Underdogs and One-hit Wonders: When is Overcoming Adversity Impressive?

Jerker Denrell, Warwick Business School, University of Warwick
Chengwei Liu, ESMT Berlin
David Maslach, College of Business, Florida State University

May 17, 2022

Abstract

Success tends to increase and failure tends to decrease the chances of future success. We show that this impact of past outcomes can change how diagnostic success or failure are regarding the competence of an individual or a firm. Succeeding under adverse circumstances is especially impressive when initial failure reduces the chances of success more for low-quality agents than for high-quality agents. Succeeding after initial failure (being a "successful underdog") can also indicate higher expected quality than succeeding twice if initial success increases the chances of success of all agents to a high level. In different circumstances, the outcome after success can be especially informative about quality, implying that failing after an initial success (a "one-hit wonder") indicates lower quality than failing twice does. We find effects consistent with our model in data on Canadian professional hockey players and on data from the Music Lab experiment: initial failure combined with eventual success is associated with high quality. The results help to clarify when failure should be attributed to the person in charge or to the situation, when underdogs and individuals who overcome adversity are especially impressive, and when a naïve "more is better" heuristic for evaluating performance can be misleading.

Keywords: Matthew Effect, performance evaluation, luck, organizational learning.

1 Introduction

Consider two venture capital managers, Robin and Grace, who are on the short list to manage a large new fund of your venture capital firm. Robin graduated from an elite university, worked for an asset management company, and then became a venture capital fund manager. During her first three years, her fund invested in a start-up that quickly became very successful. The success of this investment made her fund well-known and attracted many start-ups. Robin's fund continued to do well the next three years, investing in several companies that became successful. Grace had a similar background and also became a venture capital fund manager, but her fund had a slow start. None of the investments made during the first three years were a success. During the next three years, however, the fund invested in several companies that became very successful. After six years, both funds were valued the same. Which candidate should you hire to manage your new fund, Robin or Grace?

This paper analyzes how inferences about quality should be made in a system that is subject to strong reinforcing processes. Specifically, when is repeated success (Robin) a less reliable indicator of high quality than success after initial failure (Grace)? There is a large literature on how quality assessments should be done and are done in organizations. A central issue in this literature is whether outcomes are appropriately attributed to the person or to the situation (Perrow, 1984; Ross Nisbett, 1991; Morris Larrick, 1995). Disentangling situational and personal factors becomes even more difficult when past outcomes impact the situation in which people find themselves. For example, Robin initially succeeded in part because she is skilled. When her fund became well-known and attracted many capable entrepreneurs, her odds of success increased because the quality of the start-ups she selected from improved (Sørensen, 2007). Is not succeeding under such favorable circumstances less diagnostic of high quality than succeeding under adverse circumstances? Because Grace did not succeed initially, the pool of start-ups she selected from was less promising, but she was still able to identify valuable investments. Does this imply that Grace is the superior fund manager?

In this paper, we analyze formally when succeeding under adverse circumstances, caused by prior failure, is an indication of high quality. We show that even if people do not become more motivated after an initial failure (Nurmohamed, 2020; Ashworth Heyndels, 2007) or learn more quickly (Barnett Hansen, 1996; Kim Miner, 2007), those who succeed after an initial failure can have higher expected quality than those who succeed repeatedly because success after failure can be more informative about quality. Specifically, using a simple analytical model, we demonstrate formally that succeeding after an initial failure (a successful underdog) is an especially strong indication of high quality when initial failure impacts low-quality agents proportionally more than high-quality agents. If an initial success also mainly benefits low-quality agents, the expected quality of agents who succeed twice ($E[q|SS]$) can be lower than the expected quality of agents who succeeded after an initial failure ($E[q|FS]$). This is because the outcome after success is not informative of quality, but the outcome after failure is. The result $E[q|FS] > E[q|SS]$ cannot occur, however, when initial success or failure impact high-quality agents more than low-quality agents. In such cases, the outcome after failure is not informative about quality, but the outcome after success is. In this case, the expected quality of "one-hit wonders," i.e., agents who fail after an initial success, can be lower than the expected quality of agents who failed twice: $E[q|SF] < E[q|FF]$.

We show that effects consistent with our model occur in data on Canadian professional hockey players and from the Music Lab experiment (Salganik et al., 2006). Canadian hockey players who made it to the National Hockey League (NHL) but were not initially selected for the Junior League in Canada scored more points and had higher salaries than players who were selected for the Junior League. Similarly, songs in Salganik et al.'s (2006) experiment on music downloads that were initially unpopular are of higher quality than songs that were initially popular but reached the same final market share. The experimental nature of this study helps rule out alternative accounts, such as selection bias or changes in motivation or quality.

Our finding that a less-than-perfect track record can be a superior indication of high quality is significant, as studies show that people and managers ignore situational factors that are not salient. If most managers believe that Robin is superior to Grace, savvy managers who understand that Grace's imperfect track record is actually more impressive can gain an advantage (Denrell et al., 2019; Siegel et al., 2019). More generally, our findings show that fair assessments of individuals should correct for the role of reinforcing processes in producing outcomes.

2 Inferring Quality in the Presence of Reinforcing Processes

Managers typically use past performance to evaluate people and products (Strang Patterson, 2014; Jenter Kanaan, 2015; Benson et al., 2019; Audia Greve, 2021), but they do not treat all performances the same (Ouchi Maguire, 1975; Bol Smith, 2011; Desai, 2015). High performance is less impressive in an industry where most firms do well due to high demand (Holmstrom, 1982). A CEO may be fired even if the absolute level of performance is high but below expectations (Puffer Weintrop, 1991). Similarly, venture capitalists may defend prolonged periods of losses for a start-up if they believe the start-up will eventually turn the corner (Guler, 2007).

Several factors that impact whether high performance is a reliable signal of quality are well understood. Performances that are substantially influenced by luck and factors beyond managers' control are less informative of skill (Holmstrom, 1979). Performances observed today may be the result of actions taken a long time ago (Rahmandad Gary, 2020). The fact that performance evaluations can be subject to bias resulting from discrimination has received growing attention (Castilla, 2008). To avoid relying on noisy and biased performance measures, theorists argue that managers should combine financial and non-financial measures, comparisons with other agents, and subjective performance evaluations (Baker et al., 1988, 1994; Holmström, 2017).

Reinforcing processes also complicate inferences of quality from observed performance. An initial success can, at least temporarily, increase the chances of future success of an individual, firm, or product, generating a virtuous cycle in which success leads to future success (Arthur, 1989; Lieberman Montgomery, 1988; Salganik et al., 2006; Van de Rijt, 2019). Inferring quality based on the outcome of such reinforcing processes is complicated, as the difference in final outcomes is not proportional to the quality differences and because reinforcing processes magnify noise (Arthur, 1989; Lynn et al., 2009). Exceptional salaries do not necessarily indicate exceptional talent, but rather a winner-take-it-all dynamics (Frank Cook, 1995). Similarly, researchers writing about accidents have stressed that it is usually inappropriate to infer that a disaster was caused by individual incompetence proportional to the magnitude of the failure. Disaster is often initiated by a small mistake that triggers additional errors, resulting in a cascade of errors that eventually spirals into disaster (Perrow, 1984; Rudolph Repenning, 2002; Dörner, 1996). Prior work has also shown that because reinforcing processes amplify chance events and increase the variability of outcomes, an extreme outcome (including exceptional success) can be an unreliable signal of competence because it suggests that reinforcing processes are strong (Denrell Liu, 2012).

Less commonly discussed is how success or failure can make subsequent outcomes more or less diagnostic (i.e., revealing) of competence (Meyer, 1991) and what the implications are for quality evaluation. Success in the past can increase and failure can decrease the financial and social resources available to agents. This, in turn, can change the degree to which an outcome is informative about the quality of an agent. For example, suppose that past success increases the resources of all agents to a high level. Assume that this level of resources can compensate for low quality so that even low-quality agents will succeed with high probability today with this level of resources. In this case, a successful outcome today does not provide much additional information about the quality of an agent. Indeed, in the extreme case where all agents who succeeded in the past have the same success probability today, we learn nothing new about quality from observing the outcome (success or failure) today. Alternatively, suppose that a past failure decreases all agents' resources to a low level. Imagine that only high-quality agents can compensate for this lack of resources. Thus, only high-quality agents have a moderate success chance today, while all others have a low chance of success. In this case, a successful outcome today is a strong indication of high quality. This also raises the possibility that success after an initial failure could be a stronger indicator of high competence than succeeding twice.

Below, we formalize these scenarios and analyze when succeeding after a failure can be more or less diagnostic of high quality and when this change in diagnosticity implies that succeeding after failure can be a stronger indication of high quality than succeeding twice. We start by defining *diagnosticity* in the next section. In section 4, we illustrate our argument using a simple two-period model with two levels of quality. Section 5 provides a more general analysis of how reinforcing processes impact the diagnosticity of outcomes and how this changes quality evaluation. Section 6 provides two empirical illustrations, Section 7 discusses the theoretical and empirical results, and Section 8 explores implications.

3 The Diagnosticity of Success and Failure

A success can be more or less diagnostic of high quality, broadly defined as a trait that increases odds of success and includes the competence, skill, ability, merit, or effectiveness of the relevant actor (individual or firm), depending on the context. We define *diagnosticy of a success* as the probability that an agent who succeeds is of high quality: $P(H|S)$ Empirically, $P(H|S)$ can be estimated by the proportion of high-quality agents among all agents who succeeded (assuming we can measure quality independently of success)[1] For example, suppose 50% of high-quality agents succeed but only 5% of low-quality agents succeed. Suppose that there are 100 high-quality and 100 low-quality agents. Thus, there are 0.5 * 100 high-quality agents who succeed and 0.05 * 100 low-quality agents who succeed. The diagnosticity of success is then

$$P(H|S) = \frac{0.5 * 100}{0.5 * 100 + 0.05 * 100} = 0.909. \tag{1}$$

In general, the diagnosticity of success depends on how much more likely high-quality agents are to succeed as compared to low-quality agents. A success is more diagnostic if the success chance of high-quality agents is much higher than the success chance of low-quality agents. What matters, however, is not the difference between the success probabilities (0.5-0.05 in the example above), but the ratio (0.5/0.05). To see why, note that

$$P(H|S) = \frac{0.5 * 100}{0.5 * 100 + 0.05 * 100} = \frac{\frac{0.5*100}{0.05*100}}{1 + \frac{0.5*100}{0.05*100}}, \tag{2}$$

[1]Our definition of diagnosticity is consistent with other definitions that rely on the likelihood ratio. An observation D is diagnostic for hypothesis H if the likelihood ratio $P(D|H)/(D| - H)$ is different from 1 (Fischhoff Beyth-Marom, 1983). Our definition focuses on $P(H|S)$, a function of the likelihood ratio: $P(H|S) = 1/(1 + P(S|L)/P(S|H))$. Moreover, if $P(S|H) > P(S|L)$, the likelihood ratio can only deviate from 1 in one direction.

which is an increasing function of the ratio (0.5/0.05). More formally, let $P(S|H)$ and $P(S|L)$ be the success probabilities of high- and low-quality agents. The diagnosticity of success is an increasing function of the ratio $P(S|H)/P(S|L)$.

If the diagnosticity of success equals one, then success or failure perfectly sorts high- and low-quality agents: only high-quality agents succeed. If high- and low-quality agents are equally likely to succeed, success is not diagnostic at all. Diagnosticity is then equal to the base rate of high quality in the population ($P(H)$). In this case, knowing that agents i succeeded does not provide any new information about whether i is of high quality or not.

We can also define the diagnosticity of failure as the probability that someone who failed is of low quality: $P(L|F)$. Let $P(F|H)$ and $P(F|L)$ be the failure probabilities of high- and low-quality agents. The diagnosticity of a failure is an increasing function of the ratio $P(F|H)/P(F|L)$.

4 How Past Outcomes Change the Diagnosticity of Success and Failure

The main point in this paper is that success or failure can change the diagnosticity of future success and failure by making the task more or less difficult. As a result, success after an initial failure can be more diagnostic of high quality than succeeding twice is.

To explain when and why this occurs, consider a simple two-period model with two levels of quality. This initial illustrative model is as simple as possible for a model where the probability of success may change over time. In each of the two periods, an agent can succeed or fail. The agent can be of high or low quality: high-quality agents are more likely to succeed than low-quality agents. It is equally likely that an agent is of high or low quality. Importantly, we assume that the outcome in the first period, success or failure, changes the success probability in the second period. Success in the first period increases the probability of success in the second period, because success provides the agent with additional resources (financial or social). Failure in the first period decreases the success probability in the second period because failure depletes resources. Specifically, we assume that an agent starts out with resources equal to 0.5. A success increases resources to 1 and a failure decreases resources to zero. This is akin to accumulating resources with success and depleting resources with failure. We are interested in when this process implies that succeeding after an initial failure is an indicator of high quality.

Consider first the scenario in the upper row of Figure 1. A high-quality agent with resources equal to 0.5 succeeds with probability $p_h = 0.6$, while a low-quality agent with the same resources succeeds with probability $p_l = 0.4$. After an initial success, the resources increase to one. A high-quality agent then succeeds with probability $p_{h,S} = 0.7$, and a low-quality agent with probability $p_{l,S} = 0.595$. After an initial failure, the resources decrease to zero, and then a high-quality agent succeeds with probability $p_{h,F} = 0.5$, and a low-quality agent with probability $p_{l,F} = 0.3$.

Consider now an evaluator who does not know whether an agent is of high or low quality, but can observe whether an agent succeeds or fails. How should the evaluator assess the probability that an agent is of high quality, depending on the number and order of the successes? Is an agent who succeeds after a failure more likely to be of high quality than an agent who initially succeeds and then fails? The right column of Figure 1 shows how the probability of being high quality varies with the number and order of successes. If we only know that someone succeeded in the first period, the probability of being high quality is 0.6. If we know that someone succeeded in both periods, the probability of being high quality is 0.638. In this case, an agent who succeeds after a failure is equally likely to be of high quality as an agent who initially succeeds and then fails: $P(H|F,S) = P(H|S,F) = 0.526$.

If we change the success probabilities slightly, however, succeeding after an initial failure is more impressive than succeeding initially and then failing. To illustrate when this happens, consider the scenario in the middle row of Figure 1. This scenario is identical to the upper row except for one number: the success probability for a low-quality agent with zero resources is equal to $p_{l,F} = 0.2$ instead of $p_{l,F} = 0.3$. An agent who succeeds after a failure is now more likely to be of high quality ($P(H|F,S) = 0.626$) than an agent who initially succeeds and then fails ($P(H|S,F) = 0.526$).

Why is failing initially and subsequently succeeding more impressive than initially succeeding and then failing? And why did this happen when we decreased $p_{l,F}$ from 0.3 to 0.2? The reason is that this change makes a success after an initial failure more diagnostic of high quality. In the upper row, high-quality agents succeed with probability 0.5 after a failure, while low-quality agents succeed with probability 0.3. High-quality agents are 1.66 more likely to succeed than low-quality agents are. In the middle row, high-quality agents succeed with probability 0.5 after a failure, while low-quality agents succeed with probability 0.2: high-quality agents are 2.5 times more likely to succeed than low-quality agents are. Succeeding after a failure is thus highly diagnostic of high quality in the scenario in the middle row.

Of course, an initial failure is an indication of low quality. If it was *only* known that an agent has failed initially, the probability that this agent is of high quality is 0.4. The success after this initial failure, however, compensates for the initial failure. If we also know that the agent succeeded after the initial failure, this is very likely a high-quality agent: the probability increases to 0.626. Overcoming adversity, i.e., succeeding after an initial failure, is thus especially diagnostic of high quality because high-quality agents are much more likely to do so compared to low-quality agents. Succeeding after a failure is thus an especially difficult task that requires high quality.

Two other aspects of the scenario in the middle row of Figure 1 are crucial for the result that $P(H|F,S) > P(H|S,F)$. First, the outcome in the first period is not very diagnostic of quality. $P(H|S)$ is 0.6 while $P(H|S)$ is 0.4. If high-quality and low-quality agents had differed more in their success probabilities in the first period, the outcome in the first period would have been more diagnostic of high ability, and then $P(H|S,F)$ would have been greater than $P(H|F,S)$. Second, failing after an initial success is quite diagnostic of low quality. The reason is that low-quality agents are more likely to fail after initial success than high-quality agents are ($1 - p_{l,S} = 0.405$ versus $1 - p_{H,S} = 0.3$).

In summary, $P(H|F,S) > P(H|S,F)$ occurs when:

1) Succeeding after failure is highly diagnostic of high quality, i.e., high-quality agents are much more likely to succeed after a failure than low-quality agents are (0.5 versus 0.2). The reason may be that only high-quality agents can deal effectively with initial failure that depletes their resources (i.e., "when the going gets tough, the tough gets going").

2) An initial success is not very diagnostic of high quality, i.e., high-quality agents are just a bit more likely to succeed in the first period than low-quality agents are (0.6 versus 0.4). The reason may be that the first outcome is substantially affected by external events outside the control of the agents or based on a small sample size.

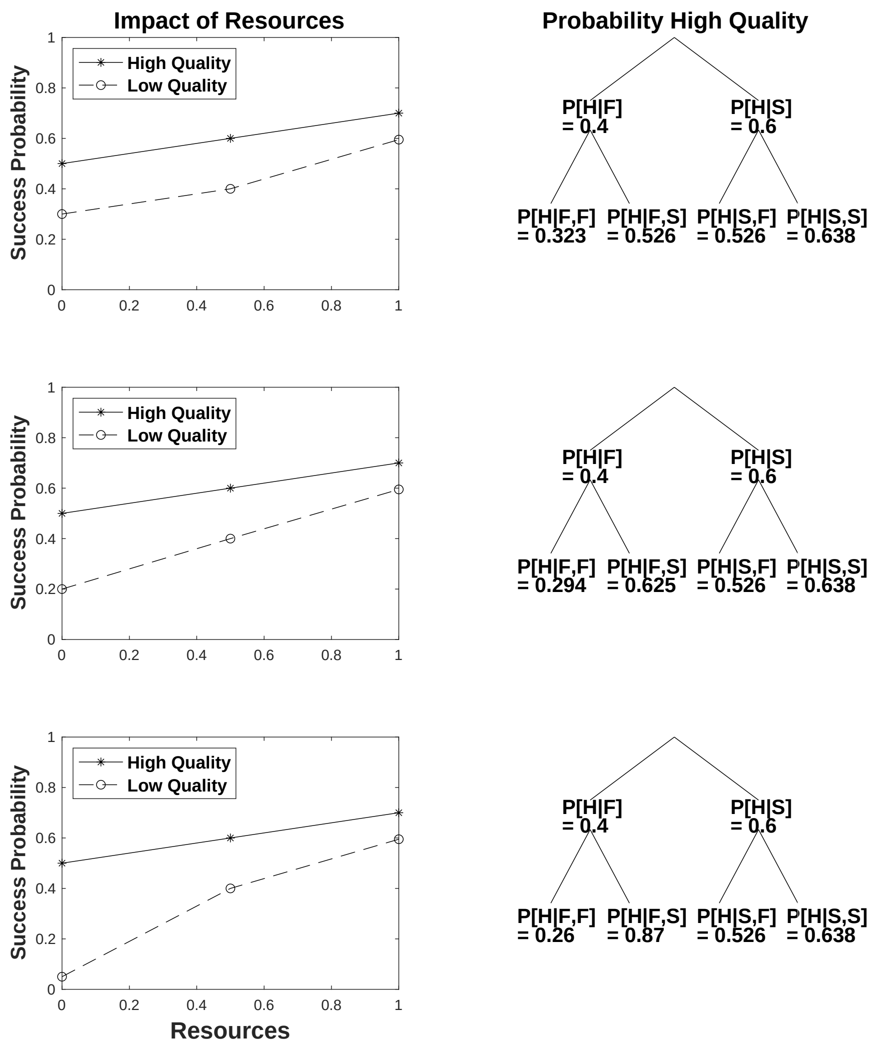

Figure 1: Left Column: Impact of resources on the success probability for high- and low-quality agents in three different scenarios. Right column: the probability of being high quality given the order of successes and failures. In all models $p_h = 0.6, p_l = 0.4. p_{h,S} = 0.7, p_{l,S} = 0.595, p_{h,F} = 0.5$. Only the value of $p_{l,F}$ differs, which is 0.3 (upper panel), 0.2 (middle Panel) and 0.05 (lower panel).

3) Failing after an initial success is diagnostic of low quality, i.e., low-quality agents are more likely to fail after an initial success than agents of high quality are (0.405 vs. 0.3).

More formally, $P(H|F,S) > P(H|S,F)$ occurs (see the Appendix) if and only if

$$\frac{1-p_h}{1-p_l}\frac{p_{h,F}}{p_{l,F}} > \frac{p_h}{p_l}\frac{1-p_{h,S}}{1-p_{l,S}}. \tag{3}$$

That is, when i) $p_{h,F}/p_{l,F}$ is large (high-quality agents are much more likely to succeed after a failure than low-quality agents are), ii) p_h/p_l is small (high-quality agents are just a bit more likely to succeed initially than low-quality agents are), and iii) $(1-p_{h,S})/(1-p_{l,S})$ is small: low-quality agents are more likely to fail after a success than high-quality agents are. Note that if p_h/p_l is small, then $(1-p_h)/(1-p_l)$ is large.

If we change the success probabilities further, succeeding after an initial failure can be even more impressive than succeeding twice. This is illustrated in the bottom row of Figure 1. This scenario is identical to the scenario in the middle row, except for one number: the success probability for a low-quality agent with zero resources is equal to $p_{l,F} = 0.05$ instead of $p_{l,F} = 0.2$. The implication is that an agent who succeeds after a failure is more likely to be of high quality than an agent who succeeds twice: $P(H|F,S) = 0.87$ while $P(H|S,S) = 0.638$. The reason is that succeeding after an initial failure is now even more diagnostic of high quality: high-quality agents are five times more likely to succeed than low-quality agents are after an initial failure. Moreover, suppose that an evaluator did not know the order of the successes and failures, only their number. In the bottom row, agents who succeeded once are more likely to be of high quality (0.698, the weighted average of $P(H|F,S)$ and $P(H|S,F)$) than those who succeeded twice (0.638).

Overall, three aspects of the scenario in the bottom row of Figure 1 are crucial for the result $P(H|F,S) > P(H|S,S)$:

1) Succeeding after failure is highly diagnostic of high quality, i.e., high-quality agents are much more likely to succeed after a failure than low-quality agents are (0.5 versus 0.05).

2) An initial success is only a weak indicator of high quality, i.e., high-quality agents are just a bit more likely to succeed in the first period than low-quality agents are (0.6 versus 0.4).

3) Succeeding after success is not very diagnostic of high quality, i.e., high-quality agents are just a bit more likely to succeed after a success than low-quality agents are (0.7 versus 0.595). As a result, the probability of being high quality only increases from 0.6 to 0.638 after a success.

Formally, $P(H|F,S) > P(H|S,S)$ occurs (see the Appendix) if and only if

$$\frac{1-p_h}{1-p_l}\frac{p_{h,F}}{p_{l,F}} > \frac{p_h}{p_l}\frac{p_{h,S}}{p_{l,S}}. \tag{4}$$

That is, when i) $p_{h,F}/p_{l,F}$ is large (high-quality agents are much more likely to succeed after a failure than low-quality agents are), ii) p_h/p_l is small (high-quality agents are just a bit more likely to succeed in the first period than low-quality agents are), and iii) $p_{h,S}/p_{l,S}$ is small (high-quality agents are just a bit more likely to succeed after a success than low-quality agents are).

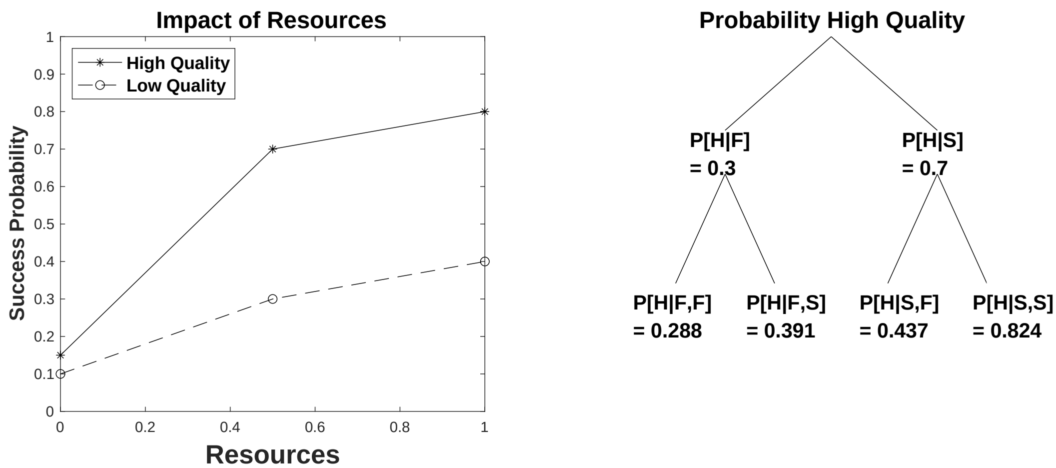

Figure 2: A) Impact of resources on the success probability for high- and low-quality agents. B) the probability of being high quality given the order of successes and failures. Here $p_h = 0.7, p_l = 0.3. p_{h,S} = 0.8, p_{l,S} = 0.4, p_{h,F} = 0.15$, and $p_{l,F} = 0.1$.

Succeeding after initial failure is not always impressive. In some cases, it is less impressive than failing after an initial success. Figure 2 shows one example where $P(H|F,S) = 0.391$ while $P(H|S,F) = 0.437$. This occurs when the success probabilities in the first period differ more than in Figure 1: $p_h = 0.7$ and $p_l = 0.3$. Thus, initial success is more diagnostic of high quality, and failure is more diagnostic of low quality. However, failure overwhelms both high- and low-quality agents, in the sense that both are likely to fail again after an initial failure: the success probabilities fall to $p_{h,F} = 0.15$ and $p_{l,F} = 0.1$. In this scenario, success after an initial failure is not diagnostic of high quality, and $P(H|F,S)$ does not increase much from $P(H|F) = 0.3$. On the contrary, the outcome after initial success is quality diagnostic, and $P(H|S,F)$ decreases substantially from $P(H|S) = 0.7$. However, $P(H|S,F)$ remains higher than $P(H|F,S)$. Formally, $P(H|S,F) > P(H|F,S)$ occurs when the inequality in equation (3) does not hold.

In all scenarios so far, failing twice indicated the lowest probability of high quality. If we change the success probabilities slightly, however, we can make failing after a success less impressive than failing twice. Specifically, consider the success probabilities in Figure 3 where initial failure reduces the success probabilities of both agents to a low level, while initial success only benefits high-quality agents. In this case $p_h = 0.6$ and $p_l = 0.4$ as in Figure 1, but $p_{h,S} = 0.9, p_{l,S} = 0.5$ and $p_{h,F} = 0.15, p_{l,F} = 0.05$. Figure 3 shows that in this case, the probability of being high quality is higher for someone who failed twice than for someone who succeeded and then failed: $P(H|F,F) = 0.386$ while $P(H|S,F) = 0.231$. The reason is that failure after an initial failure is not very diagnostic of low quality because high-quality agents are also likely to fail. Moreover, succeeding after an initial success is now highly diagnostic of high quality because high-quality agents are much more likely to succeed: only high-quality agents benefit much from the increased resources that initial success brings.

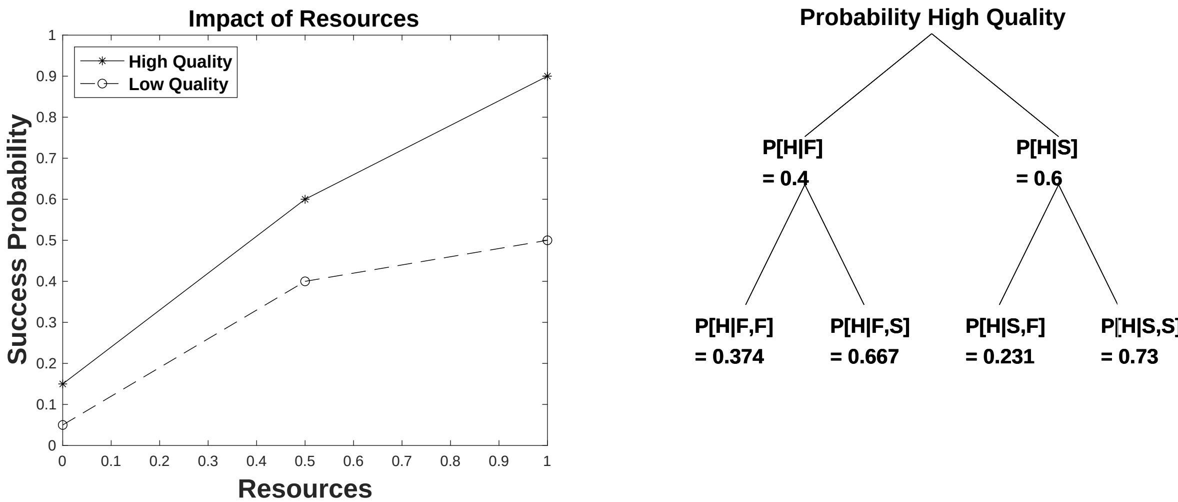

Figure 3: A) Impact of resources on the success probability for high- and low-quality agents. B) The probability of being high quality given the order of successes and failures. Here $p_h = 0.6, p_l = 0.4. p_{h,S} = 0.9, p_{l,S} = 0.5, p_{h,F} = 0.15$, and $p_{l,F} = 0.05$.

In summary, three aspects of the scenario in Figure 3 are crucial to get the result that failure after an initial success is less impressive than failing twice:

1) Failing after an initial success is very diagnostic of low quality, i.e., low-quality agents are much more likely to fail after a success than agents of high quality ($1 - p_{l,S} = 0.5$ versus $1 - p_{h,S} = 0.1$). The reason may be that only high-quality agents benefit from the increased resources that initial success brings.

2) Failing after a failure is not very diagnostic of low quality, i.e., low-quality agents are just a bit more likely to fail after a failure than high-quality agents are (0.95 versus 0.85). The reason may be that failure overwhelms all agents, and even high-quality agents are unlikely to succeed without resources.

3) An initial failure is not very diagnostic of low quality, i.e., low-quality agents are just a bit more likely to fail in the first period than high-quality agents (0.6 vs. 0.4).

More formally, let $q_h = 1 - p_h$ be the probability of failure for the high-quality agent. Similarly, $q_l = 1 - p_l$ is the probability of failure for the low-quality agent. We also define $q_{h,S} = 1 - p_{h,S}$ and $q_{h,F} = 1 - p_{h,F}$, the probability of failure after initial success or failure for the high-quality agent. Similarly, $q_{l,S} = 1 - p_{l,S}$ and $q_{l,F} = 1 - p_{l,F}$, the probability of failure after initial success or failure for the low-quality agent. It is then easy to show (see the Appendix) that the probability of being low quality is higher for an agent that fails after an initial success compared to an agent who fails twice, $P(L|S, F) > P(L|F, F)$

if and only if:

$$\frac{q_{l,S}}{q_{h,S}}\frac{1-q_l}{1-q_h} > \frac{q_{l,F}}{q_{h,F}}\frac{q_l}{q_h}. \tag{5}$$

This inequality is satisfied when i) $q_{l,S}/q_{h,S}$ is large: low-quality agents are much more likely to fail after a success than high-quality agents are, ii) $q_{l,F}/q_{h,F}$ is small: low-quality agents are just a bit more likely to fail after a failure than high-quality agents are, and iii) q_l/q_h is small: low-quality agents are just a bit more likely to fail in the first period than high-quality agents are. Note that this condition is identical to inequality (4), except that high and low quality change place, and we focus on failure probability instead of success probability.

5 Quality-Resources Interactions Impact Inferences from Success and Failure

5.1 When do past outcomes change the diagnosticity of success and failure?

We have shown that succeeding once can be more impressive than succeeding twice when past outcomes change the probability of success in such a way that success after initial failure is especially diagnostic of high quality, while success after initial success is not very diagnostic of high quality. This raises the question: when does the past outcome change the diagnosticity of success in such a way? To answer this question, suppose that the probability of success is an increasing function of both quality and resources: $g(q, r)$. Suppose success in the first period increases resources and failure decreases the resources available in the second period. Initial failure makes a subsequent success more diagnostic of high quality, and initial success makes a subsequent success less diagnostic, when an increase in resources leads to a larger *proportional increase* in the success probability for *low-quality agents* than for high-quality agents:

$$\frac{g(l, r+\delta) - g(l, r)}{g(l, r)} > \frac{g(h, r+\delta) - g(h, r)}{g(h, r)}. \tag{6}$$

The intuition is that increased resources then moves the success probability of low-quality agents closer to the success probability of high-quality agents, which makes a success less diagnostic (c.f. Wellman Henrion (1993)).[2] Conversely, decreased resources make the probability of success of low- and high-quality agents more different, which makes a success more diagnostic. Condition (6) is a necessary but not sufficient condition for $P(H|F,S) > P(H|S,S)$ (see Appendix B). $P(H|F,S) > P(H|S,S)$ does not only require that an initial failure makes a subsequent success more diagnostic, but also that the effect is sufficiently strong to compensate for the initial success.

When do low-quality agents benefit (proportionally) more from increased resources? This can occur when there are diminishing returns to resources. Suppose the probability of success is an increasing concave function of total resources, $t = q + r$–i.e., there are diminishing returns to t. For example, high-quality agents who already have substantial human capital may not benefit much from additional social or financial capital, while

[2] Formally, condition 6 implies $\frac{g(l,r+\delta)}{g(l,r)} > \frac{g(h,r+\delta)}{g(h,r)}$, or $\frac{g(h,r)}{g(l,r)} > \frac{g(h,r+\delta)}{(l,r+\delta)}$, which implies that a success is more diagnostic of high quality for low resources.

low-quality agents do. Formally, condition (6) occurs when quality and resources are "strong" substitutes, in the sense that $\partial \ln g(q,r)/\partial r \partial q < 0$ (Karlin Rubin, 1956).[3] Such a substitution effect has been observed in several business settings. For example, Chatterji et al. (2019) show that founders with high human capital benefit less from advice than founders with low human capital. Pollock et al. (2015) show that participating in blockbuster deals benefits well-known, high-status venture capital firms less than other venture capital firms. Salganik et al. (2006) showed that the long-term popularity of high-quality songs was less affected by the early popularity and attention than medium-quality songs were (Hendricks et al., 2012). In these settings, agents can succeed by having high quality and few resources or by having medium quality and high resources. It is as if success depends on the maximum: $g(q,r) = max(q,r)$.

Increased resources makes a success *more* diagnostic of high quality when the opposite condition holds:

$$\frac{g(l, r+\delta) - g(l,r)}{g(l,r)} < \frac{g(h, r+\delta) - g(h,r)}{g(h,r)}, \tag{7}$$

i.e., when an increase in resources generates a greater *proportional increase* in the success probability of *high-quality agents*.

Condition (7) occurs when quality and resources are "strong" complements in the sense that $\partial \ln g(q,r)/\partial r \partial q > 0$. Intuitively, this can happen when both high-quality and plentiful resources are necessary for success. For example, low-quality movies are very unlikely to become blockbusters even with a large budget (De Vany, 2004). Even high-quality movies, however, are unlikely to make it big unless supported by a hefty budget. In these cases, all agents who lack resources are likely to fail, and differences between agents only become apparent when resources are plentiful. In this situation, the probability of success depends on the weakest link: $g(q,r) = min(q,r)$. When agents have few resources, they all do poorly, and failure is not diagnostic of low quality. The quality of high-quality agents only starts to matter when sufficient resources are available. As a result, increased resources make a success more diagnostic.

5.2 A Model of Quality-Resource Interactions

To illustrate how inferences from success and failure depend on whether quality and resources are substitutes or complements, suppose the probability that an agent i, with quality i and resources $r_{i,t}$, succeeds in period t is

$$P_{i,t} = (wq_i^{\beta} + (1-w)r_{i,t}^{\beta})^{1/\beta}, \tag{8}$$

which is the well-known "constant elasticity of substitution" (CES) production function. This function is increasing in both q_i and $r_{i,t}$. The parameter $w \ in(0,1)$ reflects the importance of quality (when $w = 1$ only quality matters). The parameter β determines whether quality and resources are strong substitutes or complements. If $\beta > 0$ quality and resources are strong substitutes and if $\beta < 0$, then quality and resources are complements.[4] Moreover, the value of β determines whether $min(q_i, r_{i,t})$ or $max(q_i, r_{i,t})$ is given more weight. When $\beta > 1$, $max(q_i, r_{i,t})$ is given more weight.[5] It is as if the "strongest

[3]Strong in the sense that if condition (6) holds then $\partial g(q,r)/\partial r \partial q < 0$, but not the other way around.

[4]$\partial \ln P_{i,t}/\partial r \partial q = -b(q_i r_{i,t})^{b-1} w(1-w)\frac{1}{P_{i,t}^2}$.

[5]$P_{i,t}^{ces} \to max(q_i, r_{i,t}))$ as $\beta \to \infty$.

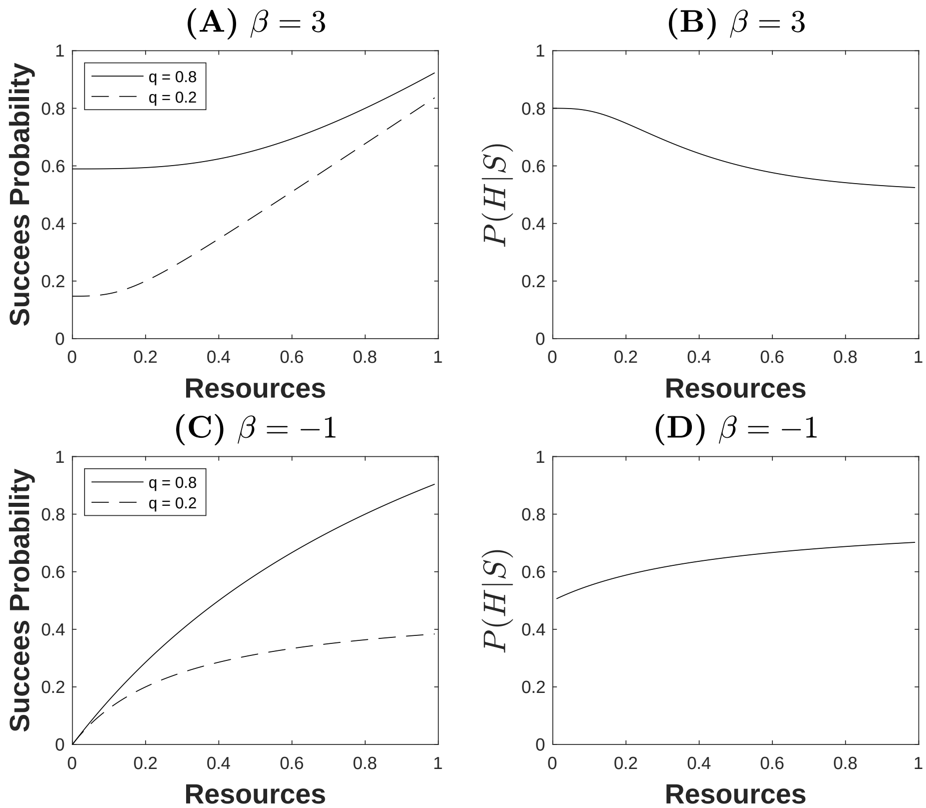

Figure 4: Top row: A) How the success probabilities for high- ($q = 0.8$) and low-quality ($q = 0.2$) agents vary with the amount of resources when $\beta = 3$ and $w = 0.4$. B) How the probability of being high quality given a success, $P(H|S)$, varies with resources when $\beta = 3$. Bottom Row: same for $\beta = -1$.

link" matters most. If $\beta < 1$, $min(q_i, r_{i,t})$ is given more weight.[6] It is as if the "weakest link" matters most.

Figure 4 illustrates the effect of β. Figure 4A shows how the success probabilities for high- ($q = 0.8$) and low-quality ($q = 0.2$) agents vary with the amount of resources when $\beta = 3$ and $w = 0.4$. When resources are plentiful, both high- and low-quality agents are likely to succeed. When resources are scarce, high-quality agents are significantly more likely to succeed than low-quality agents are. The implication is that success is not very diagnostic of high quality when resources are plentiful. Indeed, Figure 4B shows that when $r_{i,t} = 1$, $P(H|S)$ is close to 0.5. More generally, $P(H|S)$ is a decreasing function of resources. Figure 4C shows how the success probabilities vary with the amount of resources when $\beta = -1$. When resources are scarce, both high- and low-quality agents are likely to fail. When resources become plentiful, however, high-quality agents are

[6] $P^{ces}_{i,t} \to min(q_i, r_{i,t}))$ as $\beta \to -\infty$.

significantly more likely to succeed than low-quality agents are. As a result, increased resources make a success more diagnostic of high quality, i.e., $P(H|S)$ is an increasing function of resources (see Figure 4D).

The value of β determines whether a success followed by initial failure is more impressive than repeated success, because β impacts whether a past failure makes a subsequent success more or less diagnostic of high quality. To illustrate this effect of β, consider again a two-period model with two types of agents. Suppose that the high-quality agent has quality $q = 0.8$ and low quality $q = 0.2$. Suppose that the resources available at the start of period one are equal to $r_{i,1} = 0.5$, which increases to $r_{i,2} = 0.5 + \alpha$ after a success and decreases to $r_{i,2} = 0.5 - \alpha$ after a failure. The parameter α measures the impact of past outcomes on resources. If $\alpha = 0$, past outcomes do not change resources, while if $\alpha = 0.5$, success increases resources to one and failure decreases resources to zero. Figure 5 illustrates how the sequence of past outcomes changes inferences about quality. The upper row of Figure 5 shows the case where $\beta = 3$ and $w = 0.3$. In this case, succeeding after an initial failure is most impressive: $P(H|F, S) = 0.716$ while $P(H|S, S) = 0.602$. The reason is that success after initial failure, which depletes resources, is especially diagnostic of high quality, because the success probabilities of high- and low-quality agents differ more when resources are scarce (see the second column of Figure 5). The middle row in Figure 5 shows the case when $\beta = -3$ and $w = 0.3$, and quality and resources are complements. When $\beta = -3$, agents who failed after an initial success are less impressive (likely to be of high quality) than those who failed twice: $P(H|S, F) = 0.268$, while $P(H|F, F) = 0.389$. The reason is that the success probabilities of high- and low-quality agents differ substantially when resources are plentiful, while the success probabilities of both high- and low-quality agents are low if resources are scarce (see the second column of Figure 5). In both the above cases, succeeding under adversity was more impressive: $P(H|F, S) > P(H|S, F)$. The bottom row shows that when $\beta = -3$ and $w = 0.8$, we get the opposite result, $P(H|F, S) < P(H|S, F)$, for the same reasons as in Figure 2, i.e., the first outcome is too diagnostic of high quality (due to high w) to generate an inference reversal.

So far, we have shown results for only a few combinations of the parameters β and w, and for when there are two types of agents. Figures 6 and 7 show the results for all combinations of β and w (when $\alpha = 0.4$) when there are many types of agents. Specifically, the quality of an agent is drawn from a uniform distribution between zero and one. Figure 6 shows that succeeding after initial failure is more impressive than failing after initial success, $E[q|F, S] > E[q|S, F]$, for almost all values of w when $\beta > 0$.[7] Figure 6 shows the case when α, the parameter regulating the impact of past outcomes on resources, equals 0.4. Similar results hold when α is much smaller, such as $\alpha = 0.1$. The insets in Figure 6 show how the success probability varies with the resources when $\beta < 0$ and $\beta > 0$, which helps to understand the intuition. When $\beta > 0$, the inset shows that the success probabilities of high and low-quality agents differ substantially after failure. As a result, failing after success is diagnostic of high quality ($E[q|F, S]$ is high). On the contrary, after an initial success, the success probabilities of high- and low-quality agents are similar. As a result, succeeding after an initial success is not very diagnostic of high quality and $E[q|S, F]$ is close to $E[q|S]$ (i.e., we learn almost nothing new from the second outcome). When $\beta < 0$ and w is low, we also get $E[q|F, S] > E[q|S, F]$. The inset shows that when $\beta < 0$, the success probabilities of high- and low-quality agents after an initial success

[7] Except when $w = 1$ or $w = 0$.

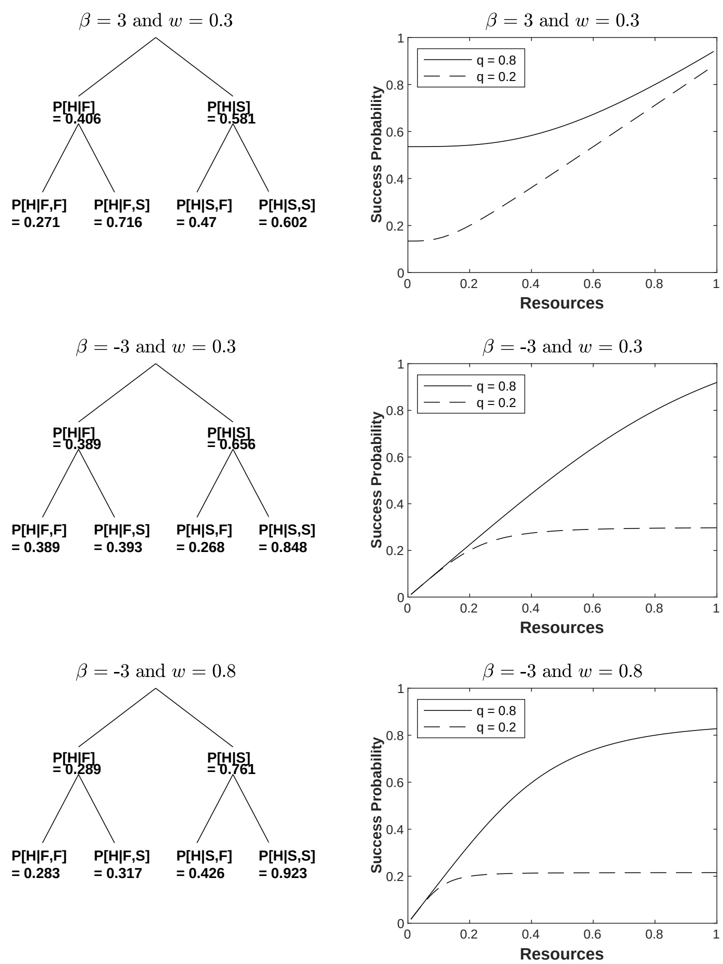

Figure 5: Top row: How the success probabilities for high- ($q = 0.8$) and low-quality ($q = 0.2$) varies with the amount of resources and the probability of being high quality when $\beta = 3$ and $w = 0.3$. Middle row: when $\beta = -3$ and $w = 0.3$. Bottom row: when $\beta = -3$ and $w = 0.8$. In all cases, $\alpha = 0.4$. Analyses are based on Bayes' rule given the success probabilities and qualities. The insets show how the success probability varies with the resources when $\beta < 0$ and $\beta > 0$.

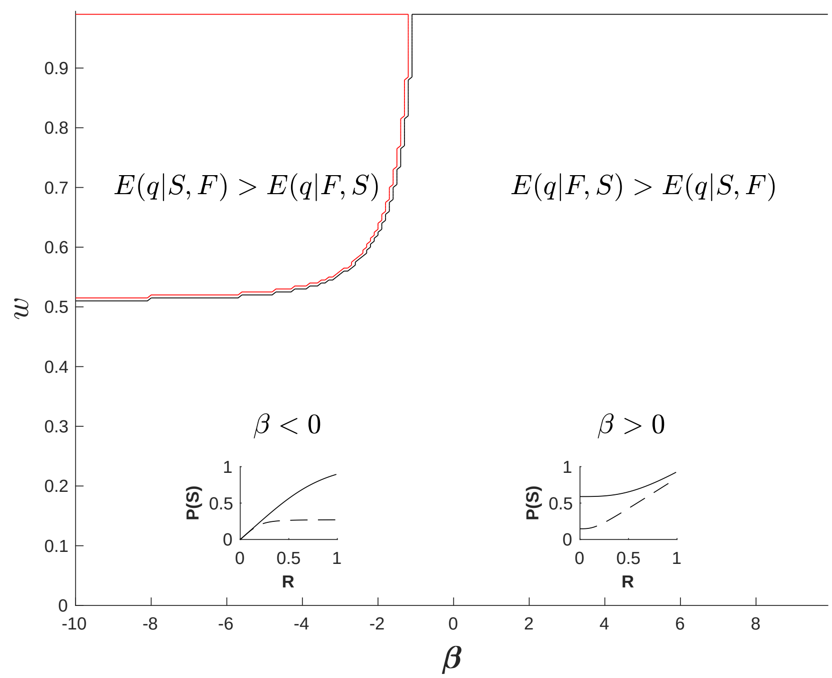

Figure 6: The values of β and w for which succeeding after initial failure is more impressive than failing after initial success, $E[q|F,S] > E[q|S,F]$. Based on numerical integration of the conditional densities (e.g., $E[q|F,S] = \int qf(q|F,S)/\int f(q|F,S))$, when $\alpha = 0.4$ and uniformly distributed quality). The insets show how the success probability varies with resources when $\beta < 0$ and $\beta > 0$.

differ substantially, and failure after an initial success is very diagnostic of low quality. As a result, $E[q|S,F]$ is substantially lower than $E[q|S]$. The success probabilities of high- and low-quality agents after failure, however, are similar, and succeeding after an initial failure is not very diagnostic of high quality. As a result, $E[q|F,S]$ is close to $E[q|F]$ (that is, we learn almost nothing new). When w is low (quality does not matter much), initial success or failure is not very diagnostic of quality, and both $E[q|S]$ and $E[q|F]$ are close to 0.5. Because $E[q|S,F]$ is substantially lower than $E[q|S]$, but $E[q|F,S]$ is close to $E[q|F]$, we get the result that $E[q|F,S] > E[q|S,F]$. When w is larger, $E[q|S]$ and $E[q|F]$ are further away from 0.5, and we get the opposite result: $E[q|S,F] > E[q|F,S]$.

Figure 7 shows when succeeding after initial failure is more impressive than succeeding twice, and when failing after initial success is less impressive than failing twice, when $\alpha = 0.4$ and $\alpha = 0.25$. A necessary condition for both of these effects is that quality impacts the success probability less than past outcomes do (the weight on quality, w, has to be less than 0.5). Succeeding after initial failure is more impressive than succeeding twice, $E[q|F,S] > E[q|S,S]$, only if quality and resources are (strong) substitutes ($\beta > 0$) and increased resources benefit low-quality agents proportionally more. In this case, succeeding after initial failure is very diagnostic of high quality, because the success probabilities of high- and low-quality agents differ a lot after failure. Failing after initial

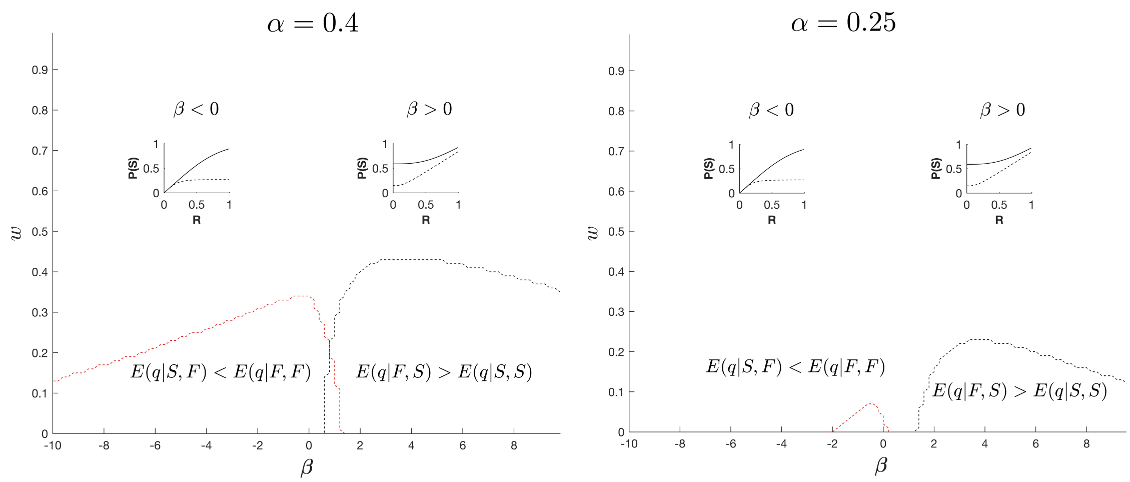

Figure 7: The values of β and w when succeeding after initial failure are more impressive than succeeding twice, and when failing after initial success is less impressive than failing twice, when $\alpha = 0.4$, $\alpha = 0.25$, and quality is drawn from uniform distribution. The numbers are based on numerical integration of the conditional densities (e.g., $E[q|F,S] = \int qf(q|F,S)/\int f(q|F,S))$. The insets show how the success probability varies with the resources when $\beta < 0$ and $\beta > 0$.

success is less impressive than failing twice, $E[q|S,F] < E[q|F,F]$, when quality and resources are complements ($\beta < 0$) and the increased resources benefit the high-quality agents proportionally more. In this case, failing after initial success is very diagnostic of low quality, because the success probabilities of high- and low-quality agents differ a lot after success. Both effects can occur simultaneously when β is close to one, but only if w is low.

5.3 Scope Conditions

A crucial assumption of the above model is that the agent started with a moderate amount of resources, $r_{i,1} = 0.5$, which then could increase (after success) or decrease (after failure). Suppose instead that the agent starts with low resources, and success increases resources, but failure does not decrease resources. Succeeding after failure could then never be more impressive than succeeding twice. The reason is that if the resources do not change after failure, succeeding after failure cannot be more diagnostic of high quality than the initial success is.

Another crucial assumption is that quality and resources combine additively and not multiplicatively. Suppose instead that quality and resources combine multiplicatively, following the Cobb-Douglas production function:

$$P^{cd}_{i,t} = q_i^a r_{i,t}^b.$$

The impact of an increase in resources is then proportional to the quality of an agent. This specification implies that resources do not change the diagnosticity of success. Regardless of the level of resources, the probability of being high quality given an observed success is the same. To see this, suppose we observe a success and want to infer whether the agent who succeeded was of high ($q = 0.8$) or low ($q = 0.2$) quality (with both levels of quality equally likely a priori). The probability that the agent was of high quality is, by Bayes rule,

$$P(H|S) = \frac{0.8^a r_{i,t}^b}{0.8^a r_{i,t}^b + 0.2^a r_{i,t}^b} = \frac{0.8^a}{0.8^a + 0.2^a},$$

which is independent of resources. Because resources do not impact inferences from success or failure, the expected quality of agents who succeed after failure is the same as those who fail after an initial success.

Finally, suppose resources and quality determine the success probability via the logistic choice model:

$$P_{i,t}^{log} = \frac{e^{c+aq_i+br_{i,t}+dq_i r_{i,t}}}{1 + e^{c+aq_i+br_{i,t}+dq_i r_{i,t}}},$$

where we assume that $a > 0$, $b > 0$, and $d < a$ (to ensure that higher-quality agents are more likely to succeed). If $d = 0$ (i.e., no interaction effect), succeeding after failure is never more impressive than succeeding twice. This is true even though decreasing resources does increase the diagnosticity of success. Thus, an agent who succeeds with fewer available resources is, everything else equal, more likely to be of high quality than an agent who succeeds with more resources available. The change in diagnosticity with decreased resources is not strong enough, however, to imply that succeeding after failure is more impressive than succeeding twice. When $d < 0$, implying that high-quality agents benefit less from increases in resources, succeeding after an initial failure can be more impressive than succeeding twice. For the same reason, when $d = 0$, failing after an initial success is never less impressive than failing twice, but when $d > 0$ (low-quality agents benefit less from increased resources), failing after an initial success can be less impressive than failing twice.

5.4 Discrete versus Continuous Outcomes

Our model focuses on binary outcomes (success / failure). Although several important outcomes are binary (e.g., get into an Ivy League school, reach an IPO), other outcomes can take on a range of values (reviewer scores) or are effectively continuous (market share, income). Some of these outcomes can be modeled as the result of a sequence of binary decisions: market share can be viewed as the outcome of a sequence of binary competitions, where you win or lose each customer. Our model is applicable to this set-up: the outcome in each period may impact the resources of the agent (a win increases resources, a loss decreases them). Simulations show that we get similar results in such a multi-period model: if resources can compensate for low quality ($\beta > 0$), then early failure indicates higher quality than early success, given the same level of eventual success. An analogous result can be demonstrated in a two-period model with continuous outcomes. Suppose the outcome in the first period is $P_{i,1} = q_i + e_{i,1}$ and the outcome in the second period is $P_{i,2} = bP_{i,1} + q_i + e_{i,2}$, where b is the impact of the past outcome and $e_{i,j}$ are normally distributed noise terms. When $b > 1$, then the expected level of quality, given total performance ($T_i = P_{i,1} + P_{i,2}$), is higher if the initial performance is low. Analyzing the case with continuous outcomes in more detail, however, is beyond the scope of this paper.

6 Empirical Illustration

6.1 A Relative Age Effect Reversal among National Hockey League Players

Our model shows that succeeding after initial failure can indicate higher expected quality when initial success significantly increases the chance of success. Professional sports provides an interesting context in which to examine whether initial failure can be associated with eventual success. Prior research has identified a relative age effect in sports. In hockey, football, and basketball, young players are organized by age groups. Players who are born right after the age group cutoff tend to receive favorable treatment, due to their relative physical maturity, an advantage that can have an enduring impact on their career development. For example, players born after December 31 are placed in a subsequent age cohort in the Canadian Junior Hockey League (CJHL). One should expect a uniform distribution of birth month among professional hockey players, as it should not influence people's inborn talent for hockey. But prior research has shown how this relative age difference generates an unrepresentative distribution of birth months among professional hockey players in Canada: almost 40% were born between January and March, and only around 10% were born between October and December (Barnsley Thompson, 1988; Musch Grondin, 2001). Subsequent studies have found a similar relative age effect in many other sports and countries (for a review, see Musch and Grondin 2001). The implication is that to be born at the right time and the right place (i.e., immediately after a selection system's cutoff) can increase the odds of success (i.e., becoming a professional player).

Our model shows that a relative age effect implies that players who make it as professionals despite being born at the wrong time and wrong place (i.e., immediately before a selection system's cutoff) may have superior skill, particularly compared to those who benefit most from the relative age effect (Fumarco et al., 2017). If this logic applies, professional players in the National Hockey League (NHL) from Canada who were born between October and December will have a higher expected skill than those born in earlier months of a year (who benefited most from a relative age effect). To examine this, we measure skills using players' average points (which combines scores and assists) and annual salary (market valuation of their contributions). Both are imperfect but useful measures of NHL players' realized skill. Additionally, we predict that Canadian NHL players who were not enrolled in CJHL are likely to have higher skill than those who were. These underdogs (without additional experience and coaching gained at CJHL) have to be very competent to become NHL players (Fumarco et al., 2017).

6.1.1 Data and Variables

A relative age effect reversal was demonstrated by Fumarco et al. (2017). Focusing on NHL elites, they found that players scoring in the top 10th percentile are much more likely to have been born immediately before the cutoff of December 31. Their data were from public sources, such as NHL websites, and the authors made their data and codes publicly available; it covered all NHL players for eight seasons between 2008 to 2016. We downloaded their data and first replicated their main results before extending it by testing our specific predictions.

In particular, we are interested in whether succeeding after initial failure predicts higher quality. All players in this dataset succeeded eventually, in the sense that they are all professional players in NHL. The challenge is how to operationalize "initial failures" and "quality."

We develop three proxies to represent initial failures. First, building on Fumarco et al. (2017) and the literature on the relative age effect, we used birth month quarters to measure the extent of initial failure or, more precisely, the disadvantage it creates. That is, in terms of physical maturity and the evaluation bias associated with it, being born between January and March is the most advantageous, whereas being born between October and December is the least advantageous, and the other two quarters fall in between. Nevertheless, birth months are an indirect indicator of failure and disadvantage. Our second measure is more direct: whether a Canadian NHL player was enrolled in CJHL (coded as 1=success) or not (coded as 0=failure). Most Canadian NHL players were also in the CJHL (92.5%). Our third measure combines the two above measures and considers players being both born between January and March and enrolled in the CJHL as being the most advantaged early on, and those born between October and December and not enrolled in the CJHL as being the least advantaged early on.

We have two proxies to represent quality. The first is the average points by each NHL player per season, and the second is their average annual salary. These two measures may represent these professional players' fully realized skill because insufficiently motivated or skilled players are unlikely to survive in this league (Teeselink et al., 2022). These two measures are also complementary, as one is the result of competition on the rink and the other is the result of competition in the labor market.

6.1.2 Results

The results, shown in Figure 8, show that an initial failure is associated with eventual success. Figures 8(C) and 8(D) replicate the results of Fumarco et al. (2017): the average points and salary of NHL players increase with birth month quarters. The only exception is that there is no significant difference in average points between players born between July and September and those born between October and December (Fumarco et al. also noted this).

Next, we examine whether not being enrolled in the CJHL predicts higher skill among Canadian NHL players. The results, as shown in Figures 8(A) and 8(B), show that succeeding (becoming an NHL player) after initial failure (not having been enrolled in CJHL) is associated with higher average scores and salary compared to those who succeeded in both.

Finally, we combine the two failure indicators: birth months and CJHL enrollment. The results, as shown in Figures 8(E) and 8(F), again show that initial failures combined with eventual success are associated with higher expected skill than repeated successes. In particular, having two initial setbacks (born between October and December and not enrolled in the CJHL) is associated with substantially higher average points and salary.

Our model shows that these results can occur without assuming that failure could increase motivation or skill, like related research on the underdog effect (Nurmohamed, 2020; Duckworth et al., 2007; Ashworth Heyndels, 2007). If selection to the CJHL is not very diagnostic of long-run potential and high-quality players are more likely to come back from an initial setback (i.e., not being selected), overcoming an initial setback is a more stringent test of high quality than selection into the CJHL is.

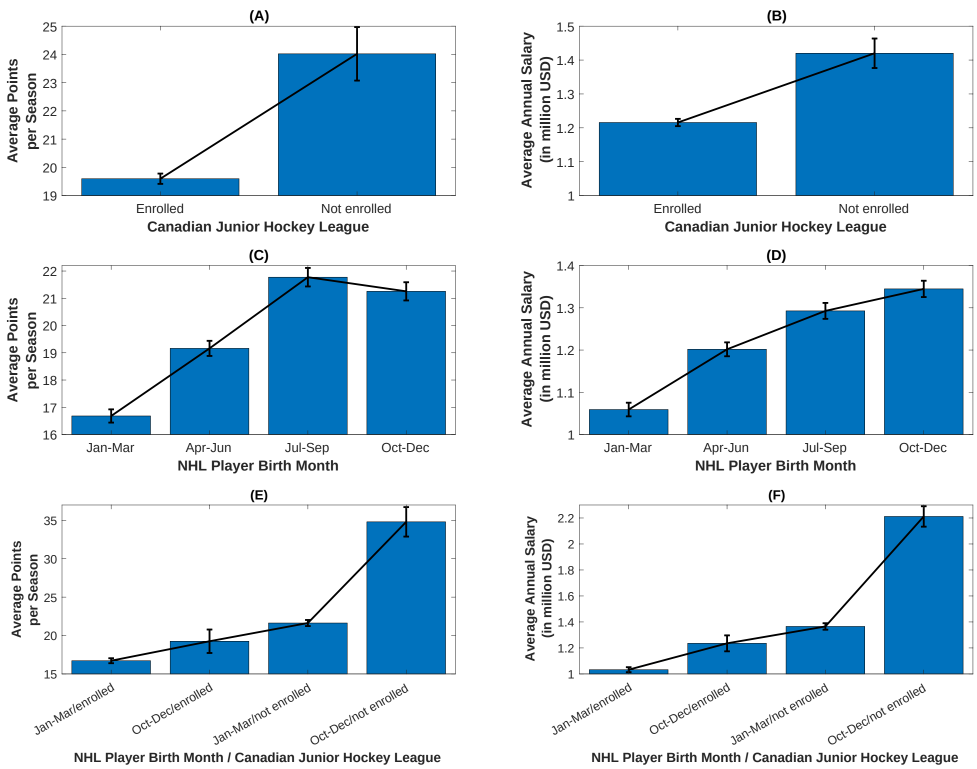

Figure 8: How NHL players' points and salary vary with their birth months and early achievement. Error bars represent standard errors.

6.2 The Underdog Effect in Music Downloads

While underdogs did better among Canadian NHL players, we cannot exclude the possibility that the mechanism is motivational. Early setbacks and competition with larger and better players may have motivated underdogs to improve more than others (Nurmohamed, 2020; Barnett, 1997). In addition, data is only available for players who continue, and differences in continuation rates can provide an alternative explanation (Cameron Heckman, 1998; Ashworth Heyndels, 2007).[8] Our second empirical illustration, using data from an experiment on music downloads (Salganik et al., 2006), excludes motivational accounts and any explanation that relies on changing traits, as "quality" does not vary over time in this data.

Salganik et al. (2006) set up a website (Music Lab) with 48 songs that participants could listen to and download. In one condition, participants independently listened to and downloaded songs without any information about the activities of other participants. The market share of downloads in this condition provides an estimate of the "quality" of a song–that is, the song-specific traits that influence the market share a song receives. Social influence was experimentally manipulated in the Salganik study by adding information on the downloading activities of other participants. In the weak social influence condition, information about how the number of others who had downloaded a particular song was available, but not particularly salient. In the strong social influence condition, information about downloads was made salient: the song that had been downloaded the most by prior participants was presented first on the web page, the second most popular was presented second, and so on. A song that became one of the most popular in the strong social influence condition was prominently displayed, and many participants were likely to listen to it and eventually download it. The experiment was repeated with eight different pool of participants (participant number ranges from 686 to 771), generating data from eight different realizations of the process (eight "worlds").

We use data from this experiment to examine the underdog effect. We predict that songs that were not displayed prominently early on but nevertheless became popular are of higher quality than songs that were prominently displayed early on and became popular. We expect this effect based on past work showing that medium-quality songs benefit more from early attention than high-quality songs do (Salganik et al., 2006; Hendricks et al., 2012), suggesting $\beta > 0$. The setup in the Salganik study differs from our two-period model, since total song downloads is the result of the decisions of many participants. Still, simulations show that a multi-period version of our model, where one participant makes a download choice in each period, predicts that a) market share early on is positively correlated with quality, b) final market share is positively correlated with quality, but c) early market share has a negative impact on predicted quality in a regression that includes both early and final market share.[9] The logic is as follows: when the eventual outcome is controlled for, having a lower early outcome is more impressive in the sense of being associated with higher predicted quality. Songs must be high quality to overcome the social influence that initially worked against them.

[8] If high-ability players are more likely to persist after failure.

[9] In the simulation, success increased resources, $r_{t+1} = r_t(1 - \lambda) + \lambda$, while failure decreased resources, $r_{t+1} = r_t(1 - \lambda)$. Initial resources were set at 0.5. The probability of success depended both on resources and quality and followed the CES production function (Equation 8). Whenever $\beta > 1$ and w was low enough, early success proportion had a negative impact on predicted quality in a regression that includes early and final success proportion.

Table 1: OLS Results for the Salganik Data, Predicting Quality*

Variable	Model 1	Model 2	Model 3
Final Market Share	0.1469		0.1774
Initial Market Share		0.0716	-0.0324
Constant	0.0178	0.0193	0.0178
N	348	348	348
R^2	0.31	0.12	0.32

*All variables significant at $p < 0.01$

6.2.1 Data

We use data from the strong social influence conditions. We downloaded Salganik et al.'s publicly available data from the lead author's website.[10] The data include 48 songs, and each song was present in eight worlds, so overall we have data on 348 song-world pairs. We calculated the final market share of each song in each world. As a measure of initial outcomes, we computed the market share of a song after 30 participants had made their choices (we demonstrate below that our results hold for choices other than 30). On average, there were about 700 participants in each world. Hence, the first 30 participants only represent 4% of the population. Still, this early choice proportion did impact subsequent choices by influencing the order in which songs were listed. Finally, we used the measure of song quality used by Salganik et al. (2006), which is the final market share of a song in the independent condition (in which download decisions were made independently, in contrast to the eight worlds where social influence was salient).

6.2.2 Results

Our focus is on how the final outcome (final market share) and the initial outcome (market share after 30 participants) predict quality. Table 1 shows the results of OLS regressions with quality as the dependent variable. The first model shows that the final market share is positively associated with quality, with a R^2 equal to 0.31. The second model shows that the initial market share is also positively associated with quality, with an R^2 equal to 0.12. The third model shows the estimated coefficients in a model that includes both the final and initial market share. In this model, the estimated effect of final market share remains positive. The estimated effect of initial market share, however, is now negative. Thus, controlling for the final market share, a high initial market share predicts lower song quality. This result supports our model prediction: succeeding (obtaining a similarly high eventual outcome) after initial failure (obtaining low choice proportion early on) is more impressive (having higher predicted quality) when early success enhances the chance of future success (due to strong social influence in this context).

How sensitive are these results to the choice of the early outcome cutoff (after 30 participants)? To examine this, we ran the same regression for all possible cut-offs of early outcomes, from the 6th to the 650th participant. The results show that the estimated coefficient is always negative, although the coefficient is not significantly different from zero for an early cutoff after 500 participants.

[10]http://www.princeton.edu/ mjs3/musiclab.shtml

7 Discussion

7.1 Three Performance Order Effects

Using a formal model and two empirical illustrations, we have demonstrated three types of performance order effects in inferences about quality from observed sequences of successes and failures: (i) a successful underdog (failure followed by success, FS) can be superior to a one-hit wonder (success followed by failure, SF); (ii) a successful underdog (FS) can also be superior to a successful favorite (SS); and (iii) a one-hit wonder (SF) can be inferior to a failed underdog (FF). We showed that such order effects can occur when past outcomes change how diagnostic success is, even if past performance does not causally impact quality, skill, or motivation. The key insight is that when (a) past outcomes change the resources available to agents, due to a Matthew effect (Merton, 1968), and (b) the availability of resources changes the relative probability that high-quality and low-quality agents will succeed, then past outcomes will change how diagnostic success is of high quality. Table 2 summarizes when each order effect occurs.

First, an underdog is superior to a one-hit wonder (FS > SF) when additional resources make subsequent success less diagnostic, but loss of resources makes a success more diagnostic of high quality. This happens when only high-quality actors can overcome an initial failure, whereas increased resources after an initial success increase all success probabilities to a relatively high level. Data from the Salganik et al. (2006) Music Lab experiment illustrated this effect: songs that were unpopular early on were of higher quality than songs that reached the same level of final market share but were popular early on.

Second, the more surprising order effect, FS > SS, occurs when resources strongly impact odds of success in such a way that additional resources increase the chances of success of even low-quality actors to a high level, while only high-quality actors can handle initial failure effectively. This order effect is illustrated by Canadian professional hockey players. Young players who were selected to the junior leagues gained more experience, resources, and connections than those who were not. The fact that a few of the underdogs managed to make it without such resources suggests that they had to be very good, as evidenced by the higher scoring capacity and higher salary.

The Music download data also provide support for a limited version of FS > SS. For example, according to the regression coefficients in Table 1, a moderately successful song that reached a final market share of 1.2 % (54 % reach this) but was unpopular early on (zero market share after 30 participants[11]) has higher predicted quality than a popular song, with final market share of 2.4 % (only 25 % reach this) and an initial market share of 6.7 % (only 10 % reach this). However, further analysis shows that the highest quality songs tend to have the highest final market share, with few exceptions, consistent with recent work showing that high quality eventually wins out in this setup regardless of early outcomes (Van de Rijt, 2019).

The third type of order effect (SF < FF) occurs when even high-quality actors cannot effectively deal with the loss of resources due to initial failure, while only high-quality agents can take advantage of increased resources due to initial success. Careers in academia provide a possible illustration. Graduates from elite Ph.D. programs, who experience more nurturing advisors and early publications in leading journals, receive better jobs with additional resources. Such resources do not necessarily compensate for

[11] 48 % of songs have zero market share initially.

Table 2: When do the three performance order effects occur?

Effect	Intuition	Resource-Quality Interactions	Impact of past outcomes
FS > SF	If failure reduces the success chances of low-quality agents to a low level, while high-quality agents' chances remain moderate, only high-quality agents have a chance to succeed after initial failure–i.e., "When the going gets tough, the tough gets going." Hence, succeeding after failure can be more diagnostic of high quality than an initial success is: (FS > S). If an initial success increases the odds of success of low-quality agents to almost the same level as high-quality agents, the outcome after an initial success is not diagnostic of quality. Thus, agents who failed after an initial success have a similar quality as those who succeeded initially (SF ≈ S).	Changes in resources impact low-quality agents proportionally more (resources and quality are strong substitutes, $\beta > 0$)	Some impact
FS > SS	If failure reduces the success chances of low-quality agents to a low level, while high-quality agents chances remain moderate, only high-quality agents have a chance to succeed after initial failure. Hence, succeeding after failure can be more diagnostic of high quality than an initial success is: (FS > S). If success benefits low-quality agents the most, then low-quality agents become almost as likely to succeed after an initial success as high-quality agents are. Hence, succeeding after a success is not diagnostic of quality, and SS is not much higher than S (SS ≈ S).	Changes in resources impact low-quality agents proportionally more (resources and quality are strong substitutes, $\beta > 0$)	Large impact
SF< FF	If only high-quality agents benefit from increased resources after an initial success, the odds of success for high-quality agents will be substantially higher than for low-quality agents after an initial success. Because low-quality agents are much more likely to fail, failing after an initial success is very diagnostic of low quality. As a result, failing after initial success can be less impressive than an initial failure (SF < F). If failure impacts high-quality agents the most, the odds of success of both low- and high-quality agents will be low after an initial failure, and both are likely to fail again. Hence, repeated failure is not diagnostic of low quality, and FF is not much lower than F (FF ≈ F).	Changes in resources impact high-quality agents proportionally more (resources and quality are strong complements, $\beta < 0$)	Large impact

lack of skill (particularly in less resource-intensive fields). On the other hand, even highly qualified graduates from low-prestige programs may be unlikely to produce much initially. Thus, a lack of output among those from elite programs (SF) may indicate less merit than a lack of output among graduates from less prestigious programs (FF).

Research on disaster dynamics provides another example where failure can trigger a cascade of errors that overwhelms even the best operator (Perrow, 1984). There is a long-standing debate in organization science about whether and when accidents should be attributed to the person or persons in charge, to the wider system, or to extenuating external circumstances (Desai, 2015; Haunschild Sullivan, 2002). Researchers in the Normal Accident camp have argued that a sequence of failures should be attributed to the system or external factors, instead of the operators, because a trivial initial error can trigger a cascade of errors in a tightly coupled system (Perrow, 1984; Rudolph Repenning, 2002). According to this perspective, failure is not informative of individual skill or effort. Researchers in the high-reliability camp have argued that even tightly coupled systems can perform reliably and be virtually failure-free if the individuals in charge are sufficiently mindful and resilient (Weick Roberts, 1993; Weick Sutcliffe, 2001). If this account is correct, the presence of any failure indicates that the operators were not sufficiently mindful.

Our results help to clarify when and why failure should or should not be attributed to individuals by showing how resources and system characteristics should impact the attribution processes. If we assume that the performances of tightly coupled systems are determined by the weakest links (that is, high skill cannot compensate for lack of resources), the least impressive performance may not be those with repeated failures (FF) but those who failed after initial success (SF). The time pressure following an initial series of failures can overwhelm even very skilled operators, making subsequent failures uninformative of skill (i.e., FF $\approx$ F). In contrast, failure after an initial success (SF) may be a cause for concern: the additional time available should allow high-skilled operators to avoid subsequent failures. This reasoning is consistent with Perrow's (1984) discussion of how initial failure can overwhelm operators. When discussing an explosion at the oil tanker Dauntless Colocotronis, Perrow argues that "operators of the system had no way of knowing that the very slight jar to the ship made a gash that would supply flammable or explosive substances to the pump and engine rooms" (1984: 74). Even a trained fire crew that eventually boarded drew the wrong conclusions about the causes and location of the explosions. Perrow concludes that in these cases, cascading errors result from "interactions that we as 'operators' could not anticipate or reasonably guard against" (1984: 75). Similarly, Rudolph Repenning (2002) stress that errors often lead to interruptions that reduce the time available to deal with further complications: "Accumulating interruptions also decrease the time available to process available stimuli and extract the pattern necessary to trigger inert knowledge" (p.10).

If the initial error does not overwhelm operators, however, mindful operators can anticipate the potential problems and have time to avert disaster, then a cascade of failures (FF) would be diagnostic of low skill to deal with interruptions, which is closer to the setting high-reliability theorists focus on (Weick Roberts, 1993; Weick Sutcliffe, 2001). In such a setting, however, our model also predicts that organizations with flawless records (SS) are less skilled in dealing with failure than those who experienced failure but dealt with it (FS), because succeeding after initial success does not provide additional information about an operator's skill (i.e., SS $\approx$ S). The ability to deal effectively with adverse circumstances can only be detected if the system or individual has been sufficiently "stress tested," which systems with flawless records have not (Kim Miner, 2007).

7.2 How Frequent, Large, and Persistent is the Effect?

Although the successful underdog effect is interesting, it is not frequent. When reinforcing processes are strong, success follows an initial success and failure follows an initial failure. Thus, the naïve heuristic "more is better" is seldom wrong because successful underdogs are rare. An evaluator who cares about identifying "diamonds in the rough," however, should consider successful underdogs. For example, 92.5% of all Canadian NHL players were in the junior league; only 7.5% failed to make it to the junior league. A team who ignores this category of players will likely still get good players. However, teams on the lookout for the best players, or for players who might be underestimated by others, should consider this category of players.

If successful underdogs are more impressive, how large is the effect? The model shows that the effect is large only if a) past outcomes strongly impact future resources and success probabilities, and b) this changes the diagnosticity of success. Among Canadian hockey players, the effect is sizable: players who made it to the NHL without making it to the junior league have 11% higher salary (0.1 standard deviation) and 23 higher average points (0.22 standard deviation). In this empirical setting, early success probably changed access to resources considerably. In the Music Lab experiment, early success did not change access to resources, but only attention (that is, the order in which songs were displayed changed). Still, the effect size is noticeable: a one standard deviation increase in initial market share (measured at participant no. 30) leads to a 0.16 standard deviation reduction in predicted quality. This can be compared to the effect of a one standard deviation increase in the final market share being associated with a 0.67 standard deviation increase in predicted quality.

The underdog effect is unlikely to persist if underdogs are eventually able to display their talents. For example, studies show that quality eventually wins out in the Music Lab experiment (Van de Rijt, 2019) if the experiment is continued with more participants. High-quality songs are eventually noted and rise in popularity. At this point, popularity provides an accurate signal of quality, and initial disadvantage will be insignificant (because final popularity is a sufficient statistic). Similarly, if resources become more equal for hockey players who reach the NHL, the ordering effect may disappear eventually for experienced players that have played under similar conditions for many seasons.

7.3 Relation to Past Work

Prior research on how reinforcing mechanisms can complicate inferences has shown that strong reinforcements lead to a weak association between outcome and quality (Gould, 2002; Lynn et al., 2009). If the strength of reinforcing mechanisms is also uncertain, Denrell and Liu (2012) show that very high outcomes can signal lower quality than moderately high outcomes. The reason is that extreme outcomes indicate strong reinforcing processes, and high outcomes are less impressive if reinforcing processes are strong because even a low-quality agent who happens to succeed initially may then repeatedly succeed.

Our results are different because we focus on the order of outcomes. The mechanism underlying our results is also different and more generally applicable. The Denrell and Liu (2012) mechanism assumes that evaluators rely on the outcome to make inferences about the strength of the reinforcing processes. This mechanism is not applicable when there is little uncertainty about the strength of reinforcing processes. Consider Canadian hockey players. For the Denrell and Liu (2012) mechanism to be applicable, there must be substantial variation in how players react to past success and failure: some players are much more disadvantaged by failure and benefit much more from success than others. While some variation is natural, it is likely small. If most players are subject to roughly the same reinforcing process, the Denrell and Liu (2012) mechanism is not applicable. In contrast, the mechanism developed in this paper is applicable: only highly skilled players may be able to overcome adversity. Similarly, our mechanism, but not Denrell and Liu's (2012), could explain why a start-up that succeeds after initial failure is more impressive than a start-up with repeated success in the same industry (with a similar strength of reinforcing processes).

Our finding is consistent with but goes beyond the literature on discounting in causal attribution, which argues that one should discount a cause of an event when an alternative cause is present (Morris Larrick, 1995; Miklós-Thal Zhang, 2013). When reinforcing processes are strong, past success or failure provides an alternative cause of success, different from quality. In our model, however, the two causes (past outcome and quality) are not independent. It is intuitively clear that one should discount cause A if B is also present (e.g., a student succeeded on the exam but was helped by another student). Prior formal analyses of discounting have focused on this case of independent causes (Morris Larrick, 1995; Miklós-Thal Zhang, 2013). It is less intuitively obvious whether to discount high quality as a reason for high performance when the alternative cause (high past performance, for example) also reflects high quality.

Finally, our account of why underdogs can be superior is different from previous motivational and learning accounts. The motivational account emphasizes that underdogs persevere (Duckworth et al., 2007), striving to prove others wrong (Goldschmied Vandello, 2012; Nurmohamed, 2020; Vandello et al., 2007), while success reduces motivation (Audia et al., 2000). The learning account argues that failure can be informative (Kim et al., 2009; Stan Vermeulen, 2013) and that struggle and rivalry lead to improvements (Barnett, 1997; Ashworth Heyndels, 2007). Both accounts assume that failure has a causal effect: agents' traits change after failure. Our account is not causal: we do not assume that failure changes motivation or skill. In our account, the outcome after failure is a better signal of existing traits, such as quality.

8 Implications: Searching for Hidden Gems

Our findings show that a "more-is-better" heuristic may underestimate the quality of individuals and products that succeed in adverse circumstances and those who fail after an initial failure. This will lead to missed opportunities and unfair evaluations if managers fail to consider whether success or failure is diagnostic of quality. Indeed, a large literature in social psychology suggests that people often do not properly account for situational influences. People tend to "take high nominal performance as evidence of high ability and do not discount it by the ease with which it was achieved" (Swift et al., 2013, p.1). To be sure, sophisticated investors and hiring committees understand the need to evaluate candidates based on information other than past performance. Most people also take into account well-known mitigating circumstances (e.g., illness, bias in favor of particular groups). This is consistent with recent research showing that although there are large individual differences, people generally take into account important situational factors but tend to ignore less known or more subtle situational factors (Scopelliti et al., 2018; Zunino et al., 2021). For example, most sports managers ignore the relative age effect in the NHL (Deaner et al., 2013). This suggests that people may fail to appreciate the order effects discussed here.

If people do not take into account the diagnosticity of outcomes, people and products will be under- or overestimated, which can create "behavioural opportunities" for discerning strategists to exploit (Denrell et al., 2019). More generally, it leads to unfair evaluations and inefficient allocations. For example, suppose that the new fund that Grace or Robin will lead does not benefit from the reinforcing processes that helped Robin to succeed in the last three years. The choice of fund manager then only depends on an assessment of their investing capabilities. Our model shows that Grace may be the best person to lead the new fund if an initial failure impacts the odds of success of low-quality fund managers more than it impacts high-quality managers. In this case, succeeding after failure can be highly diagnostic of high quality. However, the outcome after initial failure is not always more diagnostic of quality. If failure impacts high-quality more than low-quality agents (e.g., when an initial failure reduces the odds of success of all fund managers to a low level), succeeding after initial failure is relatively uninformative about quality.

How, then, can managers correctly assess the diagnosticity of success and failure and detect hidden gems? One possibility is to change how resources are allocated to better estimate quality. If all agents are given equal resources, independent of past outcomes, the diagnosticity of success and failure does not change over time, and a "more-is-better" heuristic is valid. Resources can also be reallocated from the "rich" (who succeeded in the past) to the "poor" (who failed) to identify the role of resources. For example, Salganik Watts (2008) examined the impact of social influence by manipulating download data, so that the least popular songs early on were shown to be the most popular. If the impact of resources is known, and the objective is to select the better individual, Meyer (1991) showed that the opposite reallocation strategy, giving resources from the "poor" to the "rich," is superior because it provides opportunities for individuals who failed to demonstrate their talents. The intuition is that only if the circumstances of the failed individual or products are sufficiently adverse can a success after a failure change the decision of whom to select. Consistent with this, Salganik Watts (2008) showed that only the highest quality songs can recover their popularity in the long run, despite having been falsely displayed as the least popular songs early on.

Experimenting with reallocating resources is costly if managers believe that past success indicates high quality and resources should be allocated only to high-quality individuals. An alternative approach is to use data on past outcomes to estimate the strength of reinforcing processes. For example, if data on past outcomes, their order, and a proxy for quality are available, the parameters (β,α,w) of the model in Section 5 can be estimated and used to predict quality based on past outcomes for new candidates. Romann Cobley (2015) show that such a regression approach can correct for the relative age effect in Swiss junior sprinting. Universities in the United Kingdom use a similar approach to provide what is known as contextual offers, which allow lower grades for admission for applicants from backgrounds and schools with few university graduates. A regression approach requires a proxy of quality, which may not be available if only outcomes (success or failure) are observed. It is possible to estimate quality, however, by looking at the success rate in a new round of competitions that are not (much) influenced by past outcomes. Our analysis of the NHL data was built on this assumption: we assumed that scores and salaries in the NHL reflected quality and were largely independent of early outcomes. Similarly, universities in the United Kingdom evaluate the suitability of contextual offers by examining data on university grades.[12] A regression approach also requires a good understanding of the appropriate functional form and high-quality data, and researchers have argued that in the absence of these, relying on qualitative evaluations is superior and fairer (Gorard, 2010). For similar reasons, applying a more-is-better heuristic can be more appropriate than trying to correct for the impact of past outcomes, particularly when evaluators are not certain about how quality and resources interact.

This discussion, combined with our findings, suggests that searching for hidden gems is possible but challenging. The more challenging the search process is, however, the more hidden these gems are to most managers (Denrell et al., 2019; Liu, 2021). This preserves them as attractive behavioral opportunities for informed managers who do not take outcomes at face value but evaluate them by their diagnosticity.

References

Arthur B (1989) Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal* 99(394):116–131.

Ashworth J, Heyndels B (2007) Selection bias and peer effects in team sports: The effect of age grouping on earnings of german soccer players. *Journal of sports Economics* 8(4):355–377.

Audia PG, Greve HR (2021) *Organizational learning from performance feedback: A behavioral perspective on multiple goals: A multiple goals perspective* (Cambridge University Press).

Audia PG, Locke EA, Smith KG (2000) The paradox of success: An archival and a laboratory study of strategic persistence following radical environmental change. *Academy of Management Journal* 43(5):837–853.

Baker G, Gibbons R, Murphy KJ (1994) Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics* 109(4):1125–1156.

Baker GP, Jensen MC, Murphy KJ (1988) Compensation and incentives: Practice vs. theory. *The journal of Finance* 43(3):593–616.

Barnett WP (1997) The dynamics of competitive intensity. *Administrative Science Quarterly* 128–160.

Barnett WP, Hansen MT (1996) The red queen in organizational evolution. *Strat. Mgmt. J.* 17(S1):139–157.

Barnsley RH, Thompson AH (1988) Birthdate and success in minor hockey: The key to the nhl. *Canadian Journal of Behavioural Science* 20(2):167–176.

[12] For example, Bristol University says on its webpage that "data shows that our contextual offer students achieve above average academically." (http://www.bristol.ac.uk/study/undergraduate/entry-requirements-qualifications/contextual-offers/).

Benson A, Li D, Shue K (2019) Promotions and the Peter principle. *The Quarterly Journal of Economics* 134(4):2085–2134, publisher: Oxford University Press.

Bol JC, Smith SD (2011) Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review* 86(4):1213–1230.

Cameron SV, Heckman JJ (1998) Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political economy* 106(2):262–333.

Castilla EJ (2008) Gender, race, and meritocracy in organizational careers. *American Journal of Sociology* 113(6):1479–1526.

Chatterji A, Delecourt S, Hasan S, Koning R (2019) When does advice impact startup performance? *Strategic Management Journal* 40(3):331–356.

De Vany A (2004) *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry* (New York, NY: Routledge).

Deaner RO, Lowen A, Cobley S (2013) Born at the wrong time: selection bias in the nhl draft. *PLoS One* 8(2):e57753.

Denrell J, Fang C, Liu C (2019) In search of behavioral opportunities from misattributions of luck. *Academy of Management Review* 44(4):896–915.

Denrell J, Liu C (2012) Top performers are not the most impressive when extreme performance indicates unreliability. *Proceedings of the National Academy of Sciences* 109(24):9331–9336.

Desai V (2015) Learning through the distribution of failures within an organization: Evidence from heart bypass surgery performance. *Academy of Management Journal* 58(4):1032–1050.

Dörner D (1996) *The Logic of Failure: Strategic Thinking for Complex Situations* (Metropolitan Books, NY: New York).

Duckworth AL, Peterson C, Matthews MD, Kelly DR (2007) Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology* 92(6):1087–1101, ISSN 1939-1315.

Fischhoff B, Beyth-Marom R (1983) Hypothesis evaluation from a bayesian perspective. *Psychological Review* 90(3):239.

Frank RH, Cook PJ (1995) *The Winner-Take-All Society: Why the Few at the Top Get So Much More Than the Rest of Us* (New York, NY: The Free Press).

Fumarco L, Gibbs BG, Jarvis JA, Rossi G (2017) The relative age effect reversal among the national hockey league elite. *PloS One* 12(8):1–16.

Goldschmied NP, Vandello JA (2012) The Future is Bright: The Underdog Label, Availability, and Optimism. *Basic and Applied Social Psychology* 34(1):34–43, ISSN 0197-3533, URL `http://dx.doi.org/10.1080/01973533.2011.637726`, publisher: Routledge _eprint: https://doi.org/10.1080/01973533.2011.637726.

Gorard S (2010) Serious doubts about school effectiveness. *British Educational Research Journal* 36(5):745–766.

Gould R (2002) The origins of status hierarchies: A formal theory and empirical test. *American Journal of Sociology* 107(5):1143–1178.

Guler I (2007) Throwing good money after bad? political and institutional influences on sequential decision making in the venture capital industry. *Administrative Science Quarterly* 52(2):248–285.

Haunschild PR, Sullivan BN (2002) Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Administrative science quarterly* 47(4):609–643, publisher: SAGE Publications.

Hendricks K, Sorensen A, Wiseman T (2012) Observational learning and demand for search goods. *American Economic Journal: Microeconomics* 4(1):1–31.

Holmstrom B (1979) Moral hazard and observability. *Bell Journal of Economics* 10(1):74–91.

Holmstrom B (1982) Moral Hazard in Teams. *The Bell Journal of Economics* 13(2):324–340.

Holmström B (2017) Pay for performance and beyond. *American Economic Review* 107(7):1753–77.

Jenter D, Kanaan F (2015) CEO turnover and relative performance evaluation. *the Journal of Finance* 70(5):2155–2184.

Karlin S, Rubin H (1956) The theory of decision procedures for distributions with monotone likelihood ratio. *The Annals of Mathematical Statistics* 27(2):272–299.

Kim JY, Kim JY, Miner AS (2009) Organizational learning from extreme performance experience: The impact of success and recovery experience. *Organization Science* 20(6):958–978.

Kim JY, Miner AS (2007) Vicarious learning from the failures and near-failures of others: Evidence from the us commercial banking industry. *Academy of Management Journal* 50(3):687–714.

Lieberman M, Montgomery D (1988) First-mover advantages. *Strategic Management Jounral* 9(S1):41–58.

Liu C (2021) Why do firms fail to engage diversity? a behavioral strategy perspective. *Organization Science* 32(5):1193–1209.

Lynn FB, Podolny JM, Tao L (2009) A sociological (de) construction of the relationship between status and quality. *American Journal of Sociology* 115(3):755–804.

Merton RK (1968) The Matthew effect in science. *Science* 159(3810):56–63.

Meyer M (1991) Learning from coarse information: Biased contests and career profiles. *Review of Economic Studies* 58(1):15–41.

Miklós-Thal J, Zhang J (2013) (de) marketing to manage consumer quality inferences. *Journal of Marketing Research* 50(1):55–69.

Morris MW, Larrick RP (1995) When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review* 102(2):331–355.

Musch J, Grondin S (2001) Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review* 21(2):147–167.

Nurmohamed S (2020) The underdog effect: When low expectations increase performance. *Academy of Management Journal* 63(4):1106–1133.

Ouchi WG, Maguire MA (1975) Organizational control: Two functions. *Administrative Science Quarterly* 559–569.

Perrow C (1984) *Normal Accidents: Living with High-risk Technologies* (New York, NY: Basic Books).

Pollock TG, Lee PM, Jin K, Lashley K (2015) (Un) tangled: Exploring the asymmetric coevolution of new venture capital firms' reputation and status. *Administrative Science Quarterly* 60(3):482–517, publisher: Sage Publications Sage CA: Los Angeles, CA.

Puffer SM, Weintrop JB (1991) Corporate performance and ceo turnover: The role of performance expectations. *Administrative Science Quarterly* 36(1):1–19.

Rahmandad H, Gary MS (2020) Delays Impair Learning and Can Drive Convergence to Inefficient Strategies. *Organization Science* ISSN 1047-7039, URL `http://dx.doi.org/10.1287/orsc.2020.1405`, publisher: INFORMS.

Romann M, Cobley S (2015) Relative age effects in athletic sprinting and corrective adjustments as a solution for their removal. *PLoS One* 10(4):e0122988.

Ross L, Nisbett RE (1991) *The person and the situation* (New York, NY: McGraw Hill).

Rudolph JW, Repenning NP (2002) Disaster dynamics: Understanding the role of quantity in organizational collapse. *Administrative science quarterly* 47(1):1–30.

Salganik M, Dodds P, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.

Salganik MJ, Watts DJ (2008) Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social psychology quarterly* 71(4):338–355.

Scopelliti I, Min HL, McCormick E, Kassam KS, Morewedge CK (2018) Individual differences in correspondence bias: Measurement, consequences, and correction of biased interpersonal attributions. *Management Science* 64(4):1879–1910.

Siegel J, Pyun L, Cheon B (2019) Multinational firms, labor market discrimination, and the capture of outsider's advantage by exploiting the social divide. *Administrative Science Quarterly* 64(2):370–397.

Sørensen M (2007) How smart is smart money? a two-sided matching model of venture capital. *The Journal of Finance* 62(6):2725–2762.

Stan M, Vermeulen F (2013) Selection at the gate: Difficult cases, spillovers, and organizational learning. *Organization Science* 24(3):796–812.

Strang D, Patterson K (2014) Asymmetries in experiential and vicarious feedback: Lessons from the hiring and firing of baseball managers. *Sociological Science* 1(1):178–198.

Swift SA, Moore DA, Sharek ZS, Gino F (2013) Inflated Applicants: Attribution Errors in Performance Evaluation by Professionals. *PLOS ONE* 8(7):e69258, ISSN 1932-6203, URL http://dx.doi.org/10.1371/journal.pone.0069258, publisher: Public Library of Science.

Teeselink BK, van den Assem MJ, van Dolder D (2022) Does losing lead to winning? an empirical analysis for four sports. *Management Science* .

Van de Rijt A (2019) Self-correcting dynamics in social influence processes. *American Journal of Sociology* 124(5):1468–1495.

Vandello JA, Goldschmied NP, Richards DAR (2007) The Appeal of the Underdog. URL https://journals.sagepub.com/doi/abs/10.1177/0146167207307488.

Weick KE, Roberts KH (1993) Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly* 38(3):357–381.

Weick KE, Sutcliffe KM (2001) *Managing the Unexpected* (San Francisco, CA: Jossey-Bass).

Wellman MP, Henrion M (1993) Explaining'explaining away'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(3):287–292.

Zunino D, Dushnitsky G, van Praag M (2021) How Do Investors Evaluate Past Entrepreneurial Failure? Unpacking Failure Due to Lack of Skill versus Bad Luck. *Academy of Management Journal* ISSN 0001-4273, URL http://dx.doi.org/10.5465/amj.2018.0579, publisher: Academy of Management.

Appendix A

The probability that an agent who fails and then succeeds is of high quality is, using Bayes' rule,

$$Pr(H|F,S) = \frac{p_{h,F}(1-p_h)0.5}{p_{h,F}(1-p_h)0.5 + p_{l,F}(1-p_l)0.5} = \frac{1}{1 + \frac{p_{l,F}(1-p_l)}{p_{h,F}(1-p_h)}}.$$

Similarly, the probability that an agent who succeeds and then fails is of high quality is

$$Pr(H|S,F) = \frac{(1-p_{h,S})p_h 0.5}{(1-p_{h,S})p_h 0.5 + (1-p_{l,S})p_l 0.5} = \frac{1}{1 + \frac{(1-p_{l,S})p_l}{(1-p_{h,S})p_h}}.$$

$Pr(H|F,S)$ is larger than $Pr(H|S,F)$ whenever

$$\frac{p_{l,F}(1-p_l)}{p_{h,F}(1-p_h)} < \frac{(1-p_{l,S})p_l}{(1-p_{h,S})p_h}, \implies \frac{(1-p_{h,S})p_h}{(1-p_{l,S})p_l} < \frac{p_{h,F}(1-p_h)}{p_{l,F}(1-p_l)}. \tag{9}$$

The probability that an agent who succeeds, and then succeeds again, is of high quality is

$$Pr(H|S,S) = \frac{p_{h,S}p_h 0.5}{p_{h,S}p_h 0.5 + p_{l,S}p_l 0.5} = \frac{1}{1 + \frac{p_{l,S}p_l}{p_{h,S}p_h}}.$$

Thus, $Pr(H|F,S)$ is larger than $Pr(H|S,S)$ whenever

$$\frac{p_{l,F}(1-p_l)}{p_{h,F}(1-p_h)} < \frac{p_{l,S}p_l}{p_{h,S}p_h} \implies \frac{p_{h,S}p_h}{p_{l,S}p_l} < \frac{p_{h,F}(1-p_h)}{p_{l,F}(1-p_l)}. \tag{10}$$

By symmetry, we have (using the second inequality in equation 10) that $Pr(L|S,F)$ is larger than $Pr(L|F,F)$ whenever

$$\frac{q_{l,F}}{q_{h,F}}\frac{q_l}{q_h} < \frac{q_{l,S}}{q_{h,S}}\frac{1-q_l}{1-q_h}. \tag{11}$$

Appendix B

Suppose that failure decreases resources from r_2 to $r_1 < r_2$, while success increases resources to $r_3 > r_2$. Using the first inequality in equation 10, $Pr(H|F,S)$ is larger than $Pr(H|S,S)$ whenever

$$\frac{g(l,r_1)}{g(h,r_1)}\frac{1-g(l,r_2)}{1-g(h,r_2)} < \frac{g(l,r_3)}{g(h,r_3)}\frac{g(l,r_2)}{g(h,r_2)} \implies \frac{g(h,r_2)}{1-g(h,r_2)}\frac{g(l,r_1)}{g(h,r_1)} < \frac{g(l,r_2)}{1-g(l,r_2)}\frac{g(l,r_3)}{g(h,r_3)}. \tag{12}$$

Because $g(h,r_2) > g(l,r_2)$ implies that $g(h,r_2)/(1-g(h,r_2))$ is larger than $g(l,r_2)/(1-g(l,r_2))$, 12 is only satisfied if $\frac{g(l,r_3)}{g(h,r_3)} > \frac{g(l,r_1)}{g(h,r_1)}$, i.e., if $\frac{g(l,r_3)}{g(l,r_1)} > \frac{g(h,r_3)}{g(h,r_1)}$, which is thus a necessary condition for $Pr(H|F,S) > Pr(H|S,S)$.