding the value of "grants" provided to people attempting ambitious choices. For instance, a study of grants given to entrepreneurs in Burkina Faso shows that even if immediate profits aren't improved, providing this financial support increases innovation and improves business practices, suggesting that those who get support get more time to explore and build a good foundation for their business (Grimm, Soubeiga, and Weber 2021). Similarly, a study in Kenya found that intervening with grants to youth entrepreneurs at a time of crisis helped individuals maintain their businesses and produce more profits (Domenella et al. 2021). These empirical studies suggest that if someone is inclined to take an ambitious action, then supporting them helps improve outcomes, exactly what is captured by our model (as summarized in the table). In the future, our model could be used to develop resource allocation schemes in such settings. It would also be interesting to study community-level effects when agents of varying levels of grit make decisions based on each other.

Strengths A strength of our model is that the same multi-armed bandits instance can be used to study many different facets of grit, and resulting outcomes are visible just through this two-armed bandits instance. Also, our modelling allows for a modular understanding of the effect of grit on first, behavior and second, outcome or reward associated with that behavior. This, then, allows us to study interventions that encourage similar behavior with less outcome risk.

Weaknesses Our model is not without its shortcomings. We are limited to studying an agent's choice between two options in this framework. In real settings, the stable option might not have the same return for all time. Human rationality is rarely, if ever, executed exactly as competitive ratio maximization or expected reward maximization over a prior. However, these seem like natural simplifications that provide a starting point for quantitative analysis of this trait.

6 Conclusion

We introduced a quantitative model for studying the impact of grit and financial support on decision-making between a long-term, ambitious end and immediate-reward, stable end. We developed a two-armed bandit model in which to study this and formalized notions of rationality and grit that gave rise a family of strategies for this instance. Our modeling engages with several prior works in the social science literature, and we hope our approach and framework admit future work on quantitative study of grit.

Acknowledgments

This work was supported in part by the National Science Foundation under grants CCF2212968 and ECCS-2216899, by the Simons Foundation under the Simons Collaboration on the Theory of Algorithmic Fairness, and by the Office of Naval Research MURI Grant N000142412742

References

Bernardo, J. M.; and Smith, A. F. M. 2000. *Bayesian Theory*. Wiley.

Blum, A.; and Ravichandran, K. 2025. Nearly-tight Approximation Guarantees for the Improving Multi-Armed Bandits Problem. In Kamath, G.; and Loh, P.-L., eds., *Proceedings of The 36th International Conference on Algorithmic Learning Theory*, volume 272 of *Proceedings of Machine Learning Research*, 228–245. PMLR.

Borodin, A.; and El-Yaniv, R. 2005. *Online Computation and Competitive Analysis*. Cambridge University Press. ISBN 978-0-521-61946-2. Google-Books-ID: v3faI8pER6IC.

Domenella, Y.; Jamison, J. C.; Safir, A.; and Zia, B. 2021. Can Business Grants Mitigate a Crisis? Evidence from Youth Entrepreneurs in Kenya during COVID-19. *World Bank, Washington, DC*.

Duckworth, A. 2016. *Grit: The Power of Passion and Perseverance*. Scribner.

Gibbon, P. 2020. Martin Seligman and the Rise of Positive Psychology. *Humanities*, 41(3).

Grimm, M.; Soubeiga, S.; and Weber, M. 2021. *Short-Term Impacts of Targeted Cash Grants and Business Development Services: Experimental Evidence from Entrepreneurs in Burkina Faso*. World Bank, Washington, DC.

Heidari, H.; Kearns, M.; and Roth, A. 2016. Tight Policy Regret Bounds for Improving and Decaying Bandits. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, 1562–1570. AAAI Press. ISBN 9781577357704.

Jackson, E.; and LaFore, G. 2023. *Permissivism, Underdetermination, and Evidence*, 358–370. Routledge, 1 edition. ISBN 978-1-315-67268-7.

Kamenetz, A. 2015. A Key Researcher Says 'Grit' Isn't Ready For High-Stakes Measures. *NPR*.

Morton, J. M.; and Paul, S. K. 2019. Grit. *Ethics*, 129(2): 175–203.

Patil, V.; Nair, V.; Ghalme, G.; and Khan, A. 2023. Mitigating Disparity while Maximizing Reward: Tight Anytime Guarantee for Improving Bandits. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 4100–4108. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-1-956792-03-4.

Seligman, M. 2002. *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfillment*. Free Press. ISBN 978-0-7432-4788-7.

Seligman, M. 2006. *Learned Optimism: How to Change Your Mind and Your Life*. Vintage Books. ISBN 9780394579153.

Slivkins, A. 2022. Introduction to Multi-Armed Bandits. arxiv:1904.07272 [cs, stat].

Wooten, T. 2022. *Precarious Mobility: Trying and Failing to Get Ahead in the 21st Century*. Ph.D. thesis, Harvard University. Accepted: 2022-06-07T06:32:36Z.