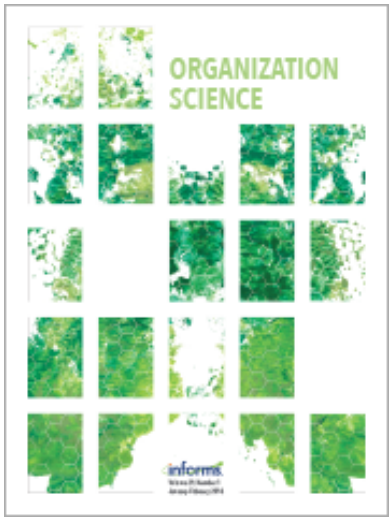


Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

When Local Learning Scales: Entrepreneurs' Initial Users and Market Expansion

Nataliya Langburd Wright

To cite this article:
Nataliya Langburd Wright (2026) When Local Learning Scales: Entrepreneurs' Initial Users and Market Expansion. Organization Science 37(1):377-401. https://doi.org/10.1287/orsc.2023.17983

https://pubsonline.informs.org/journal/orsc

ORGANIZATION SCIENCE
Vol. 37, No. 1, January–February 2026, pp. 377–401
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

When Local Learning Scales: Entrepreneurs' Initial Users and Market Expansion

Nataliya Langburd Wright[a]

[a] Columbia Business School, Columbia University, New York, New York 10027
Contact: nlw2136@columbia.edu, https://orcid.org/0000-0002-0245-0035 (NLW)

Received: July 26, 2023
Revised: May 21, 2024; January 31, 2025; June 3, 2025; September 1, 2025
Accepted: September 4, 2025
Published Online in Articles in Advance: November 17, 2025

https://doi.org/10.1287/orsc.2023.17983

Abstract. Entering new markets is crucial for technology startups to scale, yet these ventures often face high uncertainty about demand in these markets. This study examines how the composition of initial users shapes startups' new market growth amid such uncertainty. It theorizes that startups face a learning tradeoff when targeting a foreign market: Local initial users, who are more familiar to the startups, provide clearer signals due to shared language and norms; however, more representative foreign users provide more transferable insights about the target market. Using variation in feature timing on a product platform, this study finds that startups with a higher share of local initial users prior to platform launch experience greater growth in foreign users afterward. This effect is stronger when startups are based in more linguistically homogeneous countries and operate in globally standardized digital product categories, settings where the clarity benefits of local signals outweigh their transferability costs. In contrast to prior work suggesting that startups should begin with target users, this study reveals contexts when more familiar initial users can better foster new market growth and prevent premature scaling.

Supplemental Material: The online appendix is available at https://doi.org/10.1287/orsc.2023.17983.

Keywords: entrepreneurial strategy • scaling • entrepreneurship • learning • growth • early adopters • global • technology ventures

1. Introduction

Entering new markets is crucial for technology startups to scale. Doing so allows startups to expand their user base to achieve economies of scale (Giustiziero et al. 2023), leverage network effects (Boudreau et al. 2022), and signal their potential to investors (Cao et al. 2024), ultimately enabling them to achieve high-growth outcomes like initial public offerings or acquisitions (Conti and Guzman 2023). A common way that startups expand markets is across country borders (Bingham and Eisenhardt 2011, Reuber and Fischer 2011, Wormald et al. 2021, Tippmann et al. 2023). Indeed, international expansion has fueled the growth of technology startups such as Airbnb (United States), Canva (Australia), and Grammarly (Ukraine). In particular, artificial intelligence startups earn over half their revenue overseas (Murgia 2024), and startups with mostly foreign customers are more than twice as likely to surpass $50 million in valuation (Startup Genome 2024).

However, with little or no selling history, startups often face high uncertainty around demand in these foreign markets (McMullen and Shepherd 2006, Kerr et al. 2014, Gans et al. 2019, Moeen et al. 2020). This uncertainty makes it difficult to "plan" their product for this target market (Moore 2014). And so instead of planning ahead, startups often learn about this foreign market's demand experientially from their initial users (Ott et al. 2017). These initial users—whether sourced from experimentation, organic user demand, or a variety of other ways—can help startups refine and update their products to then grow their foreign user base.

Startups' initial users can come from local or foreign locations.[1] Local initial users and startups tend to be more familiar with one another. By coming from the same market, they often share a common background, such as language, making these local users' signals clearer, that is, less prone to communication frictions. This clarity enables startups to understand the initial users' value of demand. But these local signals might be idiosyncratic and unrepresentative of the foreign target market, making them less transferable, that is, less relevant to that target market because the value of demand may be different between the local initial and foreign target users. Foreign initial users are inherently more representative of foreign target ones and therefore can offer more transferable signals to the startup about the

target market. But these transferable signals can come at the cost of clarity because the startup is less likely to share a common background with these users. This tradeoff between clarity and transferability makes it difficult to anticipate which initial user composition can help startups build better products for their foreign target market and thereby achieve higher foreign user growth.

Such a tradeoff raises the question: How does the local (versus foreign) composition of initial users shape startups' subsequent foreign user growth? This question is crucial because it sheds light on how initial users shape the ability of startups to achieve target user growth—an important factor for future funding and performance (Cao et al. 2024). But this question is also difficult to answer because startups typically choose *when* to scale and reach their target market (Lee and Kim 2024). Notably, this makes it hard to disaggregate the role of learning from initial users from the timing of the expansion. The characteristics of startups that have local versus foreign initial users might also be fundamentally different, for example, in terms of the markets they target, and these characteristics may independently impact subsequent foreign user growth.

The study addresses this question by using data from a product platform on which many startups in the information technology (IT) sector initially launch their products to a mass set of foreign users. Crucially, the platform targets early-stage ventures and typically exercises discretion over the exact date of their launch. Thus, the timing of the launch occurs at a similarly early, but not specifically chosen, time frame within each firm's life cycle. The approach provides plausibly exogenous variation in the timing of launch to better isolate the role of learning from initial users from the startup's choice of when to launch. Moreover, the platform guidelines specifically attract early-stage technology startups seeking to grow their broader, particularly foreign, user base. These guidelines reduce differences in the types of startups featured, regardless of their initial user composition. Such unique platform characteristics enable the study to assess whether startups with a higher share of local (versus foreign) country initial users achieve higher subsequent foreign user growth after featuring on the platform.

The platform data enable the creation of a panel of 1,106 startups that launch their first product on the platform to grow their target user base in 2018–2020, matched with longitudinal data on their website traffic. By selecting into featuring on the platform—that is, having their company and product description posted on the platform's website—these startups are seeking to reach foreign target users. But they often do not select their exact feature date. As such, the study takes advantage of the variation in the feature date to assess how foreign user growth varies by startups with a higher share of local versus foreign country initial users. Although it is difficult to fully eliminate any differences that remain between startups that have a higher share of local versus foreign initial users, the study takes several steps to account for them, including startup fixed effects, startup-level time-varying controls, coarsened exact matching (CEM), and a two-stage least squares (2SLS) approach.

The study shows that startups with a higher share of local initial users achieve higher foreign user growth after they feature on the platform. This effect is stronger in headquarters (HQ) countries where fewer languages are spoken, making it more likely that startups and local initial users are familiar with one another by sharing a common language—and therefore that local signals are clearer. The effect also magnifies in HQ countries that are more similar to the foreign target market and in product categories with less locally fragmented preferences, making local signals more representative of the foreign target market and therefore more transferable. Supplementary technology tool and experimental analyses suggest that these local initial user signals enable startups to improve their product offering for their foreign target market. These results reveal that the clarity benefits of local users exceed their transferability costs in globally standardized product categories like the digital ones that comprise the majority of the platform's startup sample, although this may not be the case in more locally fragmented categories.

Together, this study contributes to strategy and entrepreneurship research in several ways. In entrepreneurial strategy, the study advances work on how startups expand into target markets. Prior work suggests that initial users are crucial for growth in a target market (Bingham and Eisenhardt 2011, Gans et al. 2019, Moeen et al. 2020, Cao et al. 2024). This study sheds light on a clarity-transferability tradeoff that informs which initial users can effectively help startups learn about target market demand to achieve this growth. Although prior work suggests that startups should start with representative target users, this research reveals conditions—like the digital sectors where product preferences converge globally—when doing so with familiar local users can better inform target market demand and improve growth in that market.

This work also has implications for research on entrepreneurial scaling by shedding light on an initial user mechanism that can make premature scaling costly and contribute to scaling disparities. Prior work notes that scaling too early may come at the expense of developing a strong core product and business model (Lee and Kim 2024, Tidhar et al. 2025). This study suggests that this cost may at least partly result from startups cutting short the time deriving clear and transferable insights from initial users—particularly familiar ones (in a local market) if the startup is working in globally standardized product categories like the digital ones central to

this study. As a consequence, such startups based in smaller markets, which often face pressures to expand quickly to larger markets that are highly valued by investors (Kaplan and Strömberg 2004), may face greater learning frictions. Expanding too early may come at the expense of gaining clearer signals from local initial users that may, in turn, help them build better products for their larger target market. These learning frictions may hamper their growth and contribute to global scaling disparities (Conti and Guzman 2023, Wright and Saiedi 2024).

Lastly, the study offers novel ways to measure the geographic market orientation of young companies. Studies often use headquarters or subsidiary locations to define geography for mature companies. However, this measure may be less informative for young technology companies that operate almost all virtually in distributed models, especially since the COVID-19 pandemic. Instead, this study uses the geographic composition of page visits, the regional orientation of website text using a supervised machine learning algorithm, and foreign technology tool adoption in order to define the market orientation and geography of startups.

2. Theoretical Framework

The following section builds on prior work in entrepreneurial strategy to develop a theory around how the composition of initial users shapes the ability of startups to learn about target market demand. This learning enables them to update their products and thereby achieve higher target user growth. Specifically, the market origin of these initial users presents startups with a tradeoff between getting clear versus transferable signals. The theory first discusses how startups learn about demand in a target market from the signals of their initial users. It then applies this general learning framework to customer segmentation based on international geography where the target market is a foreign market and the initial user composition is either local or foreign. Online Appendix A.1 offers a formal framework to complement the verbal theory presented below.

2.1. Learning About Target Market Demand from Initial Users

Startups often target new market segments in order to scale (Bingham and Eisenhardt 2011, Moore 2014). These new markets offer a larger user base to grow startups' revenue, achieve economies of scale, and leverage network effects (Boudreau et al. 2022, Giustiziero et al. 2023). Reaching these target markets also helps startups convince investors—who tend to care deeply about market size in their funding decisions (Kaplan and Strömberg 2004, Gompers et al. 2020)—about their growth potential (Cao et al. 2024).

But startups face high uncertainty around these target users' demand (Kerr et al. 2014, Gans et al. 2019, Moeen et al. 2020). Initial users help startups resolve this uncertainty, whether it be through deliberate experiments or trial-and-error learning (Bingham and Eisenhardt 2011, Ott et al. 2017, Camuffo et al. 2020, Cao et al. 2024). These users' signals, which often come from verbal communication (e.g., calls), in-person communication (e.g., facial expressions), or written feedback (e.g., comments or surveys on a product platform), can help startups refine and update their product offering to better reach and address the needs of the target users. Indeed, young ventures iterate on their core products and strategies in response to external signals that they get from their initial users (Gans et al. 2019, Ott and Eisenhardt 2020, Pillai et al. 2020, Kirtley and O'Mahony 2023) amid high uncertainty (Kerr et al. 2014) and the absence of organizational inertia (Zhang et al. 2016, McDonald and Gao 2019). As a result, these initial users can have a profound effect on startups' trajectories (Stinchcombe 1965).

Learning about target market demand from the signals of initial users, however, can be challenging for startups. In particular, two factors can make this essential learning difficult. The first is a lack of clarity around initial users' demand. This results from communication bottlenecks between the startup and its initial users. The second is a lack of transferability: The initial users might have fundamentally different preferences than the target ones, making the initial users' signals unrepresentative of the target market.

2.1.1. Clarity. Lack of clarity is a common challenge that can prevent startups from effectively learning about target market demand from the signals of initial users. Startups may struggle to understand the initial user's value of demand. This occurs when communication gaps emerge and messaging becomes "lost in translation." One reason for this misunderstanding is that the entrepreneur may not be able to easily interpret the initial user's cues. For example, in the product feedback process, a startup might misinterpret a user's comment that the product is *interesting* to mean that the user has a positive view of the product, but in reality, the phrase in the user's local language means a distaste for it. Or a startup might mistakenly assume that a user's smile during an interview means that the user likes the product, when in that user's culture, it is common to smile whether one likes a product or not, just to be polite. Also, in surveys, a startup may mistake low early ratings as product rejection, when they actually reflect local norms of reserving high scores for truly exceptional experiences. This struggle to interpret the cues can threaten the construct validity of a signal because entrepreneurs might misinterpret proxies of

demand from initial users, whether they be facial expressions, verbal phrases, or clicking behavior.

Even if the startup can accurately interpret the signals of an initial user, that user's misinterpretation of the product can also result in unclear signals. In this case, the communication bottleneck might originate from the initial user's end. For instance, the user might simply misinterpret that a tech product is providing scheduling support, although it is actually providing goal-setting support. As such, they may send negative signals based on their own expectations and misunderstanding rather than from any shortcomings of the product itself.

These communication frictions tend to be lower when the startup and initial user are familiar with one another. They understand what each other's cues and phrases mean. Familiarity often arises when startups and initial users share a common background, particularly in terms of language and culture (Coval and Moskowitz 2001, Wang 2015, Balachandran and Hernandez 2021). This enables startups to have more knowledge about the user base and therefore be able to pick up on the various meanings of their cues and demand for the product. For example, they can pick up on whether a smile or particular phrase means that the user likes the product or is simply being polite. The common background can also make it easier for the initial user to understand the startup's description of the product and therefore give more precise signals of their demand.

Startups and initial users often share a common background when they come from the same market (Coval and Moskowitz 2001, Balachandran and Hernandez 2021). For example, a startup headquartered and operating in France will likely speak a similar language and have similar cultural values as initial users who live in France relative to those who live in India. This is particularly the case when that market is more homogeneous.[2] A shared market that is more homogeneous increases the likelihood that the startup and initial users acquired similar attributes—be they language or cultural values. For example, that startup from France also likely shares the same language—French—as its initial users in that country. In contrast, a startup based in India might have a lower likelihood of knowing the same language as the average Indian initial user because there are more than a dozen languages spoken in the country. This lower likelihood of knowing the same language then makes it harder for the startup to interpret the signals of its initial users. It also makes it harder for the initial users to properly understand the startup's explanation of the product, which can prompt them to send inaccurate signals.

This theory implies that, irrespective of the target market, startups can still struggle to learn about target market demand from an initial market when there are communication bottlenecks between them and their initial users that reduce the clarity of initial user signals. Familiarity between startups and initial users can reduce these information frictions, and it often arises when these actors share a common background. This shared background ensures that the startup understands the cues of the initial user *and* that the initial user understands the startup's explanation of the product. Startups and initial users often form this common background through living and operating in the same market, be it country or industry. The probability that this common market experience allows them to cultivate similar attributes, for example, in terms of language or cultural norms, is higher when that market is more homogeneous.

2.1.2. Transferability. Another challenge to learning is the lack of transferability of startups' initial users' signals. Irrespective of the startup's familiarity with initial users and therefore its ability to interpret their signals, it may be the case that these initial user signals do not generalize to the target users. The value of demand in this initial user base might be fundamentally different from that of the target market. This might be because the initial user base is not representative of the target market, making these initial signals biased (Cao et al. 2024). Structural differences—whether they be demographic, linguistic, or economic—between the initial and target markets might result in fundamentally different preferences around the product. For example, the initial user base may have little need for the product, whereas the target market may perceive it as highly valuable. This lack of transferability would thus threaten the external validity of startups' learning from initial users.

These structural differences between the initial and target markets magnify in product categories that exhibit locally fragmented preferences as a result of demand heterogeneity. This demand heterogeneity can emerge from differences in customer preferences around functional attributes and minimum performance thresholds, for example, due to external factors like regulations (Adner and Levinthal 2001, Shermon and Moeen 2022). It also can result from differences in people's willingness to pay for products, deriving from variation in the value that a product brings to customers. This value may be a function of a variety of factors, such as customer resources, capabilities, human capital, and the availability of substitutes (Adner and Levinthal 2001, Shermon and Moeen 2022).

Such heterogeneity ultimately results in the segmentation of product markets (Klepper and Thompson 2006, Fischer et al. 2010, Manral and Harrigan 2023). For example, product categories like aircraft or enterprise software tend to have less fragmented product preferences because of the convergence of regulations and customer preferences across the globe. As a result, there is little segmentation in these categories. On the other hand, product categories like food and

entertainment exhibit highly fragmented preferences around the world and therefore span many segments (Zou and Cavusgil 2002, Ghemawat 2007, Wright et al. 2023a). Because fragmentation depends on underlying factors such as regulations and resources, systematic differences in these factors across markets increase the likelihood that each market contains distinct product segments and therefore user preferences. As a consequence, the fragmentation of the product category magnifies any structural differences between the initial and target user markets that attenuate the transferability of initial user signals.

To illustrate this dynamic, consider a startup that seeks to learn about demand in its target U.S. market from Ukrainian initial users. Structural differences between the U.S. and Ukrainian markets—be they economic, demographic, or linguistic—might only introduce a transferability penalty to Ukrainian initial signals in product categories like food, which exhibit highly fragmented preferences across the globe. The fragmented nature of the product category would exacerbate the structural differences between the initial and target markets: Ukrainian initial users might end up representing fundamentally different product segments with distinct preferences from those of U.S. users. For example, to the extent that income at least partly influences food preferences (with lower-income consumers less likely to prefer high-end cuisines), the fact that Ukraine has a lower per capita income than does the United States[3] may contribute to there being less of a high-end food segment in Ukraine than in the United States. Thus, the two countries may feature fundamentally different food segments with distinct preferences. But in a more standardized product category like enterprise software, where preferences vary little no matter the customers' attributes, these structural differences between the markets might be muted: Ukrainian and U.S. users in this case might end up having similar preferences, reducing any transferability risk.

This general framework suggests that getting both clear *and* transferable signals in the initial user market is important to learn about demand in a target market. Clarity improves as startups and initial users are more familiar with one another, which is more likely when they share a common background. This common background often arises when the startups and initial users come from the same homogeneous market. In contrast, transferability increases when there are fewer structural differences between the initial and target markets, particularly when the product category is less locally fragmented.

2.2. Local vs. Foreign Initial Users and Foreign Target Demand

This general framework of learning from initial users about the target market can apply specifically to the case when customer segmentation is based on international geography. This is a particularly salient dimension of customer segmentation to consider because technology startups often target foreign markets in order to scale (Bingham and Eisenhardt 2011, Reuber and Fischer 2011, Wormald et al. 2021, Tippmann et al. 2023, Wright and Saiedi 2024). Indeed, in exploratory interviews, startups discussed how crucial it is for them to enter new cross-border markets to grow their user base and attract investors. For example, one U.S.-based startup mentioned: "So we are planning to go global … in that enterprise [market] … Globalization is pretty key." Also, a UK-based startup noted: "Geographical expansion is very, very important for us … A big part … is capturing the U.S. market and setting up additional international offices." Similarly, a French startup recounted: "We'll continue to expand in Europe … And also the United States … we haven't really scratched the surface." Lastly, an Australia-based startup mentioned: "From an investor perspective, if we can validate the U.S. market because it's so big, then that really helps build the credibility for the next stage of our evolution."

2.2.1. Local Initial Users Offer Clearer Signals. Startups often rely on signals from local or foreign initial users to learn about foreign target demand. However, getting an even composition of both types of initial users is often difficult because of severe resource constraints facing early-stage startups (Gans et al. 2019, Miller et al. 2023). The local versus foreign composition of initial users can shape the extent to which clarity and transferability influence startups' learning about foreign target demand. When startups and initial users come from the same country, they are more likely to share common attributes like language. These common attributes increase the familiarity between the startup and initial user. As a result, startups will have an easier time deciphering the initial user's signals, for example, whether a particular phrase, gesture, or browsing behavior indicates that the user likes or does not like the product. The initial user, too, will have an easier time understanding the startup's description of the product in order to send precise signals. In these instances, startups would face less uncertainty about the signals; they would be clearer. Indeed, one Israel-based startup in exploratory interviews explained that they started with initial users in their HQ country market because:

> It's easier, same time zone, same language, same mentality … With cultural differences, until we translate American to Israelis, it takes us time to get the actual essence of the feedback. And with Israel it is just a lot easier because that's the culture that we are used to.

This comment underscored how local initial users offered clearer feedback because of the familiarity

between the startup and those users, a clarity that arose from their shared culture and language. This would suggest that local initial users offer clearer signals because startups and local initial users are more familiar with one another, irrespective of the target market. Startups tend to be more familiar with the local market because their teams often originate from these locations (Michelacci and Silva 2007, Dahl and Sorenson 2012) and therefore likely share a similar language and cultural norms, making them encounter less uncertainty and therefore more clarity in local signals.

2.2.2. Foreign Initial Users Offer More Transferable Signals. Although more familiar to the startup, these local user signals might be less representative of the foreign target market and therefore less transferable. Systematic differences between the local and foreign markets can lead to divergent customer preferences (Ghemawat 2001, Berry et al. 2010). As a result, the information that entrepreneurs gain from customers in the local initial market may be irrelevant for understanding customers in the foreign target market (Winter et al. 2012, Khanna 2014). For example, when there are large economic differences between these markets, customers' preferences for premium products in the local initial and foreign target markets might be fundamentally different. Precisely because of these divergent customer preferences between local and foreign markets, one startup based in Israel mentioned in exploratory interviews that it started with foreign users—its target market—right away: "We set up 100 customer calls mostly in the United States because the culture and the way customers both operate and buy in the States is super different than it is in Israel. It is super different than it is in Japan, India, EU, and so on." By starting with foreign users, this venture could gain more relevant insights into its target market.

If indeed user preferences systematically vary across markets, then local initial users will not necessarily help startups understand demand in foreign target markets. To this end, one startup from Australia mentioned: "If we take too long [to go abroad] … then we can potentially box ourselves in as a product and then just find it harder and harder to … adapt the product for an international market." In contrast to initial users from the local market, initial users from the foreign target market often more accurately represent the demand of their broader market, where there are few, if any, systematic differences between the foreign initial user and target markets. The differences that do remain come from the fact that the foreign target market can be more heterogeneous than the foreign initial user base. For example, a foreign target market like the United States includes users across many different income levels—a characteristic that may fundamentally shape product preferences. Foreign initial users from this same market may only represent a subset of these income levels. As a result, the signals from the foreign initial users may give a partial view of foreign target market demand.

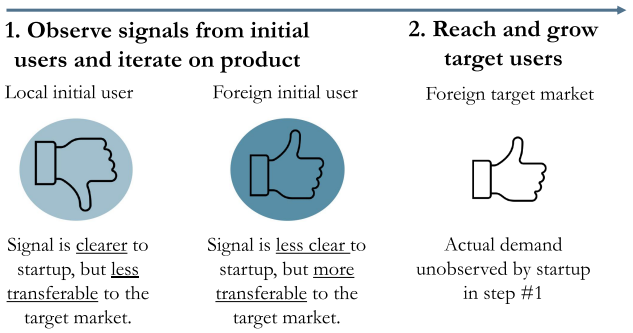
Such differences between foreign initial users and the foreign target market will generally be smaller than those between local initial users and the foreign target market. This is because foreign initial users reflect at least a subset of foreign target market demand, whereas local initial users might not even reflect a subset. For instance, foreign initial users may represent a segment of the income levels in a foreign target market, whereas local initial users might reflect fundamentally different income levels than the foreign target market. Foreign initial users' signals are therefore more representative of and ultimately transferable to the target market. By going directly to these foreign users, startups would gain a better understanding of what their target customers actually want. This transferability allows them to build products that better cater to the needs of this foreign target market.

2.2.3. Clarity-Transferability Tradeoff. Holding all else equal, the benefit of clarity that local initial users' signals might offer to enable startups to learn about foreign target demand might be offset by a higher transferability penalty because of structural differences between the local and foreign markets. These differences only matter, however, to the extent that the product category has fragmented preferences across the globe. Even amid large structural differences—be they linguistic or economic—between the local and foreign markets, local initial users could still offer transferable signals if the product category has standardized preferences across geographies.

Together, this theory suggests that learning from local versus foreign initial users presents startups with a tradeoff between the transferability and clarity of signals. Local initial users might offer clearer signals because there is greater familiarity between them and the startup, but they might give less transferable signals if the local user base is not representative of the foreign target one. Conversely, foreign initial users might offer more transferable, but potentially less clear, signals. Crucially, this theory relies on three important assumptions. Firms are in early stages, such that initial users—be they local or foreign—inform their product development. These startups also ultimately seek to conduct a broad launch of their product to their target users. Lastly, these target users tend to be from a market that is not the startup's home market (i.e., a foreign market).

Figure 1 illustrates this theoretical tradeoff under these assumptions for a startup targeting a foreign market with an unobserved positive demand (thumbs up) for its product. It shows that startups begin by (1) observing signals from their initial users and use them to iterate on their product and (2) then reach and grow

Figure 1. Tradeoff Between the Clarity and Transferability of Initial User Signals



Note. The figure illustrates a key tradeoff between clarity and transferability that startups confront when they (1) observe signals from local and foreign initial users and iterate on their product and then (2) reach and grow target foreign users: Local initial users may offer clearer, but less transferable signals, and foreign ones the reverse.

their target users. In the first step, local initial users offer clearer signals because they are more familiar to startups (with a lighter blue layer of "noise"), but they might be less transferable because they are less representative of the foreign target market. Thus, the underlying value of their demand may be different from that of the foreign target market (thumbs down rather than thumbs up). The opposite is true for foreign initial users: Their signals might be more transferable (thumbs up), but less clear (with a darker blue layer of "noise"). Whether the clarity effects outweigh the transferability ones ultimately determines which initial user signals enable startups to better update their products and grow their target users in the second step.

2.3. Summary

Overall, this theory illustrates that the local versus foreign composition of initial users presents startups with an important tradeoff: clear versus transferable signals of foreign target demand. Local initial users might offer clearer signals of demand because of the greater familiarity between them and the entrepreneur. But these local initial signals can carry a transferability cost because the local market might not be representative of the foreign target market. The relative value of clarity versus transferability then shapes whether local initial users enable startups to build products that achieve higher foreign user growth. For example, even the clearest signals in the local market are useless for gaining foreign user growth if they are so idiosyncratic that they do not transfer elsewhere. Similarly, even the most transferable signals in the foreign market are ineffective if they are unclear—the startup cannot understand them. This leads to a conjecture: *Startups with a higher share of local initial users will achieve higher foreign user growth when local signals are clearer and more transferable.* The following section proceeds to test this theory.

3. Empirical Setting: Early-Stage Product Platform

To test this theory about how the composition of initial users shapes subsequent target user growth, the study collects data from BetaList. Founded in 2010, BetaList is one of the top early-stage product platforms in the market and is based in the Netherlands. Technology startups post a description of an early-stage product and a link to the company's website on this platform in order to grow their target user base. This allows the study to track how startups' local (versus foreign) initial user composition prior to featuring on the platform shapes subsequent target user visit growth to the startups' websites after feature. Most of the business transactions of such technology ventures occur on their company website, so website visits serve as a meaningful indicator of user growth.

Startups begin by submitting an application to post their product on the platform. This application solicits information about (1) the product, including the name of the product, a short "pitch" and a longer, two-paragraph description of the product, the product categories (e.g., SaaS, social media, e-commerce), and product images; (2) the company, including the website, the country location; and (3) the team, including names and social media account links. The platform requires each startup to be in the technology sector (whether software and/or hardware), have a sufficiently appealing website, and include a mailing list on their website to allow visitors to sign up or subscribe. A (paraphrased) example of a product description includes the following:

> [Startup name] acts as a search engine on social media for finding prospective clients. The process is straightforward: you input descriptions of your product or service and your competitors' brand names into the system. Using these terms along with predictive analytics, [the startup] scours social media to identify individuals in search of a product similar to what you offer.

Crucially, the platform embodies the three assumptions that underpin the theory of this paper. First, startups are in a similar early stage such that initial users inform product development. This is primarily because the platform prohibits startups that have had a broad launch in the past. In exploratory interviews, startup founders mentioned posting on several such platforms (e.g., also Product Hunt), but doing so first on BetaList. This design feature of the platform ensures that at the time of feature, the startups have had some initial users, but the platform is their first "big" launch. This means that any users prior to feature are likely part of a small pool of initial users—sourced deliberately or organically—who offer early signals from which startups can learn to improve their product for the broad launch on the platform.

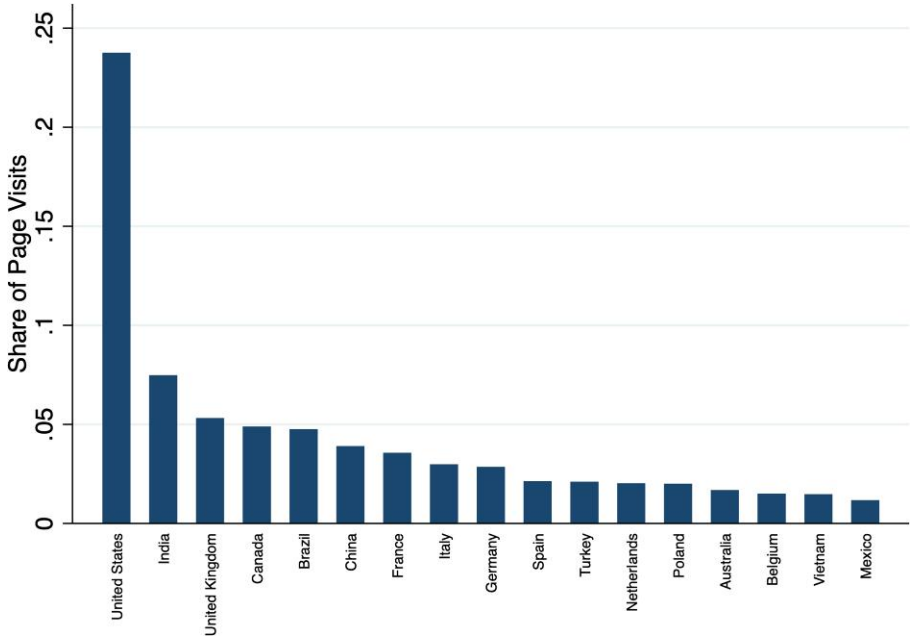
Second, the startups that self-select into featuring on the platform intend to conduct a broad launch to their target market. Indeed, the platform enables early-stage companies to reach a mass set of users. In particular, the large share of U.S. users on the platform (Figure 2) suggests that the U.S. user base is an important target market for the technology startups applying to feature on the platform. This user base therefore serves as a foreign target market for non–U.S.-based startups, which is consistent with prior research and exploratory interviews that suggest that the U.S. market is crucial for technology startups from other markets to attract investment and grow (Conti and Guzman 2023).

Lastly, the target market for startups is typically not the same as their home market. The platform guidelines specifically attract startups in the technology sector, which often need to expand into new markets in order to scale. Moreover, the majority of startups in the sample are based outside of the United States, given the European roots of the platform. Roughly 60% of the startups are based outside of North America, with the majority of this share located in Europe and Central Asia. The rest are distributed across South Asia, East Asia and the Pacific, the Middle East and North Africa, Latin America and the Caribbean, and sub-Saharan Africa. Thus, for most startups in the sample, the U.S. target market is distinct from their home market. The rich international heterogeneity of startups also provides variation to assess foreign versus local initial users and subsequent foreign user growth.

Beyond satisfying the theory's three core assumptions, the platform also helps address two endogeneity concerns in assessing how initial user composition shapes foreign user growth. The first concerns launch timing. Although broad launch dates are typically endogenous and may influence foreign user growth independently of learning, the platform generally targets early-stage ventures and controls the exact feature date. Startups can submit for free—often waiting weeks or months—or pay to expedite or schedule a specific date. Blog evidence suggests many choose the free option, leaving timing largely outside their control. As a result, the launch occurs at an early stage of the firm's life cycle, and its precise date is plausibly exogenous to startups (Athey and Stern 2013), which offers useful variation to isolate the effect of initial user composition on foreign user growth.

A second endogeneity concern is that underlying startup differences could drive both their share of local initial users and subsequent foreign user growth. For example, startups that target foreign markets less than others may both have a higher share of local initial users and achieve less subsequent foreign user growth than others. Alternatively, startups with stronger networks or greater marketing resources might reach a higher share of foreign users initially and later grow that foreign user base more than others regardless of learning from initial users. The platform context mitigates these concerns in several ways. This context attracts early-stage technology startups that often need to expand into new markets to scale. These startups specifically self-select into a platform where U.S. users make up the largest single share, suggesting that they similarly view these users as their target market. The

Figure 2. (Color online) Platform's Composition of Visitors from January 2018 to March 2021



Note. Includes countries with at least a 0.01 share.

platform also reduces frictions—such as marketing or access costs—that typically shape visibility (Adner et al. 2019). Foreign users can easily view all featured startups and click through to their websites, which forms the basis of the foreign user growth measure. Thus, the platform reduces systematic differences between startups while preserving heterogeneity in their local versus foreign initial user composition.

4. Data

The platform setting enables the creation of a startup-month panel data set. The study narrows the original sample of about 11,000 startups that launched on the platform from January 1, 2010, to September 6, 2020, by roughly a fifth to those companies that have a credible website link. Doing so ensures that it is possible to track the initial users and subsequent foreign user growth of the startups. The study further narrows this sample to startups that featured for the first time on the platform in order to remove any confounding effects from their prior feature. It also restricts the sample to startups featured from April 2018 to September 2020 to ensure that page visit data at least three months before and six months after are available for all of the startups. This reduces the sample to about half of the original. Of the remaining sample, the study removes firms that had no geographically broken-down page visit data available in the sample period of three months before and six months after. The study does this to ensure that firms were active in this period and to identify the local versus foreign initial user composition of startups prior to feature. This removes roughly a fifth of the original startup observations.

This narrowing brings the final sample to 1,106 startups featured from April 2018 to September 2020. The final sample includes only startups for which (1) website page visit data are available at least three months before and six months after feature on the platform, (2) those for whom the geographic composition of their website traffic are available, and (3) those that featured for the first time on the platform. The final data set is a startup-month panel of these 1,106 firms from January 2018 to March 2021 (three months before and six months after feature) based in 71 countries with a total of 11,060 observations. The final sample is also generally similar in its regional distribution to the original sample.[4] Of this sample, the majority (686 startups) are based outside of the United States, accounting for 6,860 observations.

Figure 3 illustrates how initial users offer signals to startups that prompt product updates, which then generate subsequent foreign user growth. The platform data allow for the assessment of how the composition of initial users (three months prior to featuring on the platform) shapes the ultimate outcome of the product iteration: foreign user growth (six months after featuring). Although the platform data cannot directly reveal the initial user signals and product iteration process, supplementary technology tool and experimental analyses help shed light on both.

The next section discusses the specific variables from the platform data.

4.1. Dependent Variables

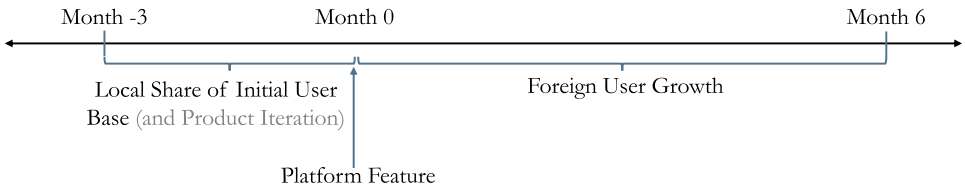
Log Foreign Country Visits: The main dependent variable used in the analysis is a proxy of startups' foreign user growth after launching on the platform.[5] This variable indicates logged page visits from users outside the focal company's HQ country in a given month to the startup's home website linked on the platform. It would capture, for example, visits from U.S.-based users to a Germany-based startup's website. These data come from the SimilarWeb database. Website page visits are a common performance indicator for early-stage (prerevenue) startups because they reveal user engagement and are predictive of venture funding (Hallen et al. 2020, Koning et al. 2022, Wright et al. 2023a, Cao et al. 2024).

Log U.S. Visits: To supplement the log foreign country visit measure, the analysis also uses logged visits from U.S.-based users for non–U.S.-based startups in a given month. This builds on prior work and exploratory interviews that suggest that the United States is a crucial target market for technology startups around the world (Conti and Guzman 2023). The United States is also the most common country of users that visit the platform (Figure 2), making it a likely target of startups that feature on the platform.

4.2. Explanatory Variables

Post: This variable indicates whether a given month-year observation is on or after the feature date of the startup on the platform. It takes a value of zero if, in a

Figure 3. (Color online) Timeline of Startup Events



given month, a startup has not yet been featured on the platform and one otherwise.

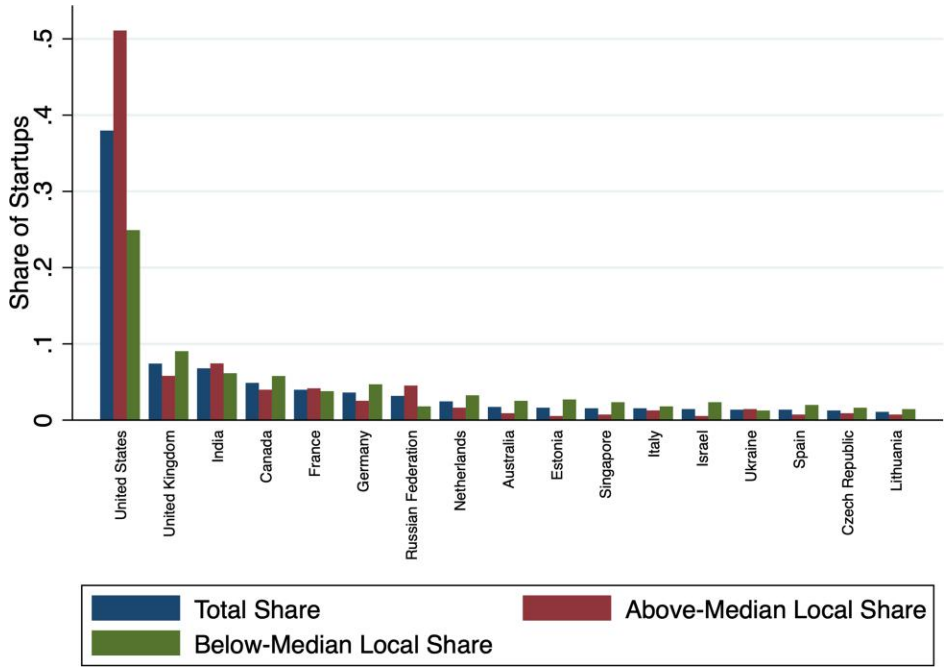
Pre-Local Share: This variable indicates the average monthly share of visits from users in the startups' HQ country during the three months prior to feature,[6] which reflects the local versus foreign composition of startups' initial user bases.[7] Figure 4 shows the distribution of startups by HQ country with above-median (red bar) and below-median (green bar) local initial user shares. The United States is the top HQ country for both. The United Kingdom and India are also in the top. The remaining primary HQ countries among those with higher shares of local initial users tend to be larger than those with lower shares of local initial users. To account for this difference in market size, subsequent analyses use firm fixed effects and coarsened exact matching on the HQ country, monthly initial visits, and the initial number of markets. The analyses also exclude startups headquartered in the United States, among the largest HQ markets. Figure A.4.1 in the Online Appendix shows the breakdown of the share of visits from local country users prior to featuring. The distribution is shifted to the left, with some startups showing no visits from their own country. The results are robust to excluding these startups.

Post × Pre-Local Share: The interaction term between featuring (post) and prior local HQ country share reveals the relationship between the prefeature local composition of startups' user base and postfeature foreign user growth. The coefficient on this variable serves as the main way to answer the research question about how the local versus foreign composition of a startup's initial user base shapes its subsequent foreign user growth.

Fixed Effects: The specifications include several fixed effects. Startup fixed effects account for quality differences between companies in the sample, which are related to both observable factors like country location or product category, as well as unobservable factors, such as the motivation of the founders. Month-year fixed effects account for seasonality in the number of visitors on the platform on a given feature date.

Number of HQ Country Languages and Ethnic Diversity: The theoretical framework suggests that when the HQ country is more homogeneous, for example, with fewer spoken languages, startups are more likely to share a common background—be it language or cultural norms—with initial users who are also from that country. This greater homogeneity makes familiarity between the startup and local initial users more likely, which increases the clarity of signals that the startup gets from these users. To measure this HQ homogeneity, the study uses the number of spoken languages in a startup's HQ country, based on data from www.cia.gov, aggregated by https://languagerc.net/languages-by-countries/. For example, France has one language associated with it: French. India, in contrast, has 14 languages associated with it, including Hindi, Bengali, Telugu, Marathi, Tamil, and Urdu, among others. France is therefore more linguistically homogeneous than India. For robustness, a country-level measure of ethnic diversity from Drazanova (2020) serves as another indicator of HQ homogeneity. Specifically, lower ethnic diversity in the HQ country could increase

Figure 4. Startup Sample by HQ Country



Note. Includes the HQ countries of 1,106 startups in the analysis sample with at least a 0.01 share.

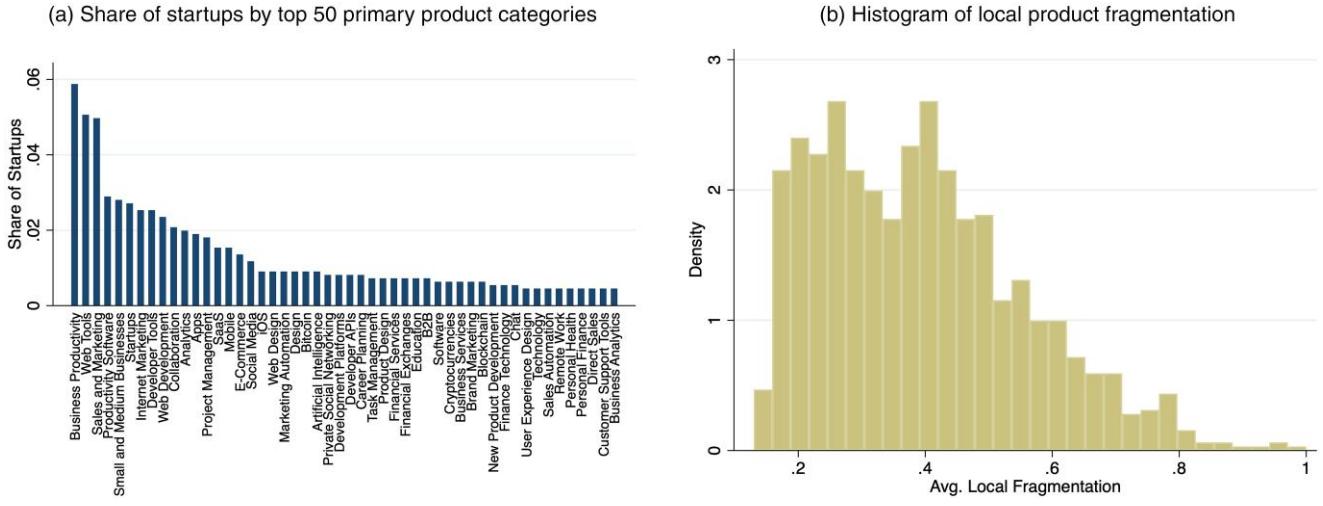
familiarity between the startup and local initial users and therefore the clarity of the initial users' signals. Both of these measures are available on an HQ country level and do not vary over the sample period.

Distance Between Local and Foreign Markets: The theoretical framework also suggests that when there are smaller differences between the local initial and foreign target markets, local signals will be more transferable. Pairwise country distances, including linguistic, demographic, and economic, between the HQ (local) country and U.S. (foreign) target markets serve as measures of these differences. Because the United States is the top origin of users of the platform (Figure 2) and is a common target market for the IT industry, it serves as a plausible common foreign target market for non–U.S.-based startups on this platform. To measure linguistic distance, the study uses data on spoken languages from the same source as was used to create the number of HQ country languages, which enables coding whether the primary language listed for a startup's HQ country is the same as in the U.S. target market (English).[8] The pairwise country distance variables for all other measures (e.g., economic and demographic) come from Berry et al. (2010): These data are available on a yearly basis up to 2021 and are matched to the year of the startup's feature date.

Local Product Fragmentation: The theoretical framework suggests that the above local-foreign market structural differences would only matter to the extent that the product category exhibits locally fragmented preferences, like in the food category, rather than in, say, enterprise software. Figure 5(a) shows that a large share of the startups in the sample are in digital product categories like business productivity and web tools, which exhibit fairly standardized preferences across countries. This distribution suggests that any cross-country differences will be muted in this sample. In the spirit of the Herfindahl-style measures of market fragmentation used in prior work (Ziedonis 2004, Sorenson et al. 2006, Shermon and Moeen 2022, Landsman et al. 2024), the study creates a measure of local product fragmentation that captures the country representation of startups in each product category as a share of the total startups in that category in the final sample.[9] This measure indirectly captures how geographically fragmented the preferences are in a particular category. The higher this ratio, the more geographic fragmentation there is. The highest ratio is one, characterizing product categories like "food and beverage" and "local businesses." The lowest ratio is 0.1, characterizing the "SaaS" product category.[10] Because startups can add up to five product category tags, the final measure reflects the average of the local fragmentation rates across these product tags for each startup. Figure 5(b) shows that the startups are skewed to the left in this local fragmentation measure, meaning that most of them are in globally standardized product categories, as we would expect in the IT sector. The study primarily uses a binary version of this measure that indicates whether the local fragmentation rate is above the median value in the final sample (*local product*) or below it (*global product*).

This measure is robust to an alternative survey-based proxy of local fragmentation. Without disclosing the research question, predictions, or results from the study, this survey-based measure asks 200 participants on Prolific to select from a full list of product categories that are available on the platform (396 in total) which ones they expect would exhibit large differences in customer preferences across countries. The final measure captures the percent of evaluators who choose each product category. The results are similar using this

Figure 5. (Color online) Share of Startups by Top 50 Primary Product Categories



Note. Figure 5(a) shows the first product category of startups in the top 50 categories.

measure for the first (generally primary) product category or the average of all product categories for each startup.[11]

Time-Varying Startup Factors: Several robustness checks control for startup-level factors that vary across time and might influence foreign user growth independently of the platform exposure. These factors include venture funding, measured by the logged U.S. dollars in funding that startups receive in a given month using PitchBook data. They also include startups' installation of tools that target foreign users, including foreign language and hosting tools from BuiltWith data, which has been increasingly used by strategy and entrepreneurship scholars to understand the innovation activity of software startups (Dushnitsky and Stroube 2021; Koning et al. 2022; Stroube and Dushnitsky 2023, 2025; Impink and Miric 2024). Specifically, the study controls for the logged number of foreign language technology tools installed in a given month on a startup's website.[12] A company headquartered in France that adopted an English technology tool in a given month would be coded as adopting one foreign language tool. If this company adopted an English and a Spanish tool, then this would count as two foreign language tools. The study also controls for the logged number of foreign hosting tools that startups adopt on their websites.[13] A company headquartered in the United States that adopted a Google Cloud Belgium tool in a given month would be marked as adopting one foreign hosting tool. If this company adopted a Google Cloud Belgium and also a Google Cloud Singapore tool, then it would be coded as having two foreign hosting tools in that month.

Summary Statistics: Tables 1 and 2 show summary statistics comparing startups with below-median and above-median local (HQ) country shares of website visits prefeature and postfeature on the platform.[14] As expected, prior to featuring, startups with a higher share of local visits see lower foreign country visit growth. They also install fewer foreign hosting tools, have website text that is less oriented toward foreign regions,[15] and are more likely to be in a local and business-to-consumer (B2C) product category.[16] They are additionally more likely to be based in larger gross domestic product (GDP) markets like the United States, consistent with prior work (Fan and Phan 2007) and Figure 4. That being said, the fact that there is not a meaningful difference in the number of foreign language tools installed suggests that both types of startups similarly target foreign user growth. This is consistent with the study's motivation: Expanding into new (foreign) markets is a common goal for startups.

4.3. Econometric Framework

To understand how foreign user growth on the platform varies by startups' local versus foreign composition of initial users, the study uses the following specification:

$$Y_{it} = \beta_1 Post_{it} + \beta_2 Post_{it} \times Pre\text{-}LocalShare_i + \gamma_i + \delta_t + \epsilon_{it}, \quad (1)$$

where i indicates startup and t is time (in months); Y_{it} indicates the logged number of visits to a startup's website from users who are foreign to the startup's HQ country; $Post_{it}$ indicates whether the month is after a startup's feature date on the platform; $Pre\text{-}LocalShare_i$ indicates the share of website visits that startups got from local users (i.e., the same country as the HQ) prior to featuring on the platform; $Post_{it} \times Pre\text{-}LocalShare_i$ indicates how the effect of featuring on the platform varies for startups with a higher versus lower share of local initial users; γ_i represents firm fixed effects to account for quality differences between startups, enabling a

Table 1. Summary Statistics of Startups Before Feature

	Low local observation	Low local mean	High local observation	High local mean	Difference
Foreign country share	554	0.94	552	0.49	0.44***
Log local country visits	554	4.43	552	6.92	−2.49***
Log foreign country visits	554	8.07	552	6.65	1.42***
Log U.S. visits	554	5.99	552	5.63	0.35*
Foreign language tools	554	0.17	552	0.17	−0.00
Log foreign language tools	554	0.08	552	0.08	−0.00
Foreign hosting tools	554	0.16	552	0.10	0.06**
Log foreign hosting tools	554	0.11	552	0.07	0.04**
Foreign-oriented text	184	0.64	177	0.51	0.13*
HQ country language count	554	6.56	552	6.98	−0.42
Local product fragmentation	554	0.39	552	0.40	−0.01
B2C product category	554	0.47	552	0.53	−0.07*
Log total visits	554	8.15	552	7.70	0.44***
U.S. HQ	554	0.25	552	0.51	−0.26***
Log HQ GDP	550	28.14	551	29.27	−1.14***
N	1,106				

Note. Shows startups with below- versus above-median local initial user share.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 2. Summary Statistics of Startups After Feature

	Low local observation	Low local mean	High local observation	High local mean	Difference
Foreign country share	554	0.91	552	0.64	0.27***
Log local country visits	554	5.19	552	6.82	−1.64***
Log foreign country visits	554	8.14	552	7.31	0.83***
Log U.S. visits	554	6.44	552	6.29	0.15
Foreign language tools	554	0.30	552	0.20	0.09
Log foreign language tools	554	0.11	552	0.10	0.02
Foreign hosting tools	554	0.24	552	0.18	0.06^{+}
Log foreign hosting tools	554	0.15	552	0.11	0.04*
Foreign-oriented text	242	0.68	250	0.53	0.15***
HQ country language count	554	6.56	552	6.98	−0.42
Local product fragmentation	554	0.39	552	0.40	−0.01
B2C product category	554	0.47	552	0.53	−0.07*
Log total visits	554	8.26	552	7.93	0.33*
U.S. HQ	554	0.25	552	0.51	−0.26***
Log HQ GDP	550	28.14	551	29.28	−1.14***
N	1,106				

Note. Shows startups with below- versus above-median local initial user share.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

within-firm analysis;[17] and δ_t indicates month-year fixed effects to account for time variation that may influence the number of users visiting websites over time.

The coefficient of interest is β_2, which indicates how featuring on the platform impacts startups with a higher share of local versus foreign initial users in terms of their foreign user growth. If this coefficient is positive, then this suggests that startups with a higher share of local initial users achieve higher foreign user growth after featuring on the platform. To understand the mechanisms behind this effect, the study assesses this specification across HQ country markets with less (versus more) linguistic and ethnic heterogeneity where we would expect that startups and local initial users are more familiar with one another, thus resulting in clearer signals. Also, it assesses this effect across smaller (versus larger) distances between the local HQ (initial) and foreign (target) markets and in less (versus more) locally fragmented product categories that would make the local market more representative of the foreign target market, resulting in more transferable signals.

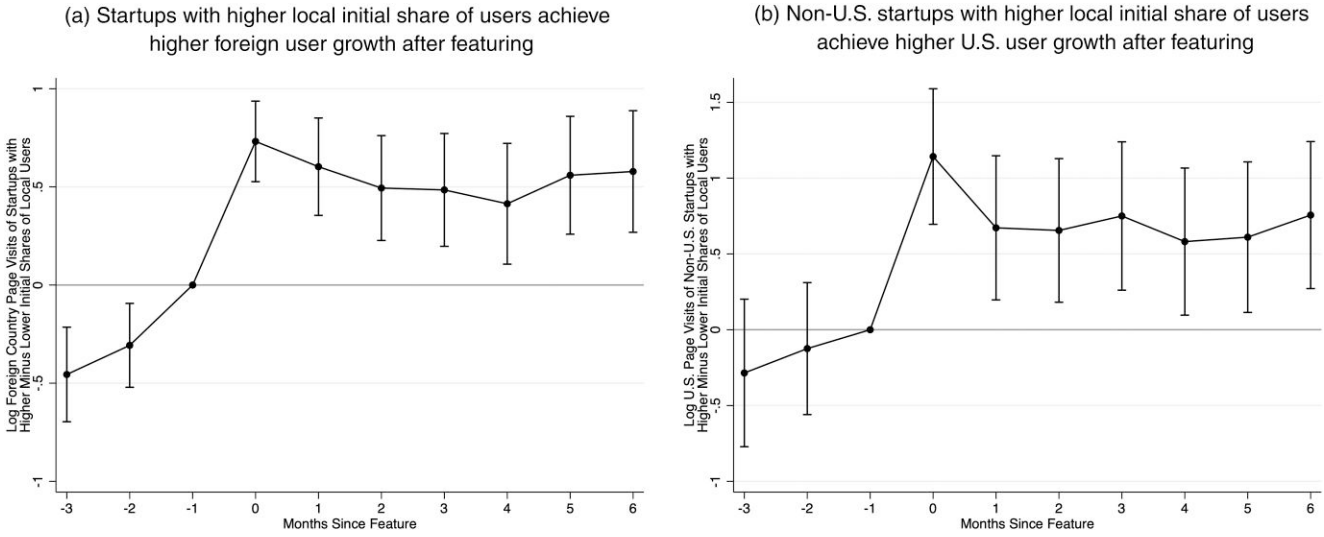
5. Results

5.1. Local Initial User Composition and Foreign User Growth

Do startups with a higher share of local initial users achieve higher foreign user growth after featuring on the platform? Tables 1 and 2 show that, after featuring on the platform, the gap between logged foreign—and particularly U.S.—visits narrows between startups with an above- versus below-median initial share of local users by more than 40%. This suggests that startups with a higher share of local initial users achieve higher foreign user growth after the platform launch. To further investigate this trend, Figure 6 plots the coefficients from applying Equation (1) with logged page visits from foreign country users as the dependent variable. Consistent with the summary tables, it shows that startups with an above-median share of local country initial users see higher growth in website visits from foreign countries after featuring on the platform than do startups with a below-median share of local initial users, despite seeing less growth in these users prior to feature. The pattern is similar when limiting the sample to startups based outside of the United States and focusing on growth in visitors from the United States, the single largest origin of users on the platform.

Table 3 applies Equation (1) with continuous versions of the initial share of local country users to assess this heterogeneity more rigorously. It shows that, at baseline (*Post*), startups see a 50% increase in foreign country visits after featuring on the platform (column 1). This growth is driven by startups with an above-median share of local initial users, which see 92% growth in these visits (column 2). In fact, those with a below-median share of local initial users do not see any meaningful growth in foreign users: The coefficient on *Post* is more than 10 times smaller in magnitude than the one for startups with an above-median share of local initial users and is not statistically significant at the 5%–10% level (column 3). Also, the difference in magnitudes between the two groups (*Post* × *Pre-Local Share*) is statistically meaningful at the 0.1% significance level, far exceeding the magnitude of the coefficient on featuring on the platform (*Post*) (column 4). In other words, startups with a higher share of local initial users see higher growth in website visits from foreign countries than do others after featuring on the platform. The results are similar when assessing only non–U.S.-based startups and their growth in U.S.-based user visits (columns 5–8).

Figure 6. Effect of Local Initial User Share on Foreign and U.S. User Growth



Could it be that startups with a higher local initial user share fundamentally differ in their quality or objectives relative to those with a lower local initial share? For example, startups with a higher local initial user share could simply have more website traffic overall prior to featuring, perhaps because they form a different type of knowledge or industry experience in their home market that enables them to later increase foreign user growth (Mitchell et al. 1994, Shane 2000, Cuervo-Cazurra 2011). Or perhaps they were simply more interested in foreign users as a target market. To help account for these possibilities, Table 4 shows that the baseline results hold even when controlling for funding and growth in foreign language and hosting tools—time-varying factors that correlate with startups' quality and market objectives that might independently impact foreign user growth. The results also hold when matching startups to a "twin" with a similar number of prefeature average page visits and country markets from the same HQ country using CEM. They continue to hold when controlling for these prefeature average page visits and other feature date characteristics in Table A.7.1 in the Online Appendix.

Together, startups with a higher share of local initial users see higher foreign user growth after featuring on the platform. These trends are consistent with the idea that startups with a higher share of local initial users build products that are better able to achieve foreign user growth after featuring on the platform.

5.2. Clarity and Transferability Mechanisms

What could be driving the higher foreign user growth that startups with a higher share of local initial users ultimately see? The theoretical framework suggests that initial users who offer clearer and more transferable signals will help startups build better products for foreign target users. These products then achieve higher foreign user growth. This would mean that startups with local initial users who offer clearer and more transferable signals should drive the baseline results.

Table 3. Startups with a Higher Local Initial User Share Achieve Higher Foreign User Growth After Featuring

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log foreign country visits (all HQ)				Log U.S. visits (non-U.S. HQ)			
Post	0.500***	0.923***	0.079	−0.076	1.085***	1.700***	0.682***	0.720***
	(0.071)	(0.108)	(0.091)	(0.084)	(0.108)	(0.180)	(0.132)	(0.122)
Post × *Pre-Local Share*				2.010***				1.531***
				(0.213)				(0.266)
_cons	6.672***	5.803***	7.537***	6.674***	3.960***	2.468***	4.931***	3.962***
	(0.050)	(0.076)	(0.064)	(0.048)	(0.075)	(0.126)	(0.092)	(0.074)
N	11,060	5,520	5,540	11,060	6,860	2,699	4,160	6,860
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	High local	Low local	Full	Non-U.S.	High local	Low local	Non-U.S.

Notes. Shows startups featured on platform from April 2018 to September 2020. The sample size drops because of singleton observations. Robust standard errors are clustered at the company level.

$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Table 4. Startups with a Higher Local Initial User Share Achieve Higher Foreign User Growth After Featuring with Funding and Foreign Technology Tool Controls and CEM

	(1)	(2)	(3)	(4)	(5)	(6)
	Log foreign country visits				Log U.S. visits	
Post	−0.077	−0.078	−0.079	−0.138	0.719***	0.816***
	(0.084)	(0.085)	(0.085)	(0.112)	(0.122)	(0.166)
Post × *Pre-Local Share*	2.009***	2.013***	2.012***	2.049***	1.533***	1.349***
	(0.213)	(0.213)	(0.213)	(0.272)	(0.267)	(0.323)
Log Deal Size	0.103		0.103	0.055	−0.051	−0.063
	(0.065)		(0.065)	(0.083)	(0.246)	(0.263)
Log Foreign Language		0.104	0.103	0.080	0.280	0.493
		(0.227)	(0.227)	(0.343)	(0.277)	(0.495)
Log Foreign Hosting		0.042	0.039	−0.141	−0.328	−0.528
		(0.305)	(0.304)	(0.362)	(0.327)	(0.419)
_cons	6.673***	6.661***	6.661***	6.296***	3.980***	3.146***
	(0.048)	(0.059)	(0.059)	(0.072)	(0.094)	(0.115)
N	11,060	11,060	11,060	7,340	6,860	4,570
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Month × Year	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Matched	Non-U.S.	Non-U.S. matched

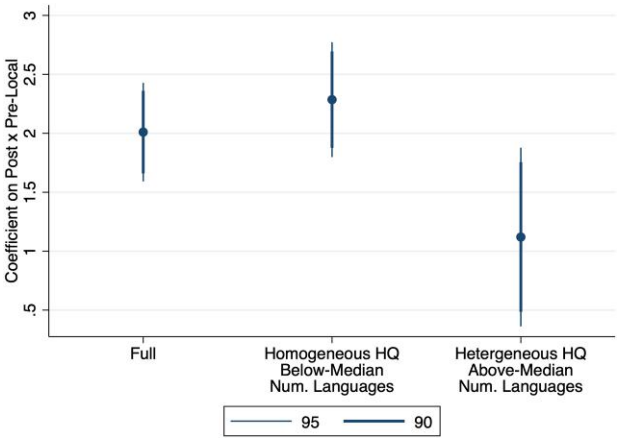
Notes. Shows startups featured on platform from April 2018 to September 2020. Robust standard errors are clustered at the company level.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The study begins by testing the clarity mechanism. Greater familiarity between startups and initial users would increase the clarity of these users' signals. This enables startups to better learn from these early signals. The homogeneity of the local market—for example, where there are fewer languages spoken—would magnify the familiarity because it increases the probability that startups and initial users from the same market share common attributes like language and cultural norms. The analysis therefore proceeds to assess how the difference in foreign user growth by startups with a higher versus lower share of local initial users varies by the number of HQ spoken languages. In effect, this means assessing variance in Equation (1) by the number of main languages spoken in the HQ country.

Figure 7 and Table 5 assess this variance by the HQ spoken language count. They reveal that the local initial user advantage (*Post* × *Pre-Local Share*) is lower in magnitude where there is an above-median number of languages spoken (Table 5, column 1). The effect more than doubles for startups based in HQ markets where there is a below-median number of languages spoken (column 2). When using a continuous measure of languages spoken, the difference (*Post* × *Pre-Local Share* × *HQ Language Count*) is statistically meaningful at the 0.1% level, reflecting nearly 20% of the baseline (*Post*) coefficient (column 3). Similar patterns emerge for ethnic diversity, although they are less statistically precise (column 4). The results are also similar when assessing U.S. user growth for startups based outside of the United States (columns 5–8). These patterns suggest that startups with a higher share of local initial users achieve higher foreign user growth when based in more linguistically and ethnically homogeneous local contexts. Greater homogeneity serves to increase the familiarity between startups and local initial users and therefore the clarity of the signals that startups get from these users.

The study next turns to the transferability mechanism. The theoretical framework suggests that initial users who offer more transferable signals will better inform startups' products, which can fuel higher foreign user growth. Signals from local initial users are more transferable when the local market is more representative of the foreign target market. This is

Figure 7. (Color online) Effect Magnifies in HQ Contexts with Fewer Languages Spoken, Increasing the Familiarity Between Startups and Local Initial Users and, Hence, the Clarity of Local Signals



Note. Shows coefficient from regressing log foreign visits after featuring on the local initial user share.

Table 5. Effect Magnifies in HQ Countries with Fewer Languages Spoken, Increasing Familiarity Between Local Initial Users and the Startup and Therefore the Clarity of Local Signals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log foreign country visits				Log U.S. visits (non-U.S. HQ)			
Post	0.379*	−0.200*	−0.084	−0.207	0.989***	0.625***	0.724***	0.452^{+}
	(0.164)	(0.097)	(0.085)	(0.240)	(0.246)	(0.139)	(0.122)	(0.246)
Post × *Pre-Local Share*	1.120**	2.285***	2.116***	2.793***	1.091*	1.730***	1.536***	2.715***
	(0.384)	(0.248)	(0.217)	(0.728)	(0.460)	(0.328)	(0.271)	(0.719)
Post × *HQ Language Count*			0.001				−0.001	
			(0.001)				(0.001)	
Post × *Pre-Local Share* × *HQ Language Count*			−0.016***				−0.001	
			(0.002)				(0.002)	
Post × *HQ Ethnic Diversity*				0.233				0.817
				(0.498)				(0.546)
Post × *Pre-Local Share* × *HQ Ethnic Diversity*				−1.504				−3.083^{+}
				(1.498)				(1.828)
_cons	6.308***	6.754***	6.674***	6.715***	3.558***	4.122***	3.962***	4.015***
	(0.105)	(0.054)	(0.048)	(0.051)	(0.149)	(0.085)	(0.074)	(0.082)
N	2,029	9,030	11,060	9,650	2,029	4,830	6,860	5,450
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	High lang	Low lang	Full	Full	High lang	Low lang	Full	Full

Notes. Shows startups featured on platform from April 2018 to September 2020. The sample size drops because of singleton observations. Robust standard errors are clustered at the company level.

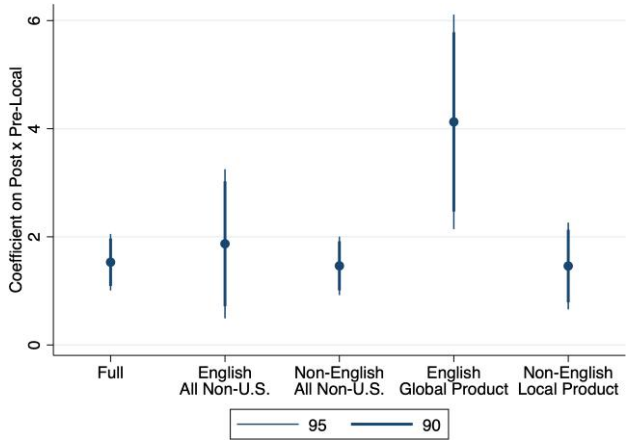
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

particularly the case when there are fewer differences between the local initial market and foreign target market. To test this, Figure 8 and Table 6 now show how Equation (1) varies by the distance between the local (HQ) market and the U.S. target market for startups based outside of the United States and in locally fragmented (versus global) product categories.

The theory suggests that the effect would magnify in less locally fragmented categories (i.e., global categories) that mute any local-foreign market structural differences. Figure 8 and Table 6 both show consistent patterns with this prediction. The local initial user effect (*Post* × *Pre-Local Share*) more than doubles when there are smaller linguistic, demographic, and economic distances between the local (HQ) and foreign (U.S.) markets in less locally fragmented product categories like SaaS (columns 3, 5, and 7). The results are similar using other proxies of market distance (Table A.5.1 of the Online Appendix). Of course, most of the startups are in digital product categories like business productivity and web tools that generally exhibit globally standardized preferences, as shown in Figure 5. Thus, the average local fragmentation is fairly low at about 0.4 (out of 1) (Figure 5). The low product fragmentation mutes any structural differences between local and foreign markets and therefore reduces the transferability penalty that local initial users might otherwise introduce. This may explain why the baseline results favor startups with a higher share of local initial users: These users offer clearer signals with little transferability cost.

Together, these results reveal that startups with a higher share of local initial users achieve higher foreign user growth. This effect magnifies in local (HQ) markets that are more linguistically and ethnically homogeneous, increasing the familiarity between the startup and local initial users and therefore the clarity of these users' signals. It also magnifies when there are smaller differences between the local initial and foreign target markets and in less locally fragmented product categories, making those local signals more representative and therefore more transferable to the foreign target market. These patterns are consistent with the theoretical framework's

Figure 8. (Color online) Effect Magnifies in HQ Contexts with Smaller Linguistic Differences from the Foreign Target Market and in Less Locally Fragmented Product Categories, Increasing the Transferability of Local Signals



Note. Shows coefficient from regressing log U.S. visits after featuring on the local initial user share for non-U.S.-based startups.

Table 6. Effect Magnifies When There Are Smaller Distances Between the HQ (Initial) and U.S. (Target) Markets, Particularly in Less Locally Fragmented Products, Making Local Signals More Representative and Therefore More Transferable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log U.S. visits						
Post	0.720***	0.700***	0.147	0.611**	−0.122	0.760***	0.483
	(0.122)	(0.189)	(0.389)	(0.184)	(0.446)	(0.191)	(0.443)
Post × *Pre-Local Share*	1.531***	1.460***	4.126***	1.455***	3.874***	1.335***	3.723**
	(0.266)	(0.407)	(0.994)	(0.393)	(1.038)	(0.392)	(1.080)
N	6,860	2,559	679	2,720	550	2,660	508
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month × Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HQ sample	All non-U.S.	Non-English	English	High demographic distance	Low demographic distance	High economic distance	Low economic distance
Product sample	All	Local product	Global product	Local product	Global product	Local product	Global product

Notes. Shows non–U.S.-based startups featured on platform from April 2018 to September 2020. The sample size drops because of singleton observations. Robust standard errors are clustered at the company level.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

conjecture that local initial users enable startups to build products that achieve higher foreign user growth when local signals are clearer and more transferable.

5.3. Product Iteration Channel

The theoretical section suggests that initial users inform startups' product iterations. These product iterations ultimately enable these companies to offer features that better address the needs of their foreign target users and thereby achieve higher foreign user growth. The analyses thus far have assessed the relationship between initial users and subsequent foreign user growth but not the product iteration in between. A supplementary analysis sheds further light on this product channel, leveraging technology tool data from BuiltWith. Specifically, this analysis uses technology tool installations on the websites of the startups in the platform sample as a proxy of product updates. The study then assesses whether (1) making changes to the number of tools on a website in the three months prior to feature predicts gaining higher foreign user growth after featuring on the platform and (2) having a higher share of local initial users predicts this tool installation.

The tool updates could reflect either a pivot in the product offering or the target customer base (McDonald and Gao 2019, Kirtley and O'Mahony 2023). The first type of update suggests that startups shift their target customer base after engagement with initial users. For example, companies with a higher share of local initial users might have been initially targeting the local market, but through engagement with initial users, shift their target to the foreign market (e.g., if local initial users do not like the product). The second type of update—which is more consistent with the theory of the paper—is that startups maintain the same target market, but they learn about ways to improve the product through the feedback and signals of initial users. Their product updates that incorporate these learnings then help augment their target user growth after featuring on the platform. A startup that has a large share of local initial users may maintain a focus on the foreign target market, but it learns about features to add to the product from these initial users to better address the needs of that target market. When the startup features on the platform, the updated product might then achieve higher foreign user growth.

To assess which type of pivot is driving the results, the study characterizes technology tools as those related to targeting a particular customer segment versus those focused on product features (for an existing customer segment). Tools that fall under the customer category include, for example, ads, hosting, language, and shipping. Changes in these tools suggest a customer pivot. For example, installing French language tools likely indicates that the company is now targeting the French-speaking market. Those that fall under the product category include, for instance, content management systems, feeds, media, and widgets (e.g., chatbots). Changes in these tools are more consistent with a product pivot, such as adding a chatbot to a website to provide a better product experience for pre-existing target customers.[18] The study then assesses whether changes in these customer- or product-oriented tools prior to feature systematically vary across startups with higher versus lower shares of local initial users and more strongly predict postfeature foreign user growth. This analysis sheds light on whether customer- or product-oriented pivots help account for the baseline results.

Table 7 reports results for this analysis. Column 1 shows that startups that make more tech updates in the three months prior to feature in terms of the logged

Table 7. Startups with More Product-Related Tech Updates Prior to Feature Achieve Higher Foreign User Growth After Feature, and a Higher Local Initial User Share Predicts More Tech Updates

	(1)	(2)	(3)	(4)	(5)
	Log foreign country visits		Log pre-tech update	Log pre-customer tech update	Log pre-product tech update
Post	0.337***	0.329***			
	(0.101)	(0.099)			
Post × *Pre-Tech Update*	0.142*				
	(0.063)				
Post × *Pre-Customer Tech Update*		−0.189			
		(0.140)			
Post × *Pre-Product Tech Update*		0.213**			
		(0.071)			
Pre-Local Share			0.266^{+}	0.054	0.270^{+}
			(0.150)	(0.073)	(0.147)
Log Pre-Visits			0.044*	0.027**	0.040*
			(0.018)	(0.009)	(0.018)
_cons	6.673***	6.672***	0.701***	0.065	0.667***
	(0.050)	(0.050)	(0.163)	(0.080)	(0.161)
N	11,060	11,060	969	969	969
Firm	Yes	Yes	No	No	No
Month × Year	Yes	Yes	No	No	No
Sample	Full	Full	Full	Full	Full
Country fixed effects	No	No	Yes	Yes	Yes
Product fixed effects	No	No	Yes	Yes	Yes

Notes. Shows startups featured on platform from April 2018–September 2020. The sample size drops because of singleton observations. Robust standard errors are clustered at the company level.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

absolute change in tool count (*Post* × *Pre-Tech Update*) achieve higher foreign user growth after featuring on the platform. The coefficient magnitude is about 42% of the baseline effect of featuring on the platform (*Post*) and is statistically significant at the 5% level. Column 2 shows that this foreign user growth is particularly driven by logged changes in product-related tools (*Post* × *Pre-Product Tech Update*) rather than customer-related tools (*Post* × *Pre-Customer Tech Update*). Column 3 shows that—on a firm level—startups with a higher share of local initial users make about 27% more absolute changes in their technology tool count prior to feature. These changes are driven by product-related tools (column 5) rather than customer-related tools (column 4).[19]

All together, these results suggest that a higher share of local initial users may prompt startups to iterate on their products—more so than pivoting their target customer segment or staying the course—in a way that enables the startups to achieve higher foreign user growth after launching on the platform.

5.4. Mechanism Robustness Analyses

The study conducts several analyses to further investigate the mechanisms and alternative explanations. Although the startup subsample analyses link clarity and transferability to foreign user growth and the technology tool analyses link product iteration to this growth, the combined process has yet to be observed. An exploratory set of four experiments that directly elicit product feedback and iteration helps shed light on the full process of how the clarity and transferability of initial user feedback shape product updates and ultimately foreign user interest. The experiments occur in four steps. To begin, one set of users gives feedback to one (French) product pitch[20] from the platform data: The first two experiments elicit this feedback from both local (French) and foreign (U.S.) users, whereas the second two only from local users. Next, another set of users evaluates the clarity and transferability of the feedback: Local (French) users score the clarity, and foreign (U.S.) users score the transferability. Subsequently, a human[21] in the first experiment and a large language model (LLM) in the remaining experiments update the original product pitch with each piece of feedback, producing multiple revised versions. Finally, a different set of foreign (U.S.) users evaluates the updated pitches in terms of their improvement from the original pitch and willingness to pay. Consistent with the main results, Online Appendix A.8 shows that foreign users score higher updated product pitches that incorporated clearer and more transferable feedback.

An alternative explanation is that it is not the clarity and transferability of initial user signals but rather the home market size of startups that independently accounts for a higher share of local initial users and subsequent foreign user growth. In particular, startups from large HQ markets might be more likely to have a

higher share of local initial users simply because there are more of them available. To this point, Table 1 shows that startups with an above-median share of local initial users are more likely to be from larger GDP and population markets. A 2SLS framework can help account for this possibility because the first stage takes into account the positive relationship between HQ market size and the composition of local initial users. Table A.9.1 in the Online Appendix shows that the results are similar in this framework. This suggests that home market size independently of initial user composition is likely not driving the results.

Yet another possibility explaining the results is that startups with a higher local initial user share are more motivated to reach foreign target users. They might therefore be actively diverting away from local country users. Indeed, Table A.10.1 in the Online Appendix shows that these startups, although achieving higher foreign user growth, do not gain higher local user growth after featuring on the platform, which results in a higher share of foreign users. This alternative substitution explanation suggests that these startups would make more foreign-oriented investments, like foreign hosting and language tools. We would also expect to see the share of foreign users increase steadily after feature as these investments take effect. In contrast to these predictions, startups with a higher share of local initial users do not adopt any more foreign language or hosting tools than others before (Table 1) or after featuring on the platform (Table A.10.2 in the Online Appendix), and their increase in foreign country share is stable after feature (Figures A.10.1 and A.10.2 in the Online Appendix). These trends are not consistent with the idea that active diversion from local users is driving the main foreign user growth effect.

If startups with a higher share of local initial users target the foreign market more, they might also be more likely to pay to schedule the feature date when they expect that there would be many foreign users on the platform. These days could particularly be foreign weekends or holidays—slack times when prior work shows people are more likely to engage in innovative behavior (Agrawal et al. 2018). To test this alternative mechanism, Table A.11.1 in the Online Appendix assesses whether startups with a higher local initial user share are more likely to feature on these dates. It finds that after controlling for the HQ country and product category, startups that have a higher local initial user share are no more likely to feature on foreign weekends and U.S. public holidays.[22] This evidence is not consistent with the idea that startups with a higher share of local initial users are selecting into heavy foreign-user feature dates.

In addition to having different target markets, startups with a higher versus lower share of local initial users may also vary in their initial scope. This scope may independently impact subsequent foreign user growth. For example, startups with a higher share of local initial users may have a narrower scope and therefore have more room to grow their foreign user base after featuring on the platform. Table A.12.1 in the Online Appendix shows that startups with a higher share of local initial users have no narrower of a scope than others, and this scope does not meaningfully predict subsequent foreign user growth. These findings are not consistent with the alternative explanation that a narrower scope independently impacts subsequent foreign user growth.

Lastly, foreign users' confusion rather than interest in startups with a higher share of local initial users may account for the foreign user growth outcome. If this were the case, we would expect that the foreign user growth would be short term. Table A.13.1 in the Online Appendix shows the same regressions conducted for 6 months after feature but now for 12 months after feature. It shows that the results persist 12 months after feature, which is more consistent with interest in the product rather than short-term confusion. Further, in a second supplementary experiment in Table A.14.1 in the Online Appendix, evaluators give higher ratings when randomly shown products of startups with a higher (versus lower) share of local initial users from the observational platform data. These higher ratings are not only in terms of interest in clicking on the startup's website but also in terms of novelty and willingness to pay. The results suggest that the clicking behavior in the observational data reveals at least some genuine interest in the product.

6. Discussion and Conclusion

6.1. Summary

The study assesses whether startups with a higher share of local initial users achieve higher subsequent foreign user growth. This choice of local versus foreign initial users presents startups with a distinct tradeoff between getting clearer versus more transferable signals. Local initial users offer clearer signals of demand, whereas foreign ones offer more transferable signals. Using a platform as a setting to test this theory, the study finds that startups with a higher share of local initial users achieve higher foreign user growth after featuring on the platform. This effect magnifies in more linguistically and ethnically homogeneous local HQ markets where startups and local initial users are more familiar with one another, and therefore these users can offer clearer signals. It also magnifies when there are smaller structural differences between the local and foreign markets in less locally fragmented product categories, making local signals more representative and therefore more transferable to the foreign target market.

Together, these patterns are consistent with the idea that the clarity benefits of local initial users exceed their transferability costs when in globally standardized product categories like the digital ones that are most common in the platform sample and among high-growth startups more broadly. This net effect, however, many not be the case in more locally fragmented product categories where local signals carry a higher transferability penalty.

6.2. Contributions to Theory

The study contributes to a growing body of work assessing how startups expand into their target markets. Prior work suggests the importance of initial users for this expansion (Bingham and Eisenhardt 2011, Gans et al. 2019, Moeen et al. 2020, Cao et al. 2024). This study illuminates which initial users are most helpful for informing this target market demand in order to ultimately achieve growth in these markets. Specifically, the results suggest that learning from initial users can enable startups to achieve higher target user growth when there is greater familiarity between the startup and its initial users, as well as when the initial market is more representative of the target market, increasing both the clarity and transferability of these users' signals. In the digital context studied in this paper—where preferences are generally standardized across the globe—local initial users may be particularly valuable sources of clear signals with little transferability penalty.

In doing so, the research also sheds light on how the digital era impacts a long-standing body of work on new market entry (Kogut and Chang 1996, Guler and Guillen 2010). Rather than rendering local geography as irrelevant (Cairncross 1997), the digital context actually makes this geographic boundary more important for companies to gain clear signals of demand that they can then more easily transfer to their new (target) market. As a result, firms can attract new market demand even without prior experience in that market. This might make experience outside of the target market less of a penalty (Zaheer 1995, Shaver et al. 1997) and even an advantage if it offers clearer signals of demand. Crucially, this depends on the local fragmentation of the product category.

Lastly, the study contributes to a growing body of research assessing the impact of intermediaries like platforms and accelerators on startup growth. Although prior work finds that factors like the founding environment, product category, size, gender orientation, and connectedness of startups shape the effects of intermediaries (Armanios et al. 2017, Fehder 2023, Wright et al. 2023b, Cao et al. 2024, Meyer et al. 2024, Wang et al. 2024), this research reveals that the composition of initial users also matters. Specifically, the ability to grow target users, which the platform affords, might accrue more to startups with a higher share of local initial users. This growth in target, and specifically foreign, users also reveals such platforms as an important avenue of startup internationalization, in addition to digital technologies like machine learning and communication technology found in prior work (Oviatt and McDougall 2005, Reuber and Fischer 2009, Brynjolfsson et al. 2019). By enabling users to gain visibility to startups (Meyer et al. 2023, 2024), these platforms may enable firms to internationalize early in their life cycle (Oviatt and McDougall 2005, Fan and Phan 2007) rather than expanding gradually (Johanson and Vahlne 1977).

6.3. Limitations and Opportunities

The study faces several limitations that open avenues for future work. Empirically, isolating the link between startups' learning from initial users and foreign user growth is challenging. Startups with more local versus foreign initial users may differ in ways that independently affect foreign growth. The platform's focus on early-stage technology startups seeking foreign expansion mitigates some concerns, while startup fixed effects, time-varying controls, coarsened exact matching, a 2SLS approach with HQ market size, and related analyses further address them. However, unobservable factors may shape initial user composition. Regression to the mean also cannot be fully ruled out, though subsample analyses suggest it is not driving results. Finally, the data do not directly capture feedback from initial users, motivating supplementary experiments in this study. Future work may exploit natural or field experiments to better identify the causal impact of the initial user composition.

The platform's features that strengthen internal validity also can limit external validity. The startup sample is concentrated in information technology, where cross-border differences in user preferences are relatively small (average product fragmentation = 0.4; Figure 5). This likely explains why startups with a higher share of local initial users show higher foreign user growth: Low transferability costs make the benefit of clear local signals outweigh any loss in representativeness. In more fragmented industries, such as manufacturing, transferability costs may be greater, enabling startups with a higher share of foreign initial users to achieve higher foreign user growth. Thus, product fragmentation emerges as a key boundary condition, with digital products limiting the transferability costs of nontarget initial user markets.

The platform setting also may exclude startups that choose to pursue different approaches to market expansion. For example, startups that choose to target proximate rather than global markets may not face a tradeoff between clarity and transferability, which is at the heart of this study. Additionally, startups that choose to gradually enter new markets rather than broadly launch to these markets also do not surface in the data.

This prevents assessing the timing of the shift from initial users to the target market, which may impact new user growth independently of the learning dynamics investigated in this study. These empirical boundary conditions are valuable areas for future work.

The study examines market segments defined by international geography, but the clarity-transferability tradeoff may also arise across industry segments. Some industries may be more familiar to entrepreneurs, making signals from initial users clearer but less representative of the target segment. For instance, an entrepreneur with manufacturing experience who is developing an IT product for finance may find manufacturing user feedback clearer yet less transferable due to structural differences between industries—even within the same country. If cross-country differences exceed cross-industry ones (Ghemawat 2001, Berry et al. 2010), the effects in this study likely represent an upper bound. Exploring this tradeoff within industry-defined markets is a promising avenue for future research.

The study additionally assesses young ventures rather than mature firms. The early-stage nature of these firms with less organizational inertia (Zhang et al. 2016, McDonald and Gao 2019) and a high sensitivity to external signals (Gans et al. 2019, Ott and Eisenhardt 2020, Pillai et al. 2020, Kirtley and O'Mahony 2023) might explain why initial users profoundly matter for their product development and understanding of target markets. Further work may assess whether mature firms—that also regularly expand into new markets (Berry and Kaul 2022)—exhibit similar effects. If so, in what circumstances? For example, one could imagine that the importance of initial users might emerge for mature firms in highly uncertain contexts. This would reveal the importance of uncertainty in explaining these results. But if not, organizational inertia may inhibit learning, suggesting that the absence of such inertia explains the results observed among young firms.

The analyses also assume that the HQ country of a startup is where most of the team is located and therefore where it has the most familiarity. However, startup teams are becoming increasingly international, for example, with the rise of distributed models and remote work (Choudhury et al. 2021, Chauvin et al. 2024). This trend may introduce heterogeneity in the clarity benefits that startups gain from initial users in a given location. In doing so, this international team composition may dilute the value of initial users from a startup's HQ market even in globally standardized product categories like enterprise software—a fruitful area to investigate in future work.

The heterogeneity between local and foreign initial users is at the core of this paper because the clarity-transferability tradeoff that it presents to startups can have important implications for target user growth. However, other heterogeneities in the initial user market and their tradeoffs may also have important implications for expanding into target markets and thus be valuable to study. Future work may assess the tradeoffs of, for example, small versus large or homogeneous versus heterogeneous initial user markets. These initial user characteristics can have profound implications for entrepreneurial strategy and growth.

Lastly, the study assesses startups' learning at a single point in time (in the three months prior to platform launch). Given the high uncertainty that young ventures face, anticipating which initial users offer the "optimal" signals is often difficult in this early period. But we would expect that over time, startups might learn to "learn." Specifically, they might anticipate that foreign user signals are less clear, so they will weigh them less or even not seek them out altogether. Theorizing and testing the evolution of this learning process is a valuable area for future work.

6.4. Practical Implications

These findings reveal several implications for practice. Importantly, they shed light on the beachhead markets[23] which can most effectively position startups to gain target market demand. The study suggests that a beachhead market that is more familiar to the startup and still representative of the target user market enables startups to update their products in ways that attract target market demand. When applied to foreign expansion decisions, startups may be able to start with a local beachhead to derive clear and transferable lessons before entering a foreign target market. This is particularly the case in the context of digital products that tend to be globally standardized, making any initial market relatively representative of the target one.

The results also show that platforms can be valuable tools for startups to grow in foreign markets. In doing so, they can help startups overcome the constraints of their local market size. But rather than completely bypassing the local market, this research suggests certain contexts in which startups may particularly benefit from focusing on local initial users to improve their product. Doing so may enable them to achieve higher foreign user growth on these platforms when the startups work in product categories that have globally standardized product preferences, like in the digital ones central to this research. In these contexts, startups gain signals from local initial users that are not only clearer but also transferable to foreign target markets.

Further, the study reveals novel ways for researchers, managers, and investors to track the geographic market orientation of young companies over time. The study uses the geographic composition of page visit data, foreign technology tool website adoption, and the regional orientation of website text using a supervised machine

learning algorithm in order to define the market orientation and geography of young companies. Such measures may be more informative for young companies rather than HQ or subsidiary locations as traditionally used to define geography for mature companies because young technology companies often operate almost entirely virtually in distributed models, especially in the post-COVID era.

In conclusion, the study shows that startups with a higher share of local initial users achieve higher subsequent foreign user growth. This is especially the case in local markets that exhibit higher familiarity between the startup and initial users and are more representative of the target market, conditions under which the clarity benefits of local signals outweigh their transferability costs. Together, these results reveal that the composition of initial users—through the clarity and transferability of their signals—can have a profound effect on the scaling trajectories of entrepreneurial ventures around the world.

Acknowledgments

The author thanks Victor Bennett, Nandil Bhatia, Wilbur Chung, Susan Cohen, Bo Cowgill, Shane Greenstein, Jorge Guzman, Ranjay Gulati, Kathy Harrigan, Pablo Hernández-Lagos, Laura Huang, Daniel Keum, Tarun Khanna, Rem Koning, Hong Luo, Amisha Miller, Thomaz Teodorovicz, Dan Wang, and the participants of the Academy of Management, Bayesian Entrepreneurship Conference, Columbia Business School Management Seminar, Duke Strategy Conference, New York University Stern Global Strategy Workshop, Smith Entrepreneurship Research Conference, and Strategic Management Society Annual Conference and Extension for valuable comments; Daniel Creech and Noa Rosinplotz for excellent research assistance; and the anonymous reviewers and Mahka Moeen at *Organization Science* for thoughtful feedback. All contents of the publication are the responsibility of the author.

Endnotes

[1] Local initial users are based in the home market of the startup. Foreign initial users are based in the startup's target market, which is distinct from the startup's home market. The study defines the startup's home market as the startup's headquarters country, where startup founders often originate and have extensive educational and work experience (Michelacci and Silva 2007, Dahl and Sorenson 2012). In some cases, however, startup founders might come from other locations beyond the headquarters country. In other cases, startups might change their headquarters country; however, this happens in less than 1.5% of firms according to Shi et al. (2024). In these cases, the founders' and startups' original locations may also theoretically characterize the startup's home market.

[2] The homogeneity of the initial market also reduces variance across initial user signals. It therefore makes it more likely that most of the initial users either like or do not like the product, rather than only some liking and others not liking the product. When signals move in the same direction, startups can more easily understand the aggregate patterns in that market's demand. Crucially, this lower variance makes it easier for any startup—not just ones that come from the same market as the initial user—to make sense of the given initial market's demand.

[3] The World Bank reports that GDP per capita in Ukraine was 5,390 USD, and it was 85,810 USD in the United States in 2024 (the latest year of data). For more information, please refer to https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.

[4] There are slightly fewer startups from Latin America and the Caribbean and more startups from the Middle East and North Africa. Each region is a fairly small share of the total sample (less than 3%). Table A.2.1 in the Online Appendix shows that the results are robust to excluding these regions from the final analysis. They are also similar when excluding either U.S. startups or European startups—the top HQ locations of startups on the platform.

[5] All log values indicate log(value + 1) to account for zero values.

[6] The results are similar if, instead of averaging monthly shares of visits, the study computes the three-month HQ share as total HQ country visits over the three months prior to feature divided by total visits in that period.

[7] The results are also robust to different measures of local versus foreign initial users. Specifically, they hold when measuring foreign initial users based on cultural, political, knowledge, demographic, economic, geographic, and linguistic differences between the HQ country market and the top user country market three months prior to feature, using measures from Berry et al. (2010) (Table A.3.1 in the Online Appendix).

[8] The primary language is defined as the first listed language of the HQ country because these languages are generally shown in order of popularity. Results are similar when using an indicator for whether English is an official language in that country.

[9] The results are similar when using the original sample of 8,396 startups (at first feature) with available HQ country and website information, as shown in Table A.5.2 in the Online Appendix. The main local fragmentation measure made from the final sample and the measure made from the original sample are highly correlated (correlation = 0.9).

[10] "SaaS" refers to software-as-a-service, a category that typically reflects enterprise-oriented software products.

[11] Prolific respondents specifically complete the following task: "Please select all of the product categories where you would expect to see LARGE differences in customer tastes across countries." The Prolific respondents were 30 years old on average, 44% had full-time jobs, and 39% were students. The categories with the highest percent (52%) were "Food and Beverage" and "Lifestyle." The categories with the lowest percentage agreement (6%–8%) include "PaaS," "IaaS," and "SaaS." "Local" product categories are ones where the first keyword is at least a 30% agreement rate and "global" product categories as the rest. As expected in the globally skewed nature of IT, this leaves 21% in a local product category and the remaining in global product categories. A similar prompt to a large language model (LLM) and a research assistant corroborate the Prolific responses.

[12] The study identifies foreign language tools as those that BuiltWith tags as a "language" tool, except for those that (1) match the primary language spoken in an HQ country using the same data set and approach as used to define linguistic structural differences or (2) contain a language name that corresponds to a specific dialect of an HQ country (e.g., "Swiss French" for Switzerland). In the rare case that a language installed on a website is not the primary language in any HQ country in the data set, the study instead uses the country where that language is an official language (e.g., Urdu is not the primary language in any HQ country in the data set but is an official language in Pakistan). The results are similar when using the official language rather than primary language in an HQ country.

[13] The study identifies foreign hosting tools as tools that BuiltWith specifically tags as a "hosting" tool, which includes the word "global" or the name of a region, country, city, or primary language

that is not associated with a startup's HQ country. The approach to identify a primary language is the same as the one used to measure linguistic structural differences.

[14] Table A.4.1 in the Online Appendix shows summary statistics of the main dependent, explanatory, and control variables of all of the firms. Tables A.4.2 and A.4.3 in the Online Appendix show these statistics for startups with above- and below-median shares of local initial users prior to feature.

[15] This foreign-oriented text variable reflects whether the predicted region of a startup's website text is different from the actual HQ region of the startup. It is created using a supervised machine learning algorithm that detects the region of a corpus of text trained on startup descriptions of about 500,000 startups with the associated HQ regions as the labels from PitchBook. After fitting the data to several supervised algorithms, including random forest, neural network, and logistic regression, the logistic regression achieves the highest accuracy rate at 62%. This measure covers nearly half of the companies in the data set where historical website data in the sample period were available from the Internet Archives. The supervised approach has an advantage over unsupervised similarity measures in that it does not assume weights for the underlying text vectors, as would an unsupervised approach. The supervised algorithm infers weights from the training data of startup descriptions. Because website snapshots are not available on a regular monthly basis and are only available for a smaller subset of firms, the analyses using this variable are inherently noisier.

[16] The study creates this B2C versus B2B product categorization by asking 50 human coders to review the product categories from the observational data and then indicate which of these categories target consumers. It then uses the ones with at least 50% agreement among the coders as keywords that indicate B2C. These keywords for B2C include, for example, "shopping," "universities," "restaurants," and "games."

[17] The results are robust to instead including HQ country fixed effects, as shown in Tables A.6.1–A.6.3 in the Online Appendix.

[18] The study identifies tools under the customer category as those that BuiltWith tags as "ads," "agency," "hosting," "language," "payment," and "shipping," and the remaining tools under the product category. These measures are highly correlated with those created when asking an LLM to code each technology tool based on its name, description, and tag using the following prompt: *Please tag each tool as a "product" or "customer" tool. Product tools improve or shape the product itself. Changes in product tools suggest a pivot in the product: The startup keeps the same target market but adapts how the product is built or delivered. Customer tools indicate or affect the target customer base. Changes in customer tools suggest a pivot in the target market after engagement with initial users. When tagging, keep in mind that some tools can influence both product delivery and customer orientation. In such cases, prioritize whether the tool mainly signals the market being served (customer) or mainly alters the product's functionality (product).* To account for variability in LLM coding, the study runs the prompt five times and averages the resulting "customer" and "product" scores. It then ultimately codes a tool as a "customer" tool if the average customer score is greater than or equal to 0.5 and 0 otherwise; the same approach applies to a "product" tool. The customer and product LLM scores have over a 0.9 correlation with those using the original approach. The final results are also similar with the measures using the LLM coding.

[19] The table clusters standard errors at the company level consistent with the baseline analyses. The results are directionally consistent when clustering at the HQ country level.

[20] The study uses a French product pitch because France is one of the top non-U.S. startup HQ countries from the platform observational data.

[21] The human who updates the original pitch with each piece of feedback is the same local user evaluating feedback in the previous step.

[22] The analyses assessing feature dates during U.S. public holidays exclude U.S.-based startups.

[23] A beachhead market is "a small market segment [an entrepreneur] can control until [the] company has sufficient resources to enter other markets" (Aulet 2017, p. 40).

References

Adner R, Levinthal D (2001) Demand heterogeneity and technology evolution: Implications for product and process innovation. *Management Sci.* 47(5):611–628.

Adner R, Puranam P, Zhu F (2019) What is different about digital strategy? From quantitative to qualitative change. *Strategy Sci.* 4(4):253–261.

Agrawal A, Catalini C, Goldfarb A, Luo H (2018) Slack time and innovation. *Organ. Sci.* 29(6):1056–1073.

Armanios DE, Eesley CE, Li J, Eisenhardt KM (2017) How entrepreneurs leverage institutional intermediaries in emerging economies to acquire public resources. *Strategic Management J.* 38(7): 1373–1390.

Athey S, Stern S (2013) The nature and incidence of software piracy: Evidence from windows. NBER Working Paper No. 19755, National Bureau of Economic Research, Cambridge, MA.

Aulet B (2017) *Disciplined Entrepreneurship Workbook* (John Wiley & Sons, Hoboken, NJ).

Balachandran S, Hernandez E (2021) Mi casa es tu casa: Immigrant entrepreneurs as pathways to foreign venture capital investments. *Strategic Management J.* 42(11):2047–2083.

Berry H, Kaul A (2022) Disaggregating multinationality: An empirical examination of aggregation, adaptation, and arbitrage activities by us multinational corporations. *Strategy Sci.* 7(2):90–105.

Berry H, Guillén MF, Zhou NAN (2010) An institutional approach to cross-national distance. *J. Internat. Bus. Stud.* 41(9):1460–1480.

Bingham CB, Eisenhardt KM (2011) Rational heuristics: The simple rules that strategists learn from process experience. *Strategic Management J.* 32(13):1437–1464.

Boudreau KJ, Jeppesen LB, Miric M (2022) Competing on freemium: Digital competition with network effects. *Strategic Management J.* 43(7):1374–1401.

Brynjolfsson E, Hui X, Liu M (2019) Does machine translation affect international trade? Evidence from a large digital platform. *Management Sci.* 65(12):5449–5460.

Cairncross F (1997) *The Death of Distance: How the Communications Revolution Will Change Our Lives* (Harvard Business Review Press, Boston).

Camuffo A, Cordova A, Gambardella A, Spina C (2020) A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Sci.* 66(2):564–586.

Cao R, Koning R, Nanda R (2024) Sampling bias in entrepreneurial experiments. *Management Sci.* 70(10):7283–7307.

Chauvin J, Choudhury P, Fang TP (2024) Working around the clock: Temporal distance, intrafirm communication, and time shifting of the employee workday. *Organ. Sci.* 35(5):1660–1681.

Choudhury P, Foroughi C, Larson B (2021) Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management J.* 42(4):655–683.

Conti A, Guzman JA (2023) What is the US comparative advantage in entrepreneurship? Evidence from Israeli migration to the United States. *Rev. Econom. Statist.* 105(3):1–45.

Coval JD, Moskowitz TJ (2001) The geography of investment: Informed trading and asset prices. *J. Political Econom.* 109(4):811–841.

Cuervo-Cazurra A (2011) Selecting the country in which to start internationalization: The non-sequential internationalization model. *J. World Bus.* 46(4):426–437.

Dahl MS, Sorenson O (2012) Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management Sci.* 58(6):1059–1071.

Drazanova L (2020) Introducing the historical index of ethnic fractionalization (HIEF) dataset: Accounting for longitudinal changes in ethnic diversity. *J. Open Humanities Data* 6(1):6.

Dushnitsky G, Stroube BK (2021) Low-code entrepreneurship: Shopify and the alternative path to growth. *J. Bus. Venturing Insights* 16:e00251.

Fan T, Phan P (2007) International new ventures: Revisiting the influences behind the born-global firm. *J. Internat. Bus. Stud.* 38:1113–1131.

Fehder DC (2024) Coming from a good pond: The influence of a new venture's founding ecosystem on accelerator performance. *Admin. Sci. Quart.* 69(1):1–38.

Fischer M, Völckner F, Sattler H (2010) How important are brands? A cross-category, cross-country study. *J. Marketing Res.* 47(5): 823–839.

Gans JS, Stern S, Wu J (2019) Foundations of entrepreneurial strategy. *Strategic Management J.* 40(5):736–756.

Ghemawat P (2001) Distance still matters: The hard reality of global expansion. *Harvard Bus. Rev.* 79(8):137–147.

Ghemawat P (2007) Managing differences: The central challenge of global strategy. *Harvard Bus. Rev.* 85(3):58–68.

Giustiziero G, Kretschmer T, Somaya D, Wu B (2023) Hyperspecialization and hyperscaling: A resource-based theory of the digital firm. *Strategic Management J.* 44(6):1391–1424.

Gompers PA, Gornall W, Kaplan SN, Strebulaev IA (2020) How do venture capitalists make decisions? *J. Financial Econom.* 135(1):169–190.

Guler I, Guillen MF (2010) Home country networks and foreign expansion: Evidence from the venture capital industry. *Acad. Management J.* 53(2):390–410.

Hallen BL, Cohen SL, Bingham CB (2020) Do accelerators work? If so, how? *Organ. Sci.* 31(2):378–414.

Impink SM, Miric M (2024) Tradeoffs to using multiple information technology providers for high tech startups. Preprint, submitted November 15, https://dx.doi.org/10.2139/ssrn.4977423.

Johanson J, Vahlne JE (1977) Process of the internationalization development firm—A model of knowledge foreign and increasing market commitments. *J. Internat. Bus. Stud.* 8(1):23–32.

Kaplan SN, Strömberg PE (2004) Characteristics, contracts, and actions: Evidence from venture capitalist analyses. *J. Finance* 59(5):2177–2210.

Kerr WR, Nanda R, Rhodes-Kropf M (2014) Entrepreneurship as experimentation. *J. Econom. Perspective* 28(3):25–48.

Khanna T (2014) Contextual intelligence. *Harvard Bus. Rev.* 92(9):58–68.

Kirtley J, O'Mahony S (2023) What is a pivot? Explaining when and how entrepreneurial firms decide to make strategic change and pivot. *Strategic Management J.* 44(1):197–230.

Klepper S, Thompson P (2006) Submarkets and the evolution of market structure. *RAND J. Econom.* 37(4):861–886.

Kogut B, Chang SJ (1996) Platform investments and volatile exchange rates: Direct investment in the U.S. by Japanese Electronic Companies. *Rev. Econom. Statist.* 78(2):221.

Koning R, Hasan S, Chatterji A (2022) Experimentation and start-up performance: Evidence from A/B testing. *Management Sci.* 68(9):6434–6453.

Landsman W, Pan J, Stubben S (2024) Equity market fragmentation and capital investment efficiency. *Management Sci.* 70(7):4381–4406.

Lee S, Kim JD (2024) When do startups scale? Large-scale evidence from job postings. *Strategic Management J.* 45(9):1633–1669.

Manral L, Harrigan KR (2023) Geographic fragmentation and declining dominance: Yet another story of AT&T's decline in the post-divestiture era. *J. Evolutionary Econom.* 33(2):605–644.

McDonald R, Gao C (2019) Pivoting isn't enough? Managing strategic reorientation in new ventures. *Organ. Sci.* 30(6):1289–1318.

McMullen JS, Shepherd DA (2006) Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. *Acad. Management Rev.* 31(1):132–152.

Meyer T, Kerkhof A, Cennamo C, Kretschmer T (2024) Competing for attention on digital platforms: The case of news outlets. *Strategic Management J.* 45(9):1731–1790.

Meyer KE, Li J, Brouthers KD, Jean RJB (2023) International business in the digital age: Global strategies in a world of national institutions. *J. Internat. Bus. Stud.* 54(4):577.

Michelacci C, Silva O (2007) Why so many local entrepreneurs? *Rev. Econom. Statist.* 89(4):615–633.

Miller A, O'Mahony S, Cohen SL (2023) Opening the aperture: Explaining the complementary roles of advice and testing when forming entrepreneurial strategy. *Organ. Sci.* 35(1):1–26.

Mitchell W, Shaver JM, Yeung B (1994) Foreign entrant survival and foreign market share: Canadian companies' experience in United States medical sector markets. *Strategic Management J.* 15(7):555–567.

Moeen M, Agarwal R, Shah SK (2020) Building industries by building knowledge: Uncertainty reduction over industry milestones. *Strategy Sci.* 5(3):218–244.

Moore G (2014) *Crossing the Chasm* (Harper Business, New York).

Murgia M (2024) AI start-ups generate money faster than past hyped tech companies. *Financial Times*, https://www.ft.com/content/a9a192e3-bfbc-461e-a4f3-112e63d0bb33.

Ott TE, Eisenhardt KM (2020) Decision weaving: Forming novel, complex strategy in entrepreneurial settings. *Strategic Management J.* 41(12):2275–2314.

Ott TE, Eisenhardt KM, Bingham CB (2017) Strategy formation in entrepreneurial settings: Past insights and future directions. *Strategic Entrepreneurship J.* 11(3):306–325.

Oviatt BM, McDougall PP (2005) Toward a theory of international new ventures. *J. Internat. Bus. Stud.* 36:29–41.

Pillai SD, Goldfarb B, Kirsch DA (2020) The origins of firm strategy: Learning by economic experimentation and strategic pivots in the early automobile industry. *Strategic Management J.* 41(3):369–399.

Reuber AR, Fischer E (2009) Signalling reputation in international online markets. *Strategic Entrepreneurship J.* 3(4):369–386.

Reuber AR, Fischer E (2011) International entrepreneurship in internet-enabled markets. *J. Bus. Venturing* 26(6):660–679.

Shane S (2000) Prior knowledge and the discovery of entrepreneurial opportunities. *Organ. Sci.* 11(4):448–469.

Shaver JM, Mitchell W, Yeung B (1997) The effect of own-firm and other-firm experience on foreign direct investment survival in the United States. *Strategic Management J.* 18(10):811–824.

Shermon A, Moeen M (2022) Zooming in or zooming out: Entrants' product portfolios in the nascent drone industry. *Strategic Management J.* 43(11):2217–2252.

Shi Y, Sorenson O, Waguespack DM (2024) The new argonauts: The international migration of venture-backed companies. *Strategic Management J.* 45(8):1485–1509.

Sorenson O, McEvily S, Ren CR, Roy R (2006) Niche width revisited: Organizational scope, behavior and performance. *Strategic Management J.* 27(10):915–936.

Startup Genome (2024) The global startup ecosystem report. Accessed April 2025, https://startupgenome.com/report/the-global-startup-ecosystem-report-2024.

Stinchcombe AL (1965) Social structures and organizations. March JG, ed. *Social Structures and Organizations* (Rand-McNally, Chicago), 142–193.

Stroube B, Dushnitsky G (2023) Built with builtwith: A technical note. Preprint, submitted June 20, https://doi.org/10.2139/ssrn.4475309.

Stroube BK, Dushnitsky G (2025) Mapping entrepreneurial inclusion across us neighborhoods: The case of low-code e-commerce entrepreneurship. *Strategic Management J.* 46(7):1762–1789.

Tidhar R, Hallen BL, Eisenhardt KM (2025) Measure twice, cut once: Unit profitability, scalability, and the exceptional growth of new firms. *Organ. Sci.* 36(1):88–120.

Tippmann E, Ambos TC, Del Giudice M, Monaghan S, Ringov D (2023) Scale-ups and scaling in an international business context. *J. World Bus.* 58(1):101397.

Wang D (2015) Activating cross-border brokerage: Interorganizational knowledge transfer through skilled return migration. *Admin. Sci. Quart.* 60(1):133–176.

Wang Y, Zhao EY, Yu T (2024) The effects of product recategorization by the platform owner on complementor firms' product updates. Working paper, Ohio State University, Columbus.

Winter SG, Szulanski G, Ringov D, Jensen RJ (2012) Reproducing knowledge: Inaccurate replication and failure in franchise organizations. *Organ. Sci.* 23(3):672–685.

Wormald A, Agarwal R, Braguinsky S, Shah SK (2021) David overshadows goliath: Specializing in generality for internationalization in the global mobile money industry. *Strategic Management J.* 2(8):1459–1489.

Wright NL, Saiedi E (2024) The mid-sized market trap in entrepreneurial scaling. Preprint, submitted December 31, https://doi.org/10.2139/ssrn.5078041.

Wright NL, Koning R, Khanna T (2023) Judging foreign startups. *Strategic Management J.* 44(9):2195–2225.

Wright NL, Nagle F, Greenstein S (2023) Open source software and global entrepreneurship. *Res. Policy* 52(9):104846.

Zaheer S (1995) Overcoming the liability of foreignness. *Acad. Management J.* 38(2):341–363.

Zhang C, Tan J, Tan D (2016) Fit by adaptation or fit by founding? A comparative study of existing and new entrepreneurial cohorts in China. *Strategic Management J.* 37(5):911–931.

Ziedonis RH (2004) Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms. *Management Sci.* 50(6):804–820.

Zou S, Cavusgil ST (2002) The GMS: A broad conceptualization of global marketing strategy and its effect on firm performance. *J. Marketing* 66(4):40–56.

Nataliya Langburd Wright is an assistant professor at Columbia Business School. She received her PhD from Harvard Business School. Her work examines how strategic market and technology choices shape global entrepreneurial growth and scaling disparities.